ECON 17100
Introduction to International Trade

DO NOT CIRCULATE THESE NOTES

Jordan Rosenthal-Kay
Autumn 2023
University of Chicago

Last updated: December 3, 2023

DO NOT CIRCULATE

Note

This material has been prepared for UChicago's Econ 17100 course. Most of this material has its roots in Krugman, Obstfeld, and Melitz (2017), Feenstra (2015), and of course *Chicago Price Theory* (Jaffe, Minton, Mulligan, and Murphy, 2019).

– Jordan Rosenthal-Kay

Contents

List of Figures

Lecture 1 – Introduction

There is a lot of international trade in the world. The Coronavirus pandemic only made the importance of trade more clear: disrupted supply chains and changes in demand led to ships congesting ports, rising prices and global turmoil. Prior to the Coronavirus pandemic, in 2019, global GDP (the total value of all goods and services sold) was around 90 trillion US dollars, while the total volume of trade (the total value of all goods traded internationally) was around 19 trillion USD. In all, international trade makes up about 20% of the global economy.

All countries trade, too, though some more than others. In entrepôts like Hong Kong and Singapore, the total value of exports exceed GDP by a factor of nearly three. Even the relatively isolated United States, which consumes much of what it produces, exports are around 10% of GDP. I could also measure the role of trade through imports, or net exports. Though net exports (exports minus imports) contribute to GDP, it is not a measure integration in the global economy. The scale of both imports and exports are informative to the importance of trade in a nation's economy. What a country imports depends on the demands of its consumers. A nation's exports depend upon its comparative advantage, a key idea of this course.

Trade theory ought to answer two questions:

1. *Why do countries trade?*
2. *What explains the content of trade?*

Hopefully, knowing *why* countries trade with each other can help us understand *what* they trade with each other, too. Knowing the answers to these questions is intellectually interesting their own right, but they're also important for trade policy. If we know what countries trade with each other, and why, we should be able to think about the effects of removing or putting up barriers to trade. Studying trade theory can help you answer questions like, Should the United States enter a free-trade agreement with Pacific and East Asian nations? Or, Will placing a tariff on imported cars protect the jobs of U.S. autoworkers?

Trade theory should help you think about more than international trade, too. Most economic activity is, in a sense, traded. When you take the train downtown to go out for dinner, you are importing services from one neighborhood to your own. Farmers in the heartland trade their produce across the country to feed people in cities, while cities produce a lot of financial services, innovation, and entertainment, some of which is traded, too. Trade theory helps you understand economic activity at many geographic scales.

Trade theory has produced two important, interrelated ideas to answer our two questions. They are,

1. Countries trade with one another because there are *gains to trade*.

2. The pattern of trade is well-explained by *comparative advantage*.

This introduction explores these ideas.

1.1 Are there gains from trade?

It might seem obvious that there are gains to trade – why else would two parties trade if neither stood to gain? However, a quick examination of how trade operates in the real world doesn't paint such a rosy picture. Trade agreements are regularly disputed. In the US in the 1990s, protests in Seattle against the World Trade Organization (WTO) turned violent. Almost twenty years later, the Obama administration's Trans-Pacific Partnership (TPP) wasn't ratified in congress and opposition to it was shared among 2016 presidential candidates from Bernie Sanders to Donald Trump. In the same year, Britain voted to leave the EU's common market. Former President Trump has suggested raising tariffs to 10% if reelected. In response to these political shifts, economists tend to be shouting from the sidelines a common refrain: free trade is good!

Yet, very few goods in the world are traded freely. Governments levy taxes on imports (tariffs). The average tariff rate in the United States is 1.5%, one of the lowest in the world. In the Central African Republic, the average tariff rate is more than ten times larger ($\sim 16\%$). Nontariff barriers like congestion at ports, bureaucratic hurdles, health inspections, and other obstacles stymie trade too: in the recent US-China trade war, these were estimated to cost over three times as much as tariffs (Chen, Hsieh, and Song, 2022). There are real barriers to trade even absent government intervention: it takes time, fuel, longshoreman labor, and a very big boat to move goods across the Pacific. In container shipping, maritime bottlenecks form at a handful of locations: Panama, Suez, the Bosphorous, and the Straits of Malacca. Can it really be efficient to grow fruit in Brazil, package it in Thailand, and then ship it to the U.S. when there are perfectly good farms here?

In short: we see tariffs routinely applied to goods at borders, and trade wars aren't uncommon. Many countries subsidize exported goods. And even for goods unaffected by trade barriers, the way they move across borders appears to contradict common sense. Why do we see such behavior "in the data?" A goal of this course is to address these questions using economics.

1.2 Comparative advantage

The first and central insight is that when countries trade, they do so on the basis of *comparative advantage*. Specializing in goods in which a country has a comparative advantage leads to gains from trade. This simply means exporting goods that have the relatively smallest opportunity cost of producing. This concept stands in contrast to *absolute advantage*, the idea that a country may be absolutely better at producing something than other countries. The legendary economist Paul Samuelson once said that comparative advantage was one of the deepest insights economics had to offer, as it may at first seem counterintuitive. Since Ricardo proposed the idea at the turn of the 19th century, many great thinkers have disputed it, and many politicians have enacted policies that that work against comparative advantage. However, it is not so hard of a concept to understand when using it to make sense of economic organization in the real world.

Consider the kitchen at a big, fancy restaurant with many cooks. The head chef plates the dishes she invented, while line cooks slice vegetables, make sauces, and sautée sundry ingredients. There is a lot of shouting involved ("yes, Chef!"), and the head chef might even send dishes back if the onions are sliced too thick or the sauce is too watery. Why doesn't she just make it herself? The head chef is likely better at slicing onions, preparing a roux, and sautéeing than every line cook she employs – heck, she created the menu herself! She has an absolute advantage in these tasks. However, the opportunity cost of the head chef working the line is losing out the person best suited to plate dishes, taste, and manage the kitchen. No kitchen is organized so that the head chef preps the vegetables, but why must the head chef specialize in plating? Couldn't she do a little bit of everything? If the head chef traded tasks with a line cook, the quality of plating would decline more than the gains in vegetable cutting. Each minute she spends cutting instead of plating gives up more of the dish's quality than can be gained from the line cook plating. That is to say, the head chef is relatively better at plating than cutting than the line cook. Similar economic forces compel countries to specialize in what they have a comparative advantage in, too.

1.3 Why study trade today?

On April 27th, 2023 U.S. National Security Advisor Jake Sullivan announced a 'New Washington Consensus' (Sullivan, 2023). What he meant was ending a trade policy that pursued 'trade liberalization as an end in itself' and instead building 'A modern American industrial strategy [that] identifies specific sectors that are foundational to economic growth' and 'protecting our foundational technologies with a small yard and high fence.' In practice, this has led to renewed state-driven investment in microchips and 'critical minerals' like lithium.

His remarks make clear we are in a new era of trade policy. No longer is free trade a policy pursued by most nations. Instead, trade policy has been crafted into an economic weapon and a tool for enforcing climate regulations. While rich nations have turned again towards protectionism, developing nations have begun to embrace trade and export-led growth strategies. As the U.S. and China entered a trade war, the African Continental Free Trade Area (AfCFTA) was signed into effect.

Understanding the effects of trade and trade policy are critical for understanding how the 21st century will unfold!

Lecture 2 – Deriving a demand curve

This is a course primarily about *supply*. However, when we do theory, we will always imagine consumers with preferences that generate downward-sloping demand in the background.

The goal of this section is to link *indifference curves* and *demand curves* so we can move between them freely as the course progresses.

2.1 Utility, relative prices and opportunity costs

We begin by considering a consumer that has preferences over two goods, X and Y. They consume bundles of goods, like $(3, 5)$ or $(2, 7)$ which reflect consuming 3 units of X and 5 units of Y, or 2 units of X and 7 units of Y. How do we know which bundle is preferred by the consumer? For that we use the concept of a *utility function*, U that relates bundles of goods with positive numbers. That is, $U(X, Y)$ is a number that reflects how many 'utils' a consumer receives from consuming the bundle (X, Y). Utils are obviously not 'real' in the sense that they can be measured. However, utils are a useful concept because they allow us to *rank* goods. For example, if,

$$U(3, 5) > U(2, 7)$$

then we know that the bundle $(3, 5)$ is preferred to $(2, 7)$. The ranking of goods, in the eyes of consumers, is enough to predict their behavior, if we assume consumers are *utility-maximizing*. Consumers don't just get to pick what bundle they want freely – they're constrained. Namely, the cost of the bundle can't exceed their income! While a consumer may prefer $(3, 5)$ to $(2, 7)$, it might be unaffordable if they have to forgo many units of Y to consume one more unit of X. That is the opportunity cost of X.

If the price of X in dollars is P_X (that is, the *nominal* price) and the price of Y is P_Y, then the opportunity cost of one unit of X is P_X/P_Y units of Y! Think of it like this: If a consumer forgoes one unit of X, they save P_X dollars. How many units of Y can they buy with that? Each dollar is worth $1/P_Y$ units of Y. If Y costs 2 dollars, and you have 1, at most you can buy only $1/2$ a unit of Y. So each unit of X is worth P_X/P_Y units of Y; P_X/P_Y is the relative price of X measured in units of Y. This idea is key: *relative prices reflect* *opportunity costs*.

2.2 Budget constraints

Suppose a consumer has M dollars of income, and can consume goods X or Y with prices P_X or P_Y. Then they have a *budget constraint*: the total value of their consumption, $P_X \times X + P_Y \times Y$ must be less than or equal to their income, M: $P_X \times X + P_Y \times Y \leq M$. We can represent this graphically as a triangular shaded region as in Figure 2.1:

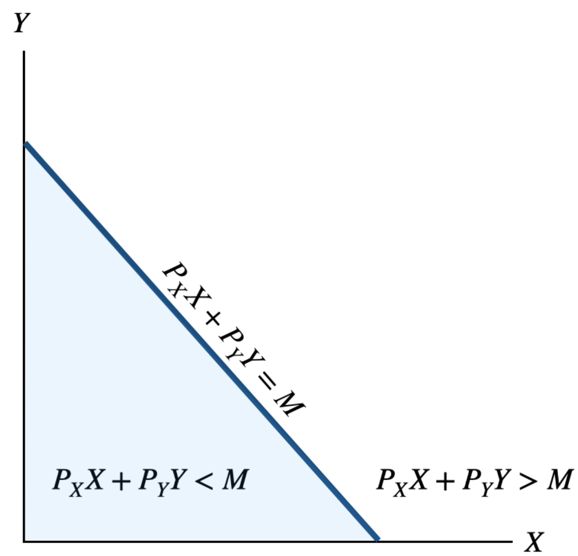


Figure 2.1: A consumer's budget constraint.

Every point on the graph reflects an (X, Y) combination. In the shaded region, such combinations are feasible – a consumer with income M can afford them at prevailing prices. Along the thick line, consumption exactly equals income. Points above the line reflect combinations of goods that are unaffordable at prevailing income and prices. The slope of the budget constraint (the blue line) is $-P_X/P_Y$. The slope of the line reflects the opportunity cost of X in units of Y. The intercepts of the line reflect consumption if all income is spent on that good. If a consumer consumes zero units of Y, they can afford M/P_X units of X: the X intercept. Similarly, the Y-intercept M/P_Y reflects the amount of Y the household could consume if they spent all their income on Y.

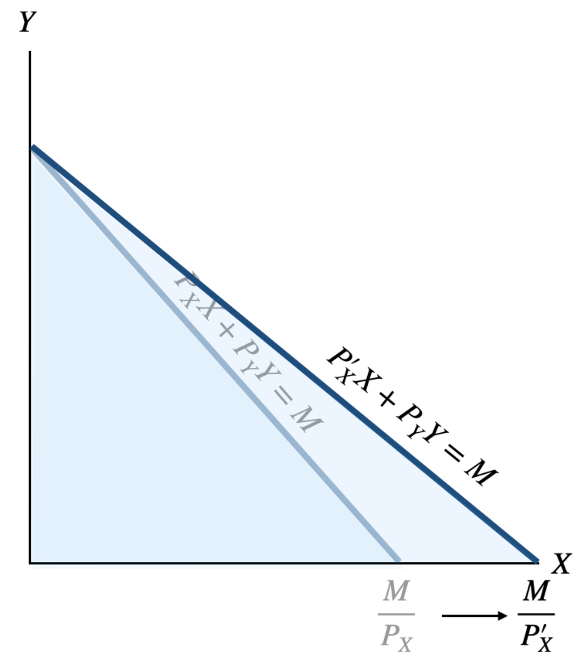


Figure 2.2: The change in a consumer's budget constraint after a price change

If the price of X falls from P_X to P'_X ($P'_X < P_X$), then they can consume more of *both* goods! If they consumed the same bundle, they would have income left over, since they spent fewer dollars on X. This extra income can be used to finance the purchase of both goods. To understand this, examine Figure 2.2. Because the price is lower, the X intercept shifts out and the slope of the line flattens! A flatter slope means the opportunity cost of X fell: to consume one more unit of X, the consumer needs to forgo fewer units of Y.

2.3 Indifference curves

A consumer with M units of income can theoretically consume any bundle in the shaded triangle, but in practice they will consume on the budget constraint. They wouldn't waste any income M. Which bundle will they pick?

To understand which bundle they will pick, we first must define *indifference curves*. Consider two bundles of goods, (X_1, Y_1) and (X_2, Y_2). A consumer is *indifferent* to consuming either bundle if $U(X_1, Y_1) =$

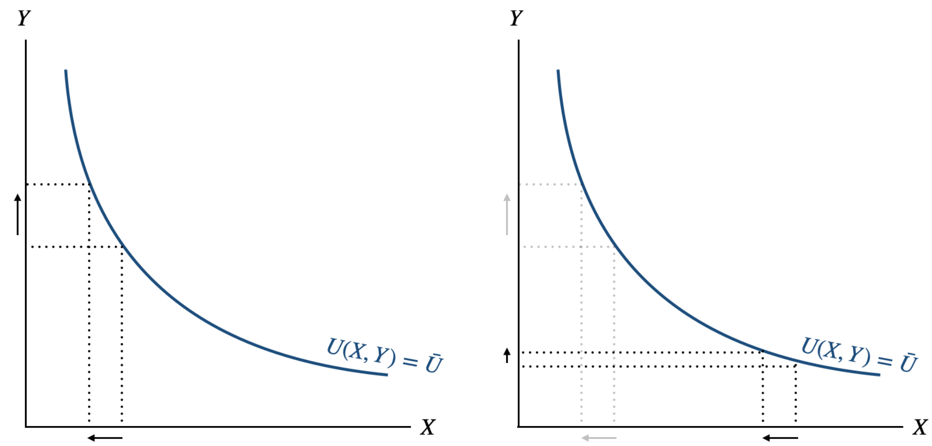


Figure 2.3: Diminishing marginal utility along an indifference curve.

$U(X_2, Y_2)$.

An indifference curve is a set of bundles $(X_1, Y_1), (X_2, Y_2), ...$ such that a consumer is indifferent to them. Suppose $U(X_1, Y_1)$ delivers $\bar{U}$ utils of utility. We can define the set of goods to which the consumer is indifferent implicitly: as any (X, Y) combination so that $U(X, Y) = \bar{U}$. This is an indifference curve.

There are two key properties of indifference curves:

1. They are concave towards the origin
2. They move outwards as utility increases.

Indifference curves must have the property that they are *concave towards the origin*. This means that they bend inwards. This reflects a key property of consumer preferences: the law of *diminishing marginal utility*. This means that the marginal utility (the utility you get from consuming one more unit of something) must be falling as you consume more and more of it. Once you've had one slice of cake, the second isn't as good. Figure 2.3 shows this graphically. In the left panel, the consumer has more Y than X. If they were to give up one unit of X and move along the horizontal arrow left, to stay equally as happy, they would need a whole lot of Y. In the right panel, things change: once they already have a lot of X, they're happy to give up a unit of it to gain just a little Y.

Indifference curves must also move out (away from the origin) as utility increases. Consider a bundle,

(X_1, Y_1) and consider another, $(X_1 + 3, Y_2 + 3)$. The second bundle is strictly better than the first! The household will always prefer,

$$U(X_1 + 3, Y_2 + 3) > U(X_1, Y_1)$$

or any arbitrary increase in either good. This means that the indifference curve for $(X_1 + 3, Y_1 + 3)$ must lie above the indifference curve for (X_1, Y_1). This is displayed in Figure 2.4: the indifference curve for the point above (X_1, Y_1) is higher, with a higher level of utility, $\bar{U}' > \bar{U}$.

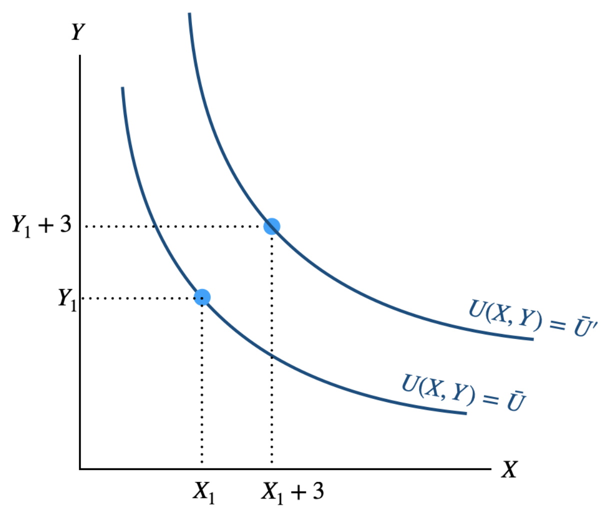


Figure 2.4: Utility rises as indifference curves move away from the origin, $\bar{U}' > \bar{U}$.

2.4 Consumer optimization

How do consumers optimize? We know already that they must choose a bundle on the budget constraint $M = P_X \times X + P_Y \times Y$. If they didn't, they'd have leftover money to spend on goods that make them happy.

Now, let's consider possible options. Suppose they started consuming at point (X_1, Y_1), which is on their budget constraint.[1] If they gave up a unit of X, they could consumer P_X/P_Y units of Y, as discussed. How much would their utility change? They would lose the marginal utility of X in utils (call it MU_X)[2] but gain

[1]Meaning $P_X X_1 + P_Y Y_1 = M$.
[2]The utility of consuming one more unit of X, $MU_X \approx U(X + 1, Y) - U(X, Y)$. For the calculus-minded, this is literally marginal: $MU_X = \partial U/\partial x$

$P_X/P_Y \times MU_Y$ because they consumed more Y. That gives a choice rule: *if* $MU_X < P_X/P_Y \times MU_Y$, *then* consume more Y. The lefthandside is the utility *lost* from giving up X. On the righthandside, the price ratio is how many units of Y we can buy per unit X, and we multiply that by the marginal utility of Y to compare the utility gain from consuming Y instead. If the righthandside is bigger, the consumer is better off choosing to consume one less unit of X in exchange for Y. Similarly, $MU_X > P_X/P_Y \times MU_Y$ means consume more X. So a consumer can only be optimizing if there's no way for them to change their choice and be happier. At the optimum, consumers must choose so that,

$$MU_X = \frac{P_X}{P_Y} \times MU_Y \implies \frac{MU_X}{MU_Y} = \frac{P_X}{P_Y}.$$

That is, consumers optimize when the ratio of marginal utilities equals the price ratio.

Let's now derive graphically what the ratio of the marginal utilities is. I start by decomposing any change in utility. I use the symbol Δ to mean a change in a variable. Any change in utility must come from either changing the consumption of X or of Y:

$$\Delta U = MU_X \times \Delta X + MU_Y \times \Delta Y.$$

This should be a straightforward formula. If you change your consumption of X by ΔX, utility should change by MU_X per unit change in X. Same thing for Y. Utility only comes from consuming X and Y in this world.

The defining feature of indifference curves is that along the indifference curve, as we change X and Y, U does not change! Therefore, along an indifference curve,

$$0 = MU_X \times \Delta X + MU_Y \times \Delta Y \implies -\frac{MU_X}{MU_Y} = \frac{\Delta Y}{\Delta X}$$

Because $\Delta Y/\Delta X$ is the formula for slope ('rise over run'), it must be such that along the indifference curve, its slope is $-MU_X/MU_Y$.

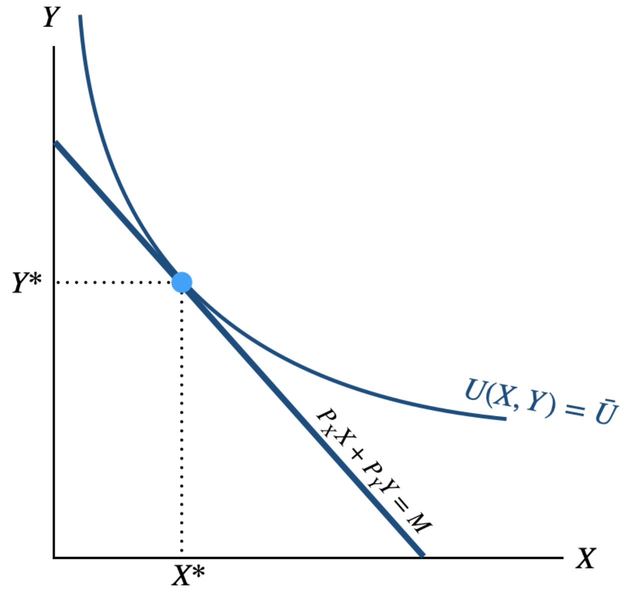


Figure 2.5: Consumer optimization. At (X^*, Y^*) the consumer maximizes their utility by setting the ratio of marginal utilities equal to the price ratio

Recall now that when consumers optimize,

$$\frac{MU_X}{MU_Y} = \frac{P_X}{P_Y}$$

i.e., the slope of the indifference curve must equal the slope of the budget constraint! This means when households optimize, there must be tangency between the budget constraint and the indifference curve.

In Figure 2.5, we see this graphically. The indifference curve moves as far away from the origin as possible until it is tangent with the budget constraint at the point (X^*, Y^*).

2.5 Putting the pieces together

A demand curve asks, what happens to quantity demanded (the optimal choice of X, X^* in Figure 2.5) as prices change. We now have all the necessary information to put this together.

First, as P_X changes to $P_X' < P_X$, the budget constraint shifts out, and the relative price of X, P_X/P_Y falls (meaning the slope of the budget constraint flattens). These are the two effects behind demand: an

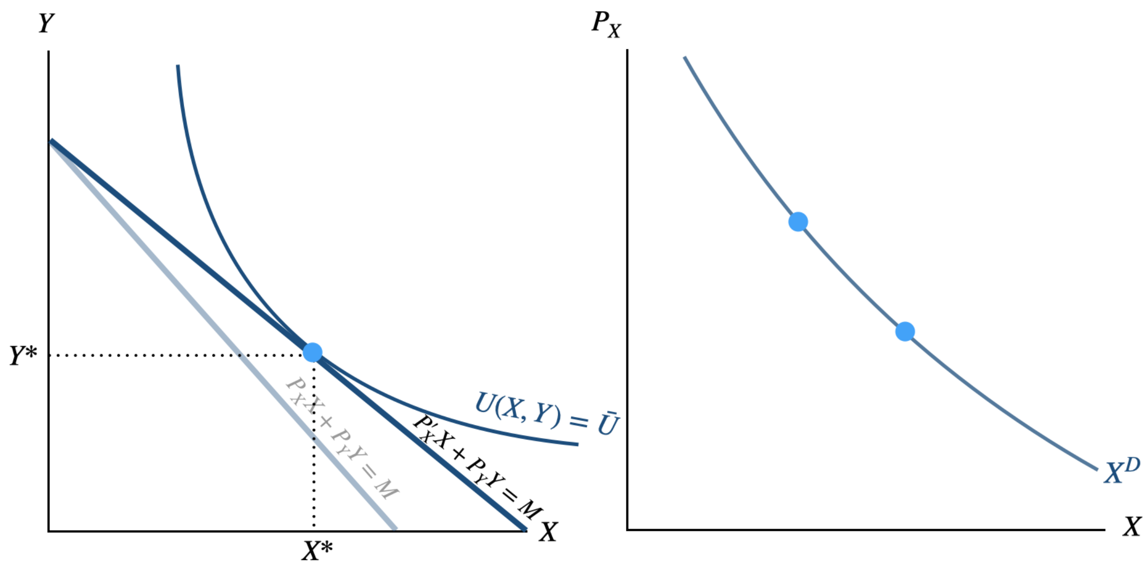


Figure 2.6: Left: Consumer optimization under a price change. Right: demand curve.

income effect (the household is richer, and wants more of everything, provided both are normal goods) and a substitution effect (X is relatively cheaper than Y, so they'll want more X (relative to Y) than what they were consuming before.

When the consumer re-optimizes, and picks a consumption bundle where the ratio of marginal utilities is equal to the price ratio, they'll end up consuming more X. This is displayed graphically in the left panel of Figure 2.6.

We just conducted an experiment. We changed prices, and examined how demand changed. We saw P_X fall and X^* rise. This gives us two data points. If we did this for all prices P_X, holding the price of Y fixed, we could trace out the relationship between P_X and X^* – this is the demand curve, displayed in right panel of Figure 2.6. The demand curve slopes down: this is called the law of demand. As prices fall, quantity demanded increases.

2.6 The price elasticity of demand

How the demand curve slopes tells us the *price elasticity of demand*: for a given percent change in price, what is the percent change in demand. The price elasticity of demand is small, the demand curve is steep, and

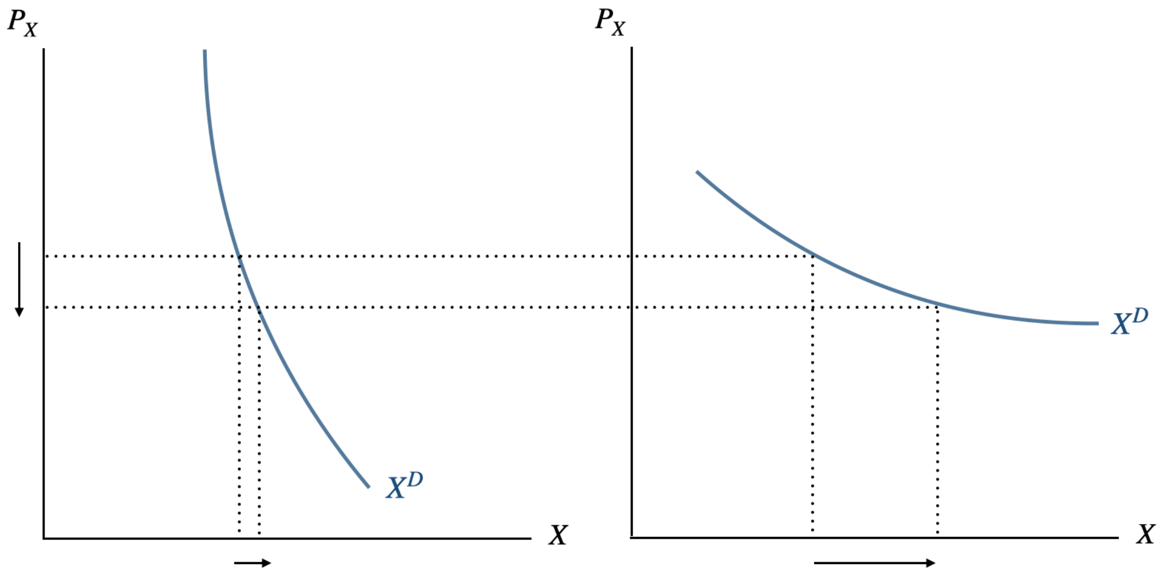


Figure 2.7: Left: Inelastic demand. Right: Elastic demand

changes in price generate less than proportional changes in demand. Goods for which demand is inelastic include necessities or habit-forming goods like cigarettes. When demand is elastic, the demand curve is very flat: small changes in price create large changes in demand.

This is displayed in Figure 2.7. The same change in price is illustrated for two different demand curves. In the left panel, the change in price causes a small change in demand. In the right panel, the same change in price causes a large change in quantity of X demanded. We say that the second demand curve is more (price-)elastic than the first.

Lecture 3 – Trade in endowments

I begin our study of gains from trade in an economist's laboratory: an economic model.

I will study an imaginary, highly stylized world to learn about some key ideas in trade before setting up a basic Ricardian economy and studying comparative advantage in the next series of lectures.

Our main goal here is to illustrate the obvious: that there are gains to trade. Moreover, we will see that the price system can help allocate resources to their best use and maximize welfare for all participants of the economy. This is Adam Smith's 'Invisible Hand' theorem: I will show in the following lecture how equilibrium prices will deliver the same allocation of resources as a benevolent social planner. Along the way, we will learn about *Pareto efficiency*.

The main tool emphasized in this chapter is the Edgeworth box, a graphical tool that allows us to illustrate gains from trade with very little math, but it does require you to turn your thinking a bit upside-down!

3.1 An endowment economy

I begin by setting up an endowment economy, before studying gains from trade and the operation of prices and markets in allocating scarce goods.

In the endowment economy, there are two goods: X and Y. For the entire economy, these are in fixed supply: there are $\bar{X}$ units of X and $\bar{Y}$ units of Y. No more can be created – there is no technology for production.

In the economy, there are also two countries, country 1 and country 2. They have indifference curves U_1 and U_2. Each country is endowed with a certain amount of goods: country 1 starts the day with an endowment (X_1, Y_1) and country 2 is endowed with (X_2, Y_2). In total, their endowments compose the entire amount of X and Y in the economy: $X_1 + X_2 = \bar{X}$ and $Y_1 + Y_2 = \bar{Y}$.

They both could simply consume their endowments and receive $U_1(X_1, Y_1)$ and $U_2(X_2, Y_2)$ utils, respectively. This is their utility under *autarky* – without trade. The question we ask, as trade economists is, can

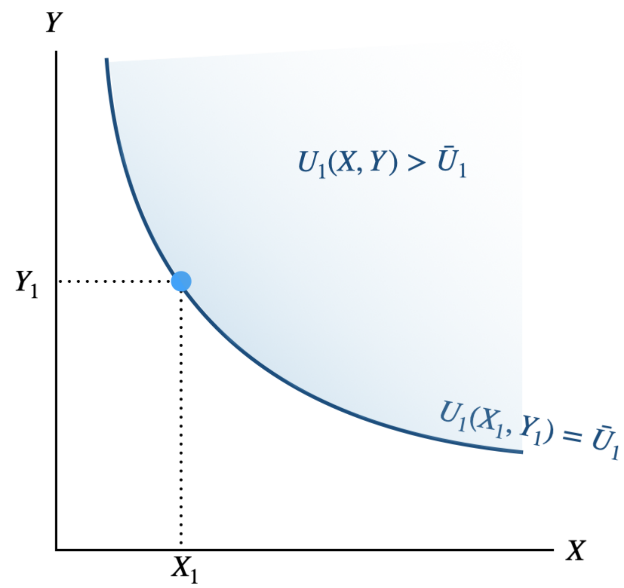


Figure 3.1: Country 1's endowment.

they trade and do better?

In Figure 3.1 we see country 1's endowment. An indifference curve associated with the level of utility $\bar{U}_1$ goes through the endowment point: every point along this curve is a bundle that would make country 1 as well-off as their endowment. In contrast, every bundle in the shaded region above the curve is associated with a level of utility higher than $\bar{U}_1$: these are bundles that make country 1 better off. The curve U_1 is the sum total of the indifference curves for all individuals in that country, and represents aggregate welfare.

Country 2 is in a similar situation. Their aggregate indifference curve goes through their endowment, and there's a set of bundles that make them better off, too.

A *Pareto improvement* would be a way to reallocate goods X and Y between both countries such that they are *both* better off. If an improvement exists, are there prices under which both parties exchange Xs for Ys to achieve that improvement? Moreover, can we reach a situation in which there are no more Pareto improvements available – a *Pareto efficient* allocation?

To answer these questions, I construct the Edgeworth box, which glues together the version of Figure 3.1 for both countries 1 and 2.

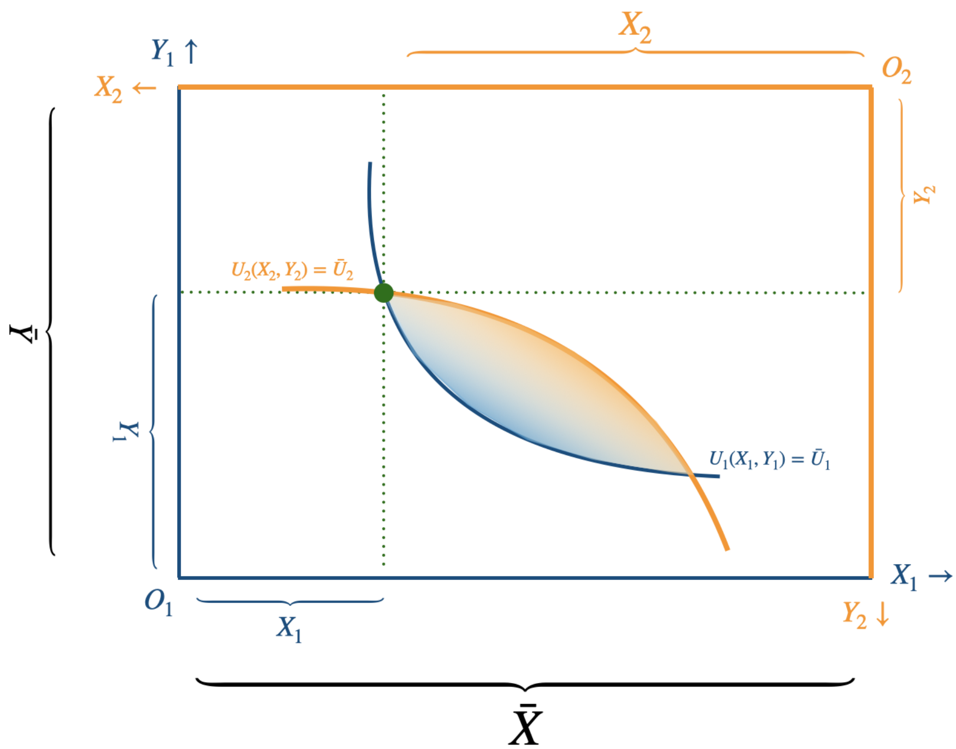


Figure 3.2: The Edgeworth box

The Edgeworth box is available in Figure 3.2. The horizontal lengths of the Edgeworth box represent the total amount of X in the economy: their length is $\bar{X}$. The vertical lengths of the box are as tall as $\bar{Y}$. The bottom-left and top-right corners are the origins of the graph, from country 1 and country 2's perspectives, respectively.

The endowment point is labeled in green inside the box. It represents *both* country 1 and country 2's endowments. Because $X_1 + X_2 = \bar{X}$, any point along the X-axis from country 1's perspective represents also represents $X_2 = \bar{X} - X_1$, country 2's amount of X. The same is true for the Y-axis.

Both country 1 and country 2's indifference curves go through the endowment point. country 1 can improve their situation by consuming any point above their indifference curve. The same is true for country 2 – you just need to turn your head upside down!

The shaded region (called the lens) represents bundles that are Pareto improvements. These are splits of the total endowment between the two countries that make them both better off.

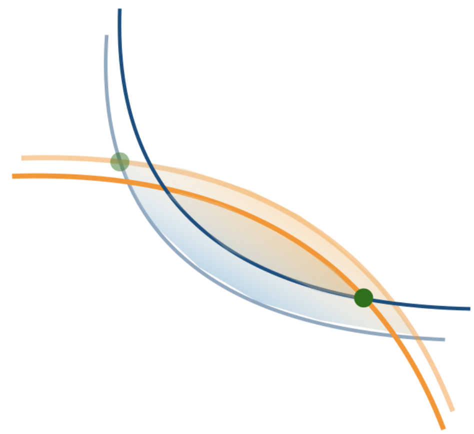


Figure 3.3: A Pareto improvement that is not Pareto efficient.

However, not every point in the lens is Pareto efficient. Consider Figure 3.3. Relative to the initial endowment point (faded) the new allocation constitutes a Pareto improvement, but there remain other Pareto improvements relative to the new allocation.

Graphically, it must be that the set of Pareto efficient allocations must be splits of the endowment that leave no lens left. This occurs when there is tangency between U_1 and U_2. What does this mean economically?

When there is tangency between U_1 and U_2, it means that,

$$\frac{MU_X^1}{MU_Y^1} = \frac{MU_X^2}{MU_Y^2}$$

where the numbered superscripts indicate countries. If it were anything else, then like a consumer optimizing, we can improve welfare by reallocating goods.

$$\frac{MU_X^1}{MU_Y^1} > \frac{MU_X^2}{MU_Y^2} \rightarrow \text{give country 1 some of 2's } Y \text{ in exchange for } X$$

likewise,

$$\frac{MU_X^1}{MU_Y^1} < \frac{MU_X^2}{MU_Y^2} \rightarrow \text{give country 1 some of 2's } X \text{ in exchange for } Y$$

However, there isn't just one point that satisfies tangency. There's a whole bunch! The set of (X, Y) points along which U_1 and U_2 are tangent is called *the contract curve*. Because every point on the contract curve

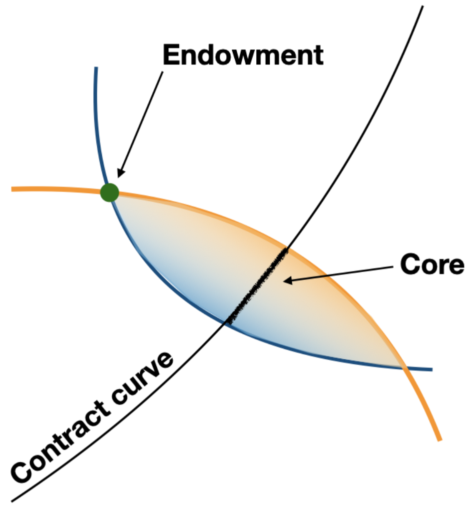


Figure 3.4: The core of an endowment economy. Every point along the contract curve is Pareto efficient. The portion of the contract curve that is a Pareto improvement over the endowment is the core.

is Pareto efficient, we can also call the set of (X, Y) along the contract curve the *Pareto frontier*. The subset of the contract curve that is in the lens of Pareto improvements is called *the core*. The core is illustrated in Figure 3.4

Given the endowment, a benevolent social planner (who can decide who gets what) would want the economy to operate along the core. This would leave no Pareto improvements on the table, and would be a welfare-improving situation over the endowment. Where, exactly, on along the core the social planner would pick depends on how much they value country 1's consumption over country 2's. This is what economic planning is – having agents (the state, a corporation, or a collection of countries) choosing the allocation of resources, instead of letting resource allocation occur in a decentralized manner through market operations.

We will show in this economy that the price system will allocate resources in a Pareto efficient manner. That is, if agents can trade freely under equilibrium prices, then the resulting allocation will be in the core of this economy.

I study the forces that equilibrate these prices and determine the resulting allocation in the next section.

3.2 An endowment economy with prices

How would a market allocate resources in this economy?

First, I define what is meant by a market. A market economy, in this environment, means a relative price P_X/P_Y under which trade occurs. Let's call this price the *terms-of-trade*. It is an exchange ratio for physical units of X with Y. I assume agents can trade freely at these prices. How trade occurs – whether in a marketplace, at ports, or over a computer – is outside of the model.

Under nominal prices P_X and P_Y, we can define aggregate nominal income based on the endowments of each country. For country 1, aggregate nominal income is,

$$\underbrace{M_1}_{\text{nominal income}} = \underbrace{P_X X_1 + P_Y Y_1}_{\text{the product of prices and the endowment}},$$

and is defined similarly for country 2. Thus, countries' budget constraints are,

$$P_X X + P_Y Y \leq M_1 = P_X X_1 + P_Y Y_1$$

The lefthandside represents any combination of X and Y, while the righthandside is fixed by the endowment. For a given terms-of-trade, countries' budget constraints can be represented in the Edgeworth box as a line that goes through the endowment point and has slope $-P_X/P_Y$.

Suppose trade began at some arbitrary price, $\tilde{P}_X/\tilde{P}_Y$. These terms of trade define income for both countries, and so their desired consumption would be the point where their indifference curves are tangent to their budget constraints, as depicted in Figure 3.5.

Equilibrium prices clear markets. This means that if $\tilde{P}_X/\tilde{P}_Y$ is an equilibrium price, then the amount of X and Y demanded at these prices would equal the amount supplied in the economy, $\bar{X}$ and $\bar{Y}$.

Figure 3.5 demonstrates that these prices cannot be equilibrium prices. For good X, country 1's demand $\tilde{X}_1$ and country 2's demand $\tilde{X}_1$ do not sum up to the total amount of X in the economy: there is a surplus of X. Similarly, at this price, too much Y is demanded! The amount of Y demanded by both countries exceeds the

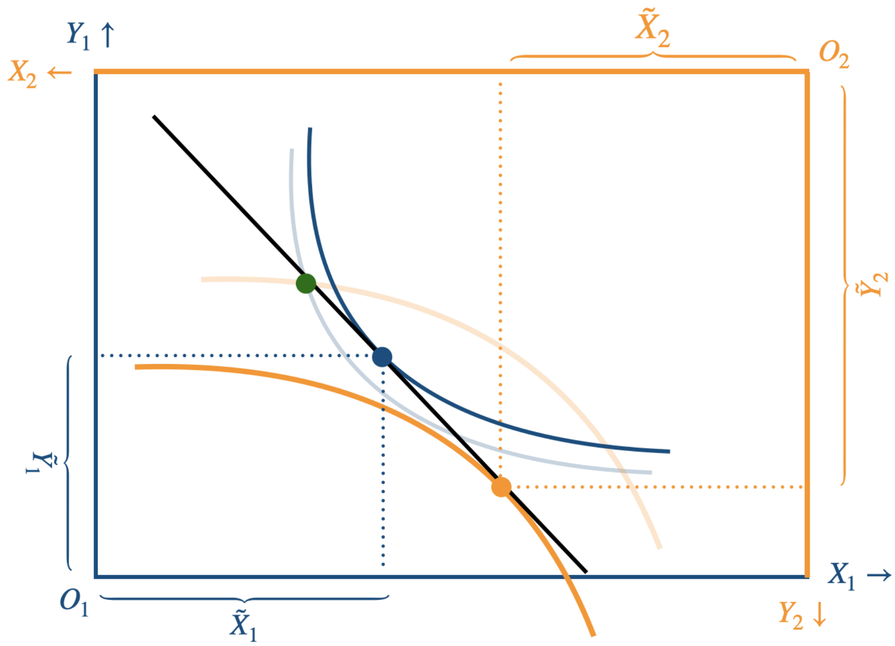


Figure 3.5: Shortages in the Edgeworth box – Prices fail to clear the market. There is a shortage of Y and a surplus of X

amount available: $\tilde{Y}_1 + \tilde{Y}_2 > \bar{Y}$: there is a shortage of Y.

What's going on? The terms-of-trade are too high: the relative price of X is too high, causing a surplus. The flip side of this is that the relative price of Y (P_Y/P_X) is too low: there is too much Y demanded and there is a shortage.

Because we can now move freely between indifference curves and demand curves, we can reconceptualize the scenario in the Edgeworth box using traditional supply and demand figures. I do this in Figure 3.6. In the endowment economy, there is no production, so the total supply of X and Y is fixed at $\bar{X}$ and $\bar{Y}$ – vertical, inelastic supply curves. Each country's indifference curve can be converted to a demand curved and then added together to create world demand.

The market clearing price must be lower than $\tilde{P}_X/\tilde{P}_Y$. By the same logic, the relative price of Y must be higher! A lower relative price of X would incentivize consumers to consume more X; likewise, a higher relative price of Y would help alleviate the shortage.

How do we know the terms-of-trade that clear the X market also clear the Y market? The geometry of

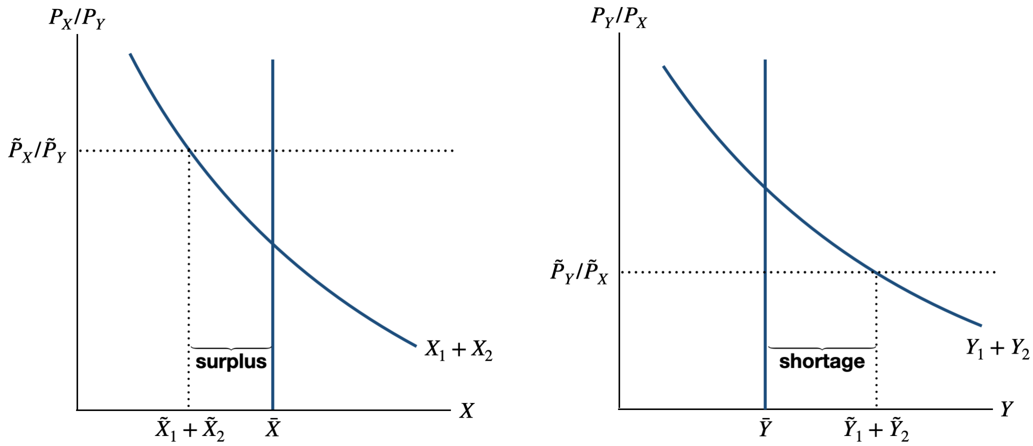


Figure 3.6: The world market for X and Y. A supply-and-demand rendering of a surplus of X and a shortage of Y. Note that the vertical axis in the right panel is inverted.

the Edgeworth box makes this clear. Both consumption bundles are always along the budget constraint, the black line with slope $-P_X/P_Y$ that goes through the endowment. Either there are two points along the line, consistent with a surplus or shortage, or both bundles overlap, meaning the same price clears both markets. If a price cleared the X market but not the Y market, or vice versa, we would require the budget constraint to either be completely vertical or completely horizontal. This only happens if one of the nominal prices is zero – a scenario we rule out.

3.3 What have we learned?

We began studying two countries in an endowment economy in autarky. Both received the utility of consuming their endowment, but as was evident from the Edgeworth box, there were gains to trade.

Absent a market, an all-knowing social planner would want to reallocate endowments to any point along the contract curve. If the social planner was restricted to find a Pareto improvement, they would pick among any Pareto efficient allocation in the core.

Under free trade and the existence of a market, where consumers in the two countries could buy and sell and world prices, there existed a terms-of-trade that both made sure the amount of X and Y demanded globally equalled global supply and picked an allocation in the core, too. Free trade and the market system picked an allocation that a planner with an unreasonable amount of knowledge could've picked, too.

We have learned that in an endowment economy, there are gains to trade! Moreover, we have learned how to actualize them: if agents face the same world prices (free trade) then exchange in a market economy results in a socially efficient allocation of resources.

Lecture 4 – Ricardian trade theory

In this chapter, we will study an economy with an ounce more realism than the endowment economy. We will also do away with the Edgeworth box as an analytical tool, though the key ideas – gains from trade, Pareto efficiency, and the ability of prices which clear markets to achieve Pareto efficicent outcomes – will linger on.

In the chapter, we will focus on *comparative advantage* – rooted in differences in labor productivity across sectors – as explaining the pattern of trade. In a sense, we will be leaning heavily on the 'labor theory of value' from classical economic theory from Smith, Ricardo, and Marx. In later chapters, we will add more realism by adding other factors of production and studying the labor market.

For now, the economies we will study can be thought of as small island nations with one person on each deciding how to allocate his or her time to the production of different goods. These have been called 'Robinson Crusoe economies', after the famous fictional castaway who lives on a desert island.

4.1 Bananas for coconuts

I begin by considering two nations in *autarky*, meaning they produce what they consume, and there is no trade.

These Robinson Crusoe economies operate as if there's one person on each island. That person can dedicate their time to harvesting bananas or coconuts. Their productivity in either task depends local, fixed features of the landscape – the height and density of the trees, and so on.

Our economy has two countries, Home and Foreign, and labor productivity (as well as the number of workers) differs between them. At Home, the one inhabitant can use an hour of his time to produce 4 bananas or 2 coconuts. Abroad, in Foreign, coconut trees are taller but bananas are scarce. Its one inhabitant can produce per unit of time 1 bananas or 4 coconuts. The *unit labor requirements* – how many units (here, hours) of labor are needed to produce one good – are summarized in Table 4.1.

	Bananas	Coconuts
Home	1/4	1/2
Foreign	1	1/4

Table 4.1: Unit labor requirements in a two-nation Robinson-Crusoe world

The unit labor requirements can help define autarky prices on each island. At Home, the *opportunity cost* of one coconut is forgoing 2 bananas, but in Foreign, the cost of a coconut is only one 1/4th of a banana. As a unit of labor produces relatively more coconuts than bananas in Foreign, compared to Home; coconuts are relatively 'cheaper' there, or, conversely, bananas are relatively more expensive.

At home, autarky prices are these opportunity costs. Without trade, consumers at home are willing to trade two bananas for one coconut: $P_C/P_B = 2$. In Foreign, $P_C^*/P_B^* = 1/4$, where I use * to indicate Foreign.

The economy's production possibilities frontier has a slope given by these autarky prices, and intercepts depend on the total amount of labor available in each economy. I will denote the size of each economy with $\bar{L}$ and $\bar{L}^*$.

Let's suppose that each nation has one person, a true Robinson Crusoe economy, so that $\bar{L} = \bar{L}^* = 1$.

4.1.1 Drawing the Ricardian Production Possibilities Frontier

Let's draw the production possibilities frontier (PPF) for each nation. The PPF captures the economy's aggregate resource constraint – that is, their national budget constraint. What constrains production of bananas and coconuts on our island is the amount of labor available, and its productivity in producing goods.

Consider Home. If all labor was allocated to the production of bananas, we would have 4 bananas (recall: 1/4 of a unit of labor makes one banana, so $\bar{L}/(1/4) = 4$). Likewise, if all labor was allocated to the production of coconuts, we would have 2 coconuts. In Foreign, they can produce at most 1 banana, but they can make 4 coconuts.

In Home, as discussed above, giving up a small amount of labor in the production of bananas loses four times that in bananas but gains twice that in coconuts. This doesn't matter how many bananas or coconuts are already in production – the trade-off is always the same. This means that the *slope* of the PPF, which

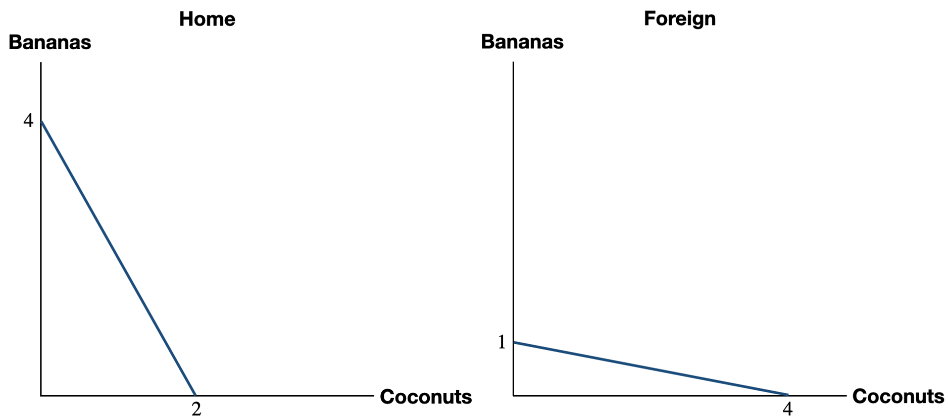


Figure 4.1: The PPFs of the Robinson Crusoe economies

represents the trade-off between bananas and coconuts is always the same: it must be a straight line. Its slope is -2, using the formula for slope $= \frac{\Delta\text{rise}}{\Delta\text{run}}$. A -4 change in 'rise' (bananas) creates a $+2$ change in 'run' (coconuts). Likewise, the slope of the Foreign PPF is $-1/4$. Figure 4.1 illustrates the two PPFs for the Robinson Crusoe world.

4.1.2 Consumption in autarky

You should have a pretty good idea of what will happen next. The resident of each island will want to consume so that their indifference curve is tangent to their budget constraint. They will consume a bundle that is both on the PPF and is such that the marginal rate of substitution between bananas and coconuts is equal to the opportunity cost of bananas in terms of coconuts – the autarky price ratio. Like with utility maximization, autarky prices in each economy are the slope of the PPF.

Utility maximization in autarky is displayed in Figure 4.2. Without trade, at Home the bundle (B, C) is consumed, while Foreign consumes (B^*, C^*). Both countries have the same preferences with quite steep indifference curves – indicating that they would happily forgo a lot of banana for one more coconut: a preference for coconuts.

As in the Edgeworth box, there ought to be gains from trade. The question is can we find world prices that clear markets and find a Pareto efficient allocation?

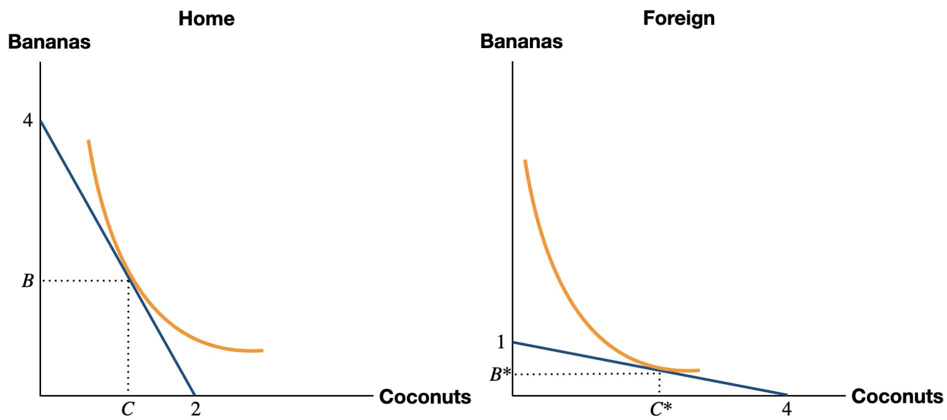


Figure 4.2: The PPFs of the Robinson Crusoe economies

4.1.3 Deriving the world supply curve

In the hope of finding equilibrium prices, let's try and describe the world market, as before in the Edgeworth box. However, now production decisions depend on prices: the world supply curve is not given by the endowment, but on the decision of producers on how to allocate time (or, equivalently, labor/man-hours) between the two productive activities.

Let's begin by constructing the world supply curve for coconuts. I will do this by slowly increasing the relative price of coconuts, and asking what the quantity supplied in the world would be.

First, suppose the relative price of coconuts in the world market P_C^W/P_B^W was very small, near zero. The inhabitants of both Robinson Crusoe economies would want to then produce bananas and trade then for coconuts in world markets – each bananas can be transformed into more coconuts in the world market than if each nation was in autarky. How would they do this using prices? They would produce bananas, sell them for P_B^W dollars each in the world market, and then use that income to buy coconuts at P_C^W.

Residents in both Home and Foreign would want to do this up until the terms-of-trade $P_C^W/P_B^W = 1/4$. At that price, residents in Foreign are *indifferent* to buying and selling in domestic and global markets – the price ratio is the same! There's no arbitrage to take advantage of by using world instead of domestic markets. In fact, for any terms-of-trade above 1/4, coconuts, not bananas, fetch a better relative price in world markets than in autarky for Foreign. At a price of, say, $1/3 > 1/4$, each coconut produced in Foreign can be exchanged for more bananas in the world market than at home.

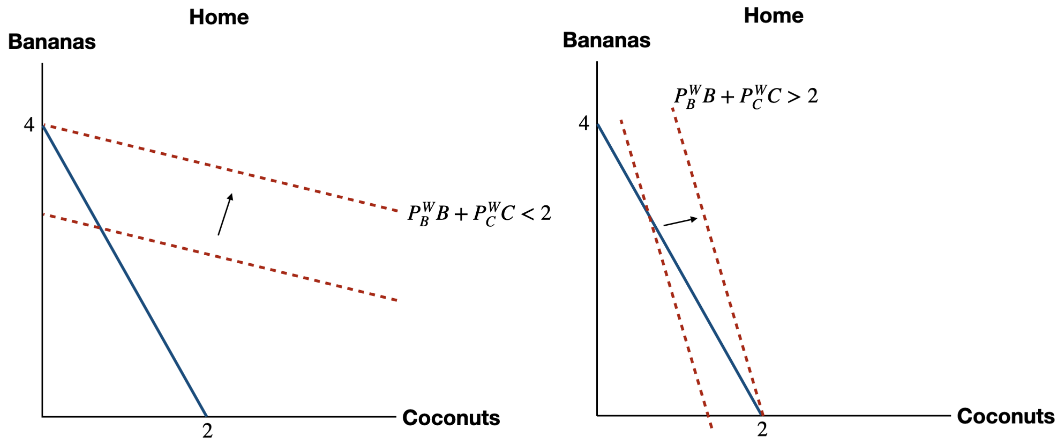


Figure 4.3: Ricardian profit maximization in two scenarios. The red dashed line represents producer revenue under world prices.

Similarly, once the terms-of-trade exceeds 2, residents of Home would also rather switch to producing coconuts. Total world supply for both goods at different terms of trade is made explicit in Table 4.2.

Terms-of-trade	Home	Foreign	Total Bananas	Total Coconuts
$0 < P_C^W/P_B^W < 1/4$	Bananas	Bananas	5	0
$1/4 < P_C^W/P_B^W < 2$	Bananas	Coconuts	4	4
$2 < P_C^W/P_B^W$	Coconuts	Coconuts	0	6

Table 4.2: World supply for both goods at different terms of trade

These supply decisions can be displayed graphically, too. When terms of trade are P_C^W/P_B^W, the goal of producers is to maximize profit by choosing how much B and C to produce.

$$\max_{B,C} \underbrace{P_C^W C + P_B^W B}_{\text{income}} \quad \text{subject to} \quad \underbrace{P_B B + P_C C = \bar{L}}_{\text{autarky PPF}}$$

Note the equation for the autarky PPF. To derive this equation, note that $P_B B =$ total amount of labor spent producing B, and $P_C C =$ the total amount of labor spent on producing C. This the equation for the autarky PPF simply states the resource constraint: all labor must be spent producing either B or C.

Ricardian profit maximization is displayed in Figure 4.3. Geometrically, it amounts to pushing away from the origin the income under world prices line (in the figure, dashed in red) as far as possible while remaining on the PPF.

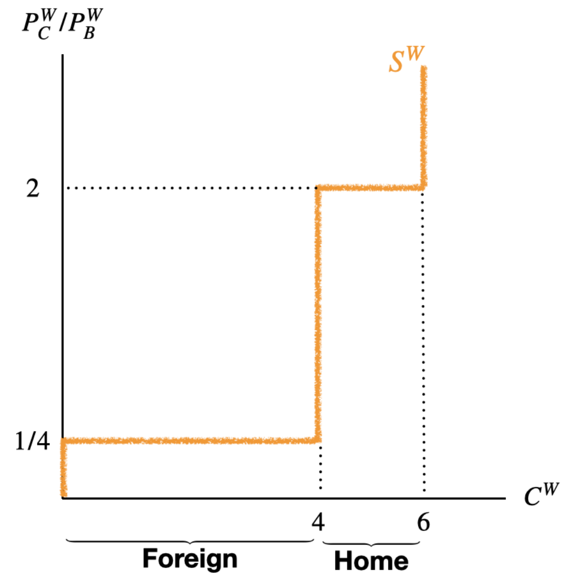


Figure 4.4: The world supply for coconuts. Foreign produces all coconuts until the price is high enough to incentivize Home to join in.

As P_C^W/P_B^W increases, the red dashed line gets steeper, until when $P_C^W/P_B^W > 2$, it is steeper than the PPF and the profit-maximizing point is the bottom only-coconuts corner of the PPF. The world supply curve for coconuts is displayed in Figure 4.4.

4.1.4 Demand and market clearing in the Ricardian economy

So far, we have only developed the global supply curve in the Ricardian economy. What remains is to find a demand curve and market-clearing prices. We need to find terms-of-trade such that there are no shortages or surpluses of either bananas or coconuts in this economy.

The level of demand is determined by income. When there is trade, income depends on world, not autarky prices. Income under world prices given production choices (B, C), $P_C^W C + P_B^W B$ is often called the "consumption possibilities frontier" (CPF). In autarky, the consumption and production possibilities frontiers are the same: you can only eat what you make. Under trade, production and consumption are divorced!

The household budget constraint (CPF) is no longer equal to the PPF. Utility maximization subject to the CPF determines demand, and profit maximization subject to the PPF determines supply.

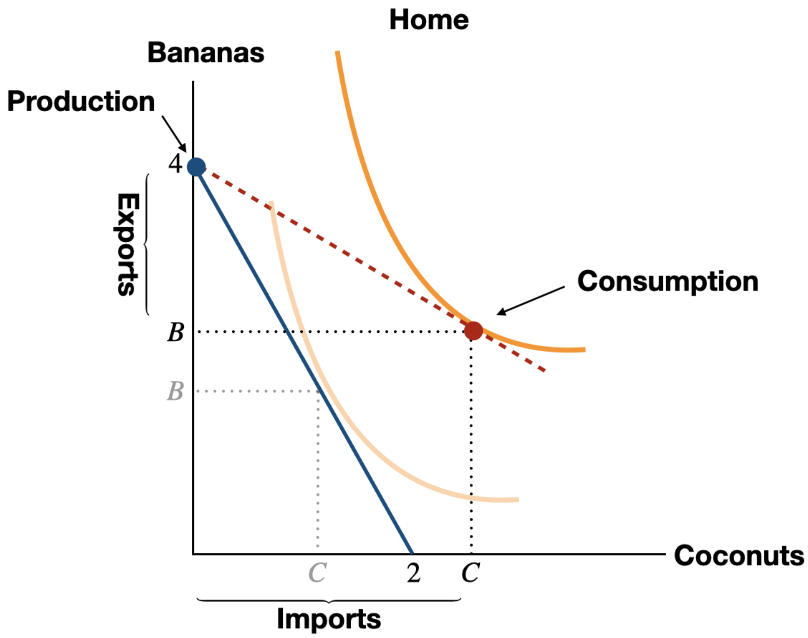


Figure 4.5: Production and consumption when world prices are 'flat.' The transparent indifference curve reflects consumption and production in autarky.

As before, utility maximization requires the marginal rate of substitution to equal the price ratio. Graphically, this looks like Figure 4.5 displays what happens when utility and profits are maximized under world prices.

First, it is evident that when trading under world prices, utility is higher: there are gains to trade! In the Figure, the old autarky equilibrium is made transparent: you can see the new indifference curve is higher, and the consumer (at these prices) is consuming more of both goods! Because coconut is cheaper in world markets, they consume more of it (a substitution effect). By the same logic, they are richer producing only bananas, since they fetch a higher price in world markets than at home, and with that increased wealth they also eat more bananas (an income effect outweighing a substitution effect).

Second, consumption and production are divorced. This requires some goods to be purchased in world markets, or imported. For Home, all coconuts imported at these prices are financed by the sale of bananas in world markets.

Market clearing in this economy means that the total amount of bananas and coconuts consumed equals the total amount supplied. Another way of saying this is that imports = exports for both bananas and coconuts.

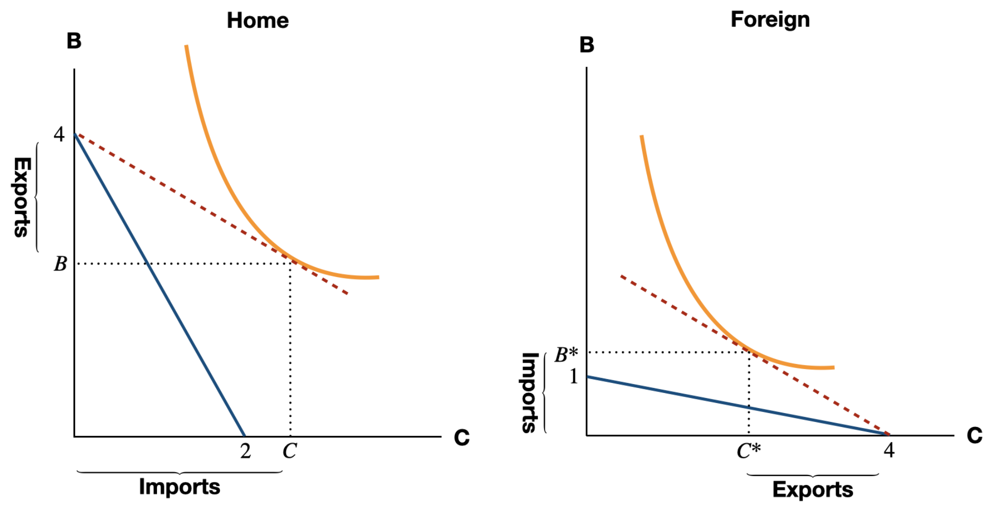


Figure 4.6: Market clearing importing and exporting in a Ricardian economy

Using Table 4.2, it should be evident that if the terms-of-trade P_C^W/P_B^W are too low, both countries produce only bananas, and both would want to export bananas to finance coconut consumption. However, no one is producing coconuts! So there is a shortage of coconuts and a surplus of bananas.

Likewise, if $P_C^W/P_B^W > 2$, both countries will want to produce only coconuts, but consume some of both goods. These prices create a surplus of coconuts and a shortage of bananas.

The only prices that can work are $1/4 < P_C^W/P_B^W < 2$. At these prices, both coconuts and bananas are produced, and so it will be possible to clear markets. Exactly what prices depend on the shape of the demand curve.

When imports equal exports, the situation with both countries looks like Figure 4.6. Equilibrium prices have to be flatter than the steeper autarky prices but steeper than the flatter autarky prices – only in this scenario will both goods be produced.

4.2 Ricardo's example: wine for cloth

I'll now follow Ricardo's famous example from his 1817 book, *On the Principles of Political Economy and Taxation*.

In this section, we'll see another example of all the Ricardian machinery in action. Moreover, this is an example where comparative and absolute advantage do not align, but the comparative advantage still forms the basis for trade.

For Ricardo, his two nations were England and Portugal, and they traded wine and cloth. The manhours of labor required to produce either a barrel of wine or a ream of cloth are listed in Table 4.3.

	Cloth	Wine
England	100	120
Portugal	90	80

Table 4.3: Labor requirements in Ricardo's wine and cloth economy.

Here, Portugal is just plain better in producing both: they can produce a unit of wine or cloth with fewer man-hours than England can. Both goods are 'absolutely' cheaper there; that is, they have an absolute advantage in producing both goods. Why should there still be gains from trade in this case?

Suppose Portugal and England have the same population. Suppose moreover that $\bar{L} = \bar{L}^* = 3600$. Then, the maximum amount of each good that each country can produce is given by the Table 4.4.

	Cloth	Wine
England	36	30
Portugal	40	45

Table 4.4: Maximum amount producible in each country.

Table 4.4 helps us define the intercepts of each country's PPF when doing Ricardian graphical analysis. In Figure 4.7. The slope of the PPF reflects autarky prices, P_W/P_C where now W means wine and C means cloth. Autarky prices make clear that in exchange for cloth, wine is cheaper in Portugal than England. This also means that in exchange for wine, cloth is more expensive in Portugal. If the Portuguese could exchange wine for cloth at English prices, a barrel of wine could be exchanged in the market for $120/100 = 6/5$ reams of cloth, while in the autarky economy, the same barrel of wine could only be transformed into $80/90 = 8/9$ths of a ream of cloth. Even though Portugal has an absolute advantage in the production of both goods, the possibility of trade creates an opportunity for arbitrage: to take advantage of lower prices elsewhere! Autarky prices reflect *comparative* not absolute advantage.

As discussed in the previuous section, if both countries can trade freely, in equilibrium prices must incentivize

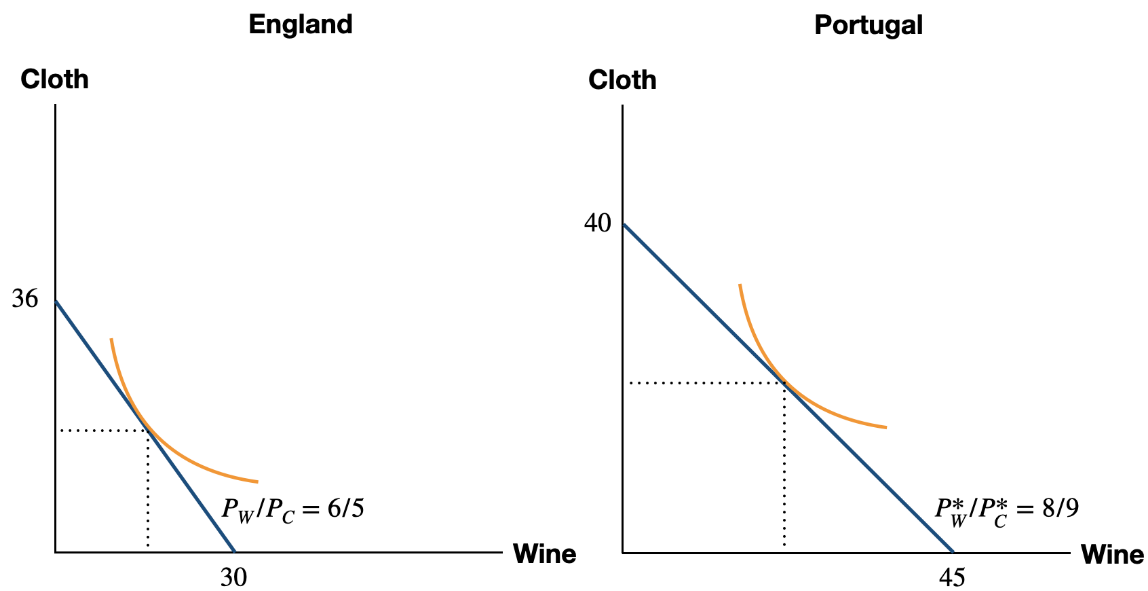


Figure 4.7: PPFs and autarky indifference curves in Ricardo's famous example

complete specialization and the production of both goods. This requires equilibrium terms-of-trade to be between $8/9$ and $6/5$: under these prices England would specialize in cloth and Portugal would specialize in wine. Countries specialize in the goods in which they have a comparative advantage!

Moreover, prices need to make sure markets clear and that imports = exports. The exact price that clears both markets depends on demand, but it does exist (and importantly there is only one market clearing price). It is determined by the intersection of world supply and demand, as shown in Figure 4.8.

The market clearing price, determined by the intersection of world supply and demand balances imports and exports. The resulting equilibrium is shown in Figure 4.9. Make sure you can identify the imports and exports in this diagram.

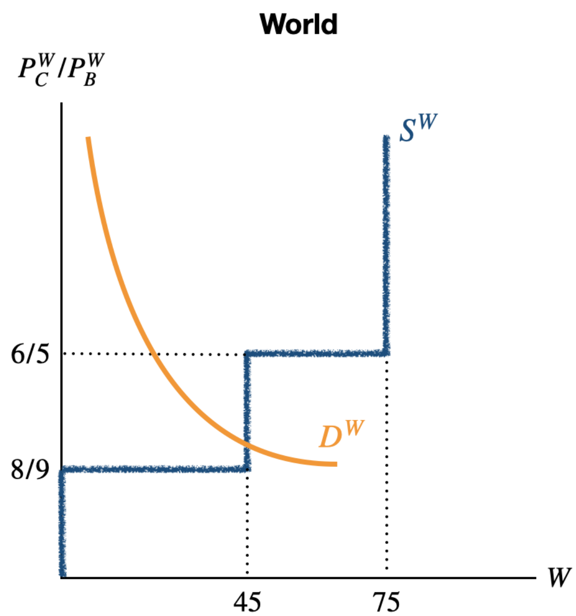


Figure 4.8: The world market in Ricardo's example.

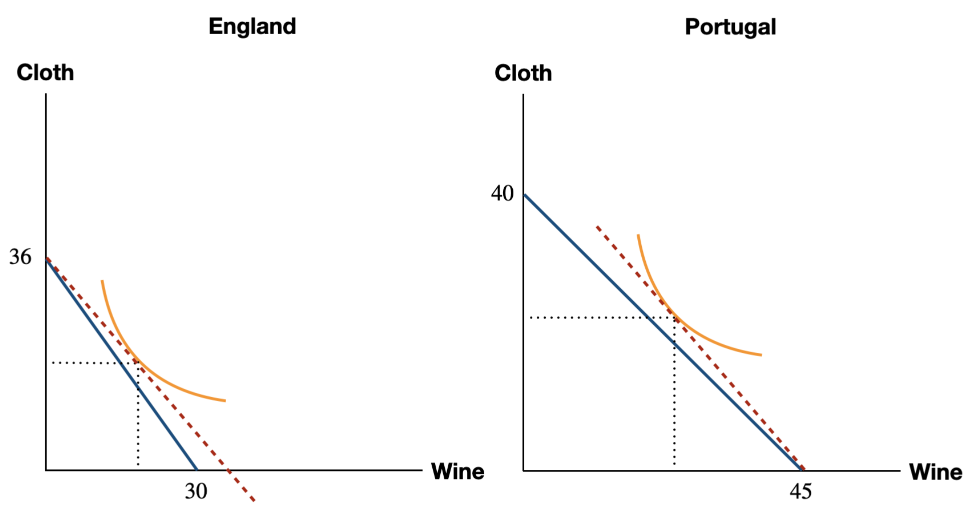


Figure 4.9: The free trade equilibrium in Ricardo's famous example

Lecture 5 – The specific factors model

In the specific factors model, we build on the Ricardian model to add more realism and policy relevance.

The Ricardian model is useful to illustrate comparative advantage as explaining the pattern of trade, and the ability for trade to generate Pareto improvements even when one country has an absolute advantage in the production of all goods. However, the Ricardian model is not useful for policy analysis or thinking about more nuanced issues in trade. Moreover, it makes a lot of predictions that just don't hold up to data. For example, it predicts *complete specialization*, something we see no country doing! Second, it abstracts from major questions in trade, like *who* benefits from trade, and how does trade affect other domestic markets. In the U.S., there is a lot of news coverage on trade's effects on workers in export sectors, like manufacturing. The Ricardian model fails to provide an understanding of these markets.

In this lecture, I will add some more realistic details to the Ricardian model, and see what insights and predictions remain robust to this realism. In particular, I will add (1) Specific factors: elements like land and capital that matter in production and (2) Decreasing-returns-to-scale production technology. I hope to convince you that these two details buy us a lot of realism and broaden our toolkit.

In return for these two details, we will be gain: (1) Winners and losers from trade (2) Incomplete specialization (3) A labor market (4) The ability to use standard supply-and-demand diagrams to describe the trading equilibrium!

All of this will require introducing some new economic concepts, like, returns to scale, and the difference between average and marginal products.

5.1 Returns to scale and the concave PPF

In the basic Ricardian setup, we had a production function that related output (e.g., X) to labor. It looked like,

$$X = a_x L_x$$

where a_x was the productivity of labor (L_x) in the X-producing industry. Unit labor requirements were simply $1/a_x$.

This is not a very realistic production technology. It fails to capture many important things. First, labor is the only factor of production. A cursory look outside reveals that most businesses combine labor with capital, land, and intermediate inputs to produce their goods and services. I am writing this from a cafe in which labor (a barista), capital (an espresso machine), and intermediate inputs (roasted coffee beans) are combined to produce espresso, which is then combined with capital and land to produce seating. The experience of having an espresso at a table is produced in a final step where the espresso and seating are combined with a host of intermediate inputs: cushions, jazz, WiFi, etc, to produce the 'good' I'm consuming. It's a misnomer, calling it espresso, when what I'm paying for is better thought of as The Espresso Experience.

It's quite clear that as I enjoy The Espresso Experience, that another worker wouldn't increase the output of the experience by a lot. I definitely couldn't replicate the experience for a friend if one more unit of labor was available. Nor would the marginal cushion, unit of WiFi speed, or napkin be able to replicate the experience. There are obvious diminishing returns to the use of each factor in the production of The Espresso Experience. However, if I could duplicate everything and create a perfect replica of the worker, espresso machine, cushions, and so on, I could produce another Espresso Experience.

That is to say, the production technology exhibits *constant returns to scale* in aggregate. The Espresso Experience production function might look something like,

$$X = F(L, K),$$

where F is an arbitrary function, L is labor used in the Espresso Experience, and $\bar{K}$ represents all the fixed capital combined with labor to create the experience. What distinguishes labor and capital is that the baristas can quit and work elsewhere if they so wanted, but the capital, $\bar{K}$ is already installed. It is specific to the industry and immobile across industries: the unit of capital called 'an espresso machine' is not the same as a furnance for heating and bending autoglass. In other words, the espresso machine won't walk across the street for a higher wage. In the long-run, capital is mobile, and we will use the Heckscher-Ohlin model in the next lecture to study capital mobility.

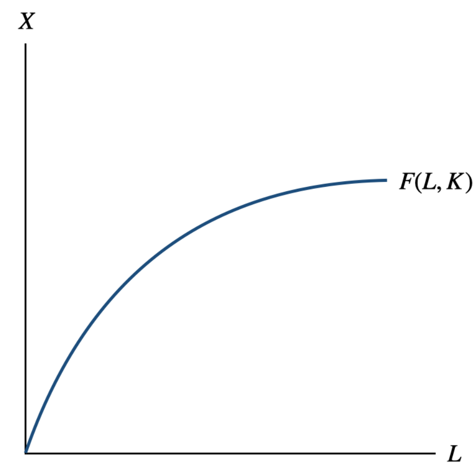


Figure 5.1: The production function as a function of L, holding $\bar{K}$ fixed in the specific factors model

What it means to display constant returns to scale is that if you doubled all the inputs, output would double, too. Formally, if you have any positive number $\lambda > 0$, returns to scale means,

$$\lambda X = F(\lambda L, \lambda K).$$

Scaling up L and $\bar{K}$ by λ scales up output, X, by the exact same amount. A consequence of returns to scale means that there are not constant returns to any given input. In particular, if $\bar{K}$ is held fixed, scaling up L would not result in scaling up output by the same amount:

$$\lambda X \geq F(\lambda L, K).$$

Moreover, we will assume these returns are diminishing on the margin: with a fixed input, the each additional worker contributes less and less to production. This means production functions look like Figure 5.1. The Figure shows how each each additional unit of L means that X increases by less and less: diminishing marginal returns to labor.

5.1.1 Different industries, different specific factors

We will now suppose each country has two industries, X and Y. These could be, for example, manufacturing and agriculture. What matters is that each industry has different specific factors.

In the case of manufacturing and agriculture, our immobile, sector-specific factors might be capital $\bar{K}$ and land $\bar{T}$. Manufacturing requires capital, like machines. Agriculture combines land and labor to make cereals, fruits, and so on. Agricultural land can't "work" in manufacturing. It is immobile. Likewise, capital used in the manufacturing production process – like an autoglass bending furnace – cannot be used in agriculture. The profits in each industry accrue to the owners of capital and land – landlords earn profit from renting their land out to farmers, who finance those rental payments by selling agricultural goods. Capitalists earn rents from the use of their capital in manufacturing. Unlike capital $\bar{K}$ and land $\bar{T}$, workers of which there are $\bar{L}$, are mobile between the two industries.

The two goods have the following production functions:

$$X = F(L, K), \quad Y = G(L, T),$$

where F and G have constant returns to labor and specific factors.

5.1.2 Deriving the specific factors model PPF

How much X and Y the economy can produce depends on the total amount of labor in the economy. If all labor was allocated to X production, the economy could produce $F(\bar{L}, K)$ units of X. Similarly, if every worker worked in the Y-producing industry, the economy could produce $G(\bar{L}, T)$ units of Y.

However, because each production function has diminishing marginal returns to labor, the PPF is no longer a straight line. In fact, it is bent away from the origin.

To see this, imagine the following thought experiment: suppose the economy starts with all labor in X-production. If we remove one laborer from X production to Y production, X output should only fall by a little. As there are diminishing marginal returns to labor, the last worker employed in X production didn't add

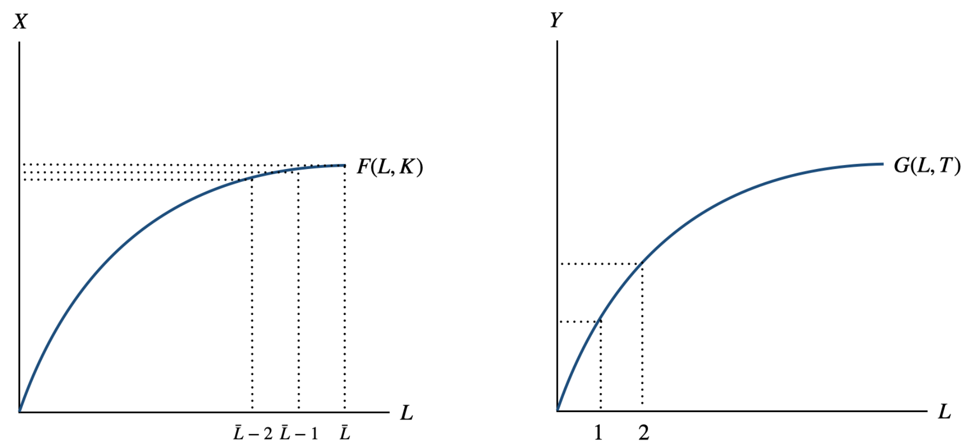


Figure 5.2: Reallocating labor from X to Y

much, so removing that worker from the production process won't take away a lot either. However, moving that worker to the Y industry should cause a big increase in output! Because of diminishing marginal returns to labor, the first worker employed is the most productive.

If we did this again, and removed another worker from X-production, X output would fall a little more than it did before, because the 2nd to last worker employer was a little more productive than the last. Likewise, the second worker added to Y production doesn't add as much. This exercise is displayed in Figure 5.2.

As in the Ricardian model, how much X or Y we can produce is constrained by the amount of labor in the economy. Along the production possibilities frontier, every worker is allocated to either X or Y production.

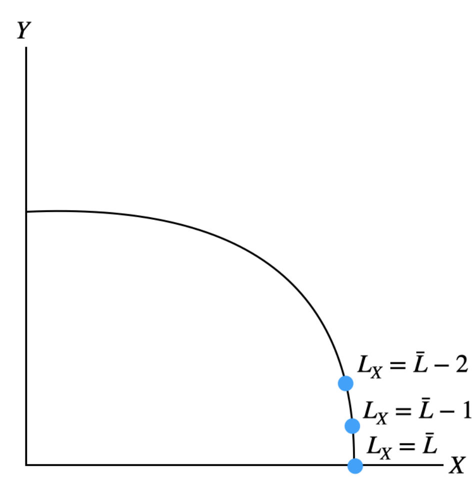


Figure 5.3: The Specific Factors Model PPF.

In Figure 5.3, I graph the Specific Factors Model PPF from this experiment of reallocating labor between the two sectors. Recall that the PPF is a graph whose axes are the quantities of the two goods. When all labor is allocated to X production, the economy operates along the X-intercept. As the first worker moves from X to Y, Y output increases a lot, and X output falls by a small amount.

This also means that the opportunity cost of X in terms of Y is falls as we decrease X. This opportunity cost is called the marginal rate of transformation, or the marginal rate of technical substitution ($MRTS$). The slope of the PPF is $-MRTS$.

5.2 The specific factors model autarky equilibrium

In autarky, because production and consumption must balance, the autarky equilibrium is described by a relative price for X in terms of Y, P_X/P_Y. In equilibrium, this price must such that the amount of X and Y producers want to produce is equal to the amount of X and Y demanded by consumers. The first is determined by profit maximization, and the latter by utility maximization.

Profits in the specific factors model are maximized when the price ratio is equal to the opportunity cost of X in terms of Y.

If producers forgo producing a unit of X, labor can reallocate towards the production of Y. How much Y goes up as X goes down depends on the $MRTS$. This is because the $MRTS$ represents the slope of the PPF, $\Delta Y/\Delta X$ along the PPF. It captures opportunity costs – how much ΔY can adjust in response to ΔX given technological constraints. Change in profit labor reallocating to produce Y instead of X is equal to,

$$\Delta\text{Profit from a change of } \Delta X = P_Y \times MRTS \times \Delta X - P_X \Delta X$$

This means that,

$$\Delta\text{Profit from a change of } \Delta X > 0 \quad \text{if} \quad MRTS > \frac{P_X}{P_Y}$$

As before, if the price ratio is less than $MRTS$, then labor ought to reallocate from X-production to Y-production. If the price ratio rise above the $MRTS$, labor ought to reallocate from the Y industry into the

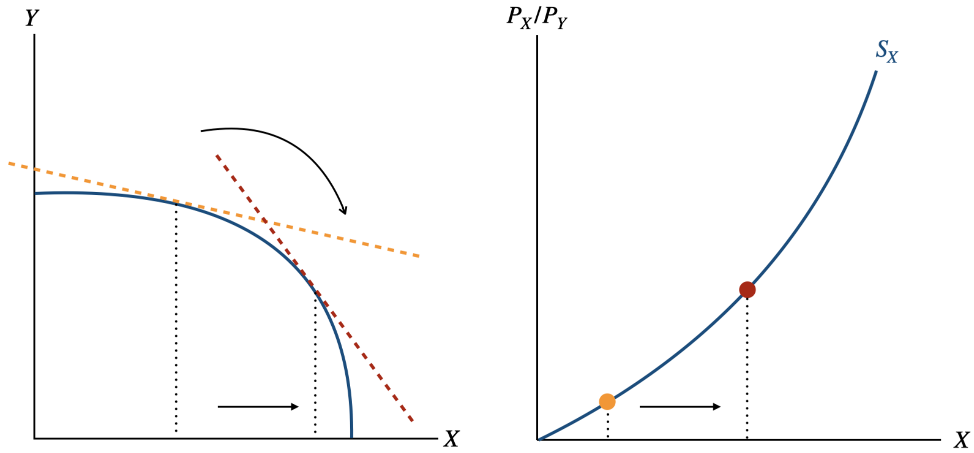


Figure 5.4: Deriving the supply curve for X in the Specific Factors model. As the relative price of X steepens, producers supply more X.

X one. Therefore, when,

$$MRTS = \frac{P_X}{P_Y},$$

profits are maximized. This is a tangency condition! As the relative price of X, P_X/P_Y becomes steeper, the profit-maximizing point on the PPF shifts rightwards. This allows us to trace out the supply curve for X, which relates its relative price to the quantity supplied, as displayed in Figure 5.4.

Consumers maximize utility when, $MRS = P_X/P_Y$, so by consequence, when $MRS = MRTS$, both profits and utility are maximized. This occurs when there is tangency between the PPF and the country's indifference curve.

Tangency between the PPF and the indifference curve – the specific factors model autarky equilibrium, is displayed in Figure 5.5.

5.3 The labor market

We've already run into a conceptual issue: the autarky equilibrium of Section 5.1.2 may be consistent with profit maximization if there was *one* producer trying to produce. This assumption was appropriate with our Ricardian Robinson Crusoe economies, but not very appropriate for the real world.

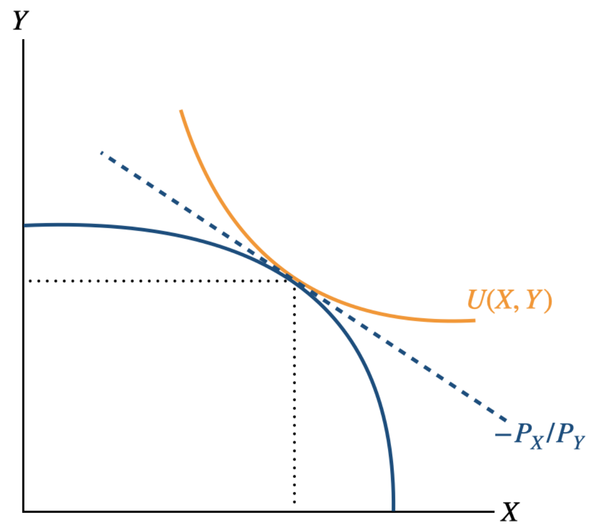


Figure 5.5: The autarky equilibrium in the specific factors model.

In fact, in the Specific Factors Model, we assume that there are two industries that compete with each other. One industry produces X, and the other produces Y. If there are two industries, do we necessarily have to end up at the same allocation as if there was only one industry?

In this section, we will show that under *perfect competition*, the economy will produce at the profit-maximizing equilibrium. This will deepen the 'Invisible Hand' theorem established in Lecture 3: the price system and perfect competition will result in the same allocation as if there was a social planner. Let's dive in.

5.3.1 The average and marginal product of labor

Each industry produces its output according to the production functions $X = F(L, K)$ and $Y = G(L, T)$. That is they combine labor with their specific factor (capital and land) to produce output.

We assume they take prices P_X and P_Y as given, as well as a new price: the cost of labor, the wage W. Their goal is to maximize profit. Consider the X industry. Their profit is,

$$\underbrace{\pi_X}_{\text{profit}} = \underbrace{P_X X}_{\text{revenue}} - \underbrace{W L_X}_{\text{costs}}.$$

They can choose how much to produce X, or, equivalently, how many workers to hire L_X, since any amount

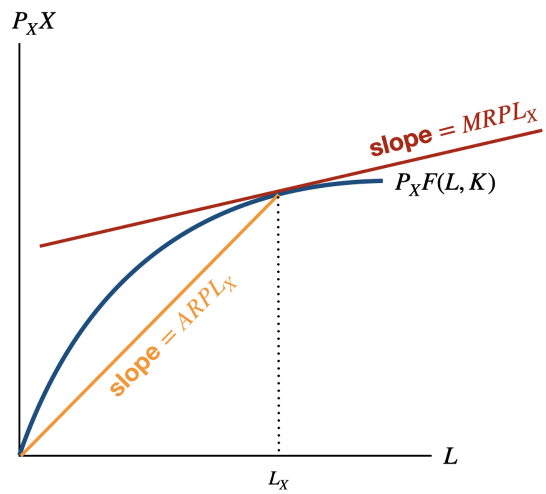


Figure 5.6: The average and marginal revenue products of labor

of X corresponds to some amount of L_X. To know how they make their profit-maximizing decision, we need to define the average and marginal product of labor.

The average (revenue) product of labor (ARPL) is the amount of revenue that a worker creates on average:[1]

$$ARPL_X = \frac{P_X F(L_X, K)}{L_X}.$$

The marginal product of labor is that amount of output created by the additional worker, $MRPL_X \approx P_X F(L_X + 1, K) - P_X F(L_X, K)$.[2] Both can be displayed graphically, in Figure 5.6.

The Figure depicts the average and marginal revenue products of labor for a quantity of labor L_X as the slopes of two lines. The line which connects the origin to the point $(L_X, P_X F(L_X, K))$ has a slope equal to,

$$\text{slope of line from origin} = \frac{\text{rise}}{\text{run}} = \frac{P_X F(L_X, K)}{L_X} = ARPL_X$$

Similarly, the line tangent to the point represents the *marginal* revenue product of labor. As F exhibits decreasing marginal returns to labor, we can see graphically the slope of the the line connecting the origin to a the point will always be *steeper* than the line tangent to that point. From that, we can conclude that $ARPL_X > MRPL_X$. This is because the marginal worker is less productive than the last, but the average

[1]To get revenue-per-worker, or the average revenue product of labor, simply multiply the average physical product (output per worker) by the price of output.

[2]As with marginal utility, for the calculus minded, this can be thought of as literally marginal: $P_X \frac{\partial F}{\partial L}$.

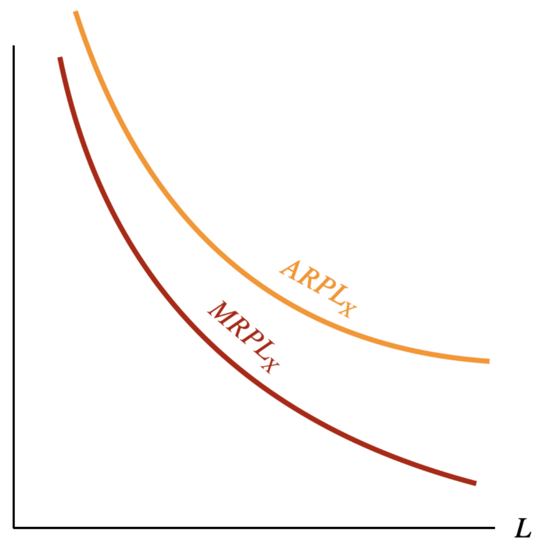


Figure 5.7: The average and marginal revenue products of labor

includes the productivity of all the workers hired before, who were more productive. We can show this graphically as well, as shown in Figure 5.7

5.3.2 Profit maximization

How does an employer maximize profit? They employ workers until $MRPL = W$.

Suppose an industry hired L_0 workers, and at L_0, $MRPL > W$. This would imply that the additional revenue generated by hiring one more worker and having them produce would generate more revenue than costs – more profit. So, to maximize profit, the industry ought to expand and hire (at least) another worker. Suppose they hired $L_1 > L_0$ workers where at L_1, $MRPL < W$. This means that the last worker they hired cost the industry more (W) than the revenue they brought in – thus, to maximize profit, they ought to contract and fire some workers. Therefore, only when $MRPL = W$ can the employer be profit-maximizing.

How much profit does the employer earn? We can express this using the average product of labor. For

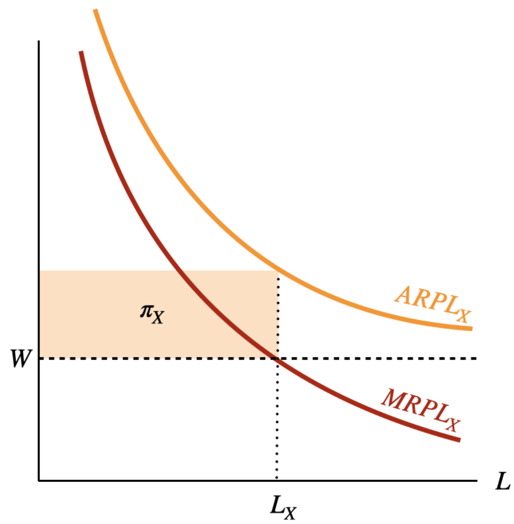


Figure 5.8: The profit maximizing choice of L is L_X. Profit at this choice is given by the area of the orange rectangle.

industry X

$$\begin{aligned}\pi_X &= P_X F(L_X, K) - W L_X \\ &= \left(\frac{P_X F(L_X, K)}{L_X} - W\right) L_X \\ &= (ARPL_X - W)\, L_X\end{aligned}$$

We can show the profit graphically, by finding a rectangle with height $ARPL_X - W$ and width L_X. Then the area of this triangle would equal profit. I show this in Figure 5.8. At wage W, the X-industry chooses labor L_X, which generates profit π_X.

5.3.3 Wage determination

In Section 5.3.2, the wage was exogenous, and we determined how much labor an industry would hire to maximize profit.

As the profit-maximizing choice for each industry is always to choose L such that $W = MRPL$, we can think of the $MRPL$ curves as industries' labor demand curves! They tell us at a given price (W) how much labor the firm would want to hire.

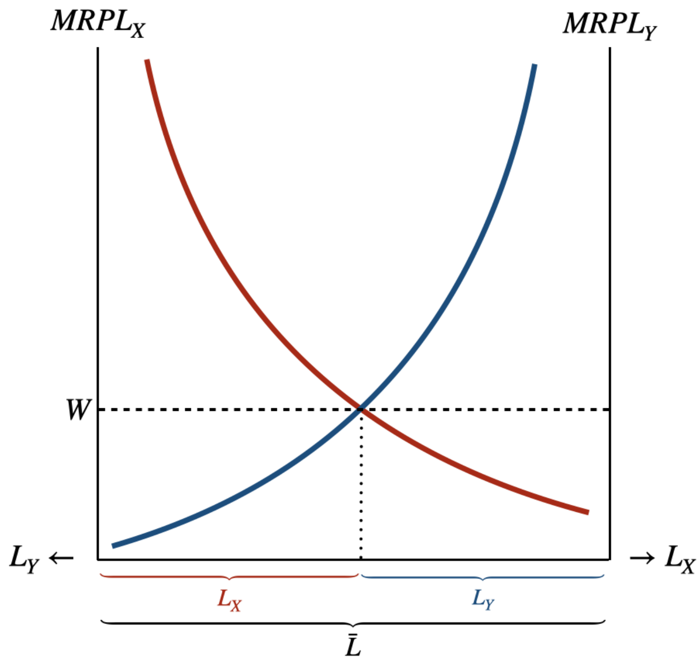


Figure 5.9: Wage determination in the specific factors model

To determine how the wage is set, we need to know how labor is supplied to each sector. Workers in the specific factors model choose which industry to work in by maximizing their utility – which is driven here by their income! They work in whatever industry offers the highest wage.

In equilibrium, both industries operate to satisfy demand for both goods, and so therefore they must offer the same wage. Labor supply in this model is *perfectly elastic* to each industry: if either industry offered a penny below the equilibrium wage, all workers would quit and want to work in the other industry. The only way to sustain both industries producing is if they offer the same wage. However, we assume aggregate labor supply is *perfectly inelastic*: $\bar{L}$ does not change in response to W – this is because there is no migration between countries in this model!

The conditions for equilibrium in the labor market are: (1) full employment (otherwise we would not be on the PPF) and (2) both industries offer the same wage. These conditions are enough for us to figure out what the equilibrium wage will be. I will do so graphically.

In Figure 5.9, I show how the equilibrium wage is determined in the specific factors market. The length of the L-axis represents the total amount of labor in the economy, $\bar{L}$. The origin on the left side represents zero labor the X-industry while the origin on the right side represents zero labor for the Y-industry. Just like the Edgeworth box, any point on the line represents a split of $\bar{L}$ into L_X and L_Y. The left vertical axis measures $MRPL_X$, and on the right, it measures $MRPL_Y$. The Y-industry simply has its L-axis flipped, and the diagram shows the labor demand curves for each industry.

The equilibrium where all workers are employed and both industries set marginal revenue products of labor equal to the wage is shown where the two labor demand curves cross in the diagram. This is how the equilibrium wage is determined.

5.3.4 Pareto efficiency in the specific factors model

We now return to our original question: is the free market allocation (determined by the labor market) the same as if there was a social planner? The answer is yes.

When both industries maximize profit, we know they equate marginal *revenue* products of labor. The marginal revenue product is the price times the marginal product of labor (MPL), which measures how much physical output does one more worker produce. So the labor market equilibrium can be described as,

$$P_X MPL_X = P_Y MPL_Y \implies \frac{P_X}{P_Y} = \frac{MPL_Y}{MPL_X}$$

This would be the same as the equilibrium point where the slope of the PPF ($-MRTS$) is equal to the price ratio.

We know two things about the PPF – one, all labor is used, so $L_X + L_Y = \bar{L}$, and that its slope is given by how Y changes as we trade it for X. Since X and Y are produced using labor, any change in them must come from changes in labor scaled by the MPL,

$$\Delta X = MPL_X \times \Delta L_X, \quad \Delta Y = MPL_Y \times \Delta L_Y$$

With some rearranging, we can see that the marginal rate of technical substitution depends on the marginal products of labor, and how labor reallocates between sectors.

$$MRTS = \frac{\Delta Y}{\Delta X} = \frac{MPL_Y}{MPL_X}\frac{\Delta L_Y}{\Delta L_X}$$

Similarly, we know that, along the PPF, $\Delta \bar{L} = 0$, so,

$$0 = \Delta L_X + \Delta L_Y \implies \frac{\Delta L_Y}{\Delta L_X} = -1.$$

Therefore, $-MRTS = MPL_Y/MPL_X$. Consequently, the following two statements are the same,

$$-MRTS = \frac{P_X}{P_Y}, \quad \text{(economy-wide profit maximization)}$$
$$\frac{MPL_Y}{MPL_X} = \frac{P_X}{P_Y}, \quad \text{(individual producer profit maximization)}$$

We have learned that we can describe the economy as if there was one producer maximizing profits, even though there are multiple producers and a labor market in the background!

5.4 The specific factors model trade equilibrium

Now we know how the Specific Factors Model operates under autarky, it's time to open up the economy to trade. We'll do so by again having two economies: Home and Foreign. We will assume Home and Foreign have access to the same technology, F and G but differ in the endowment of specific factors K, T and mobile factors, $\bar{L}$

Trade does the same in the Specific Factors Model as it does in the Ricardian model: it allows production and consumption decisions to be different: producers can export and consumers can import. World prices must be such that one nation's exports must equal another nation's imports.

However, we've lost a key ingredient: *comparative advantage*. How do we know what world prices will be and who will specialize in what? What even *is* comparative advantage in this model?

While there is no strict sense of comparative advantage, due to the concave PPF, we can loosely think of

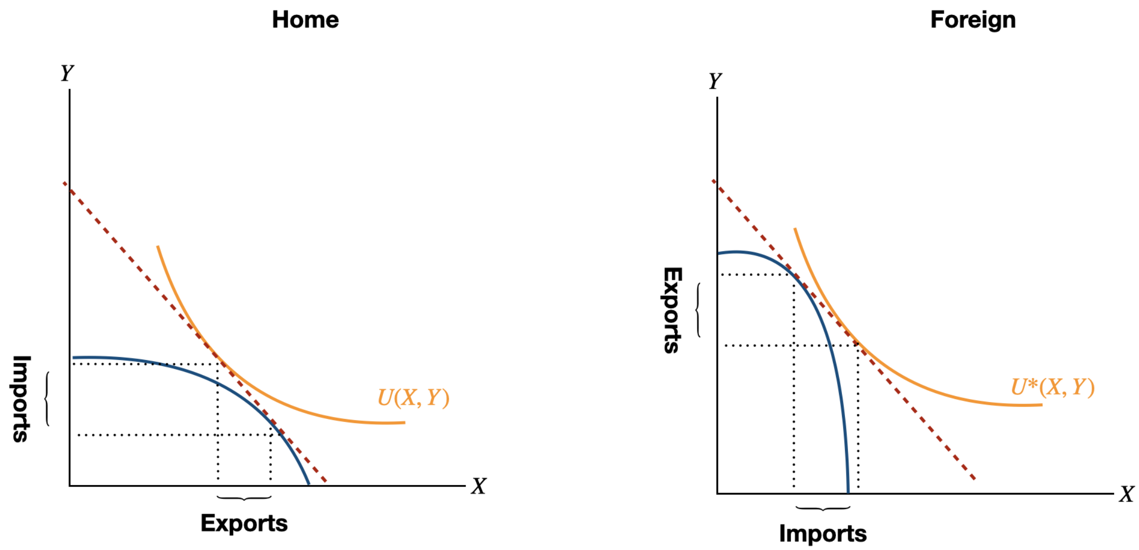


Figure 5.10: A depiction of the Specific Factors Trading Equilibrium

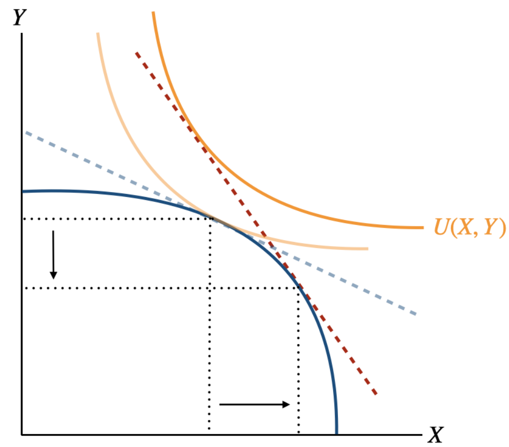
comparative advantage rooted in relative specific factor endowments. Countries will tend to export the good that uses the specific factor in which they are relatively more abundant. E.g., if $\bar{K}/\bar{T} > \bar{K}^*/\bar{T}^*$, Home's PPF will be on average flatter than Foreign's (they can produce relatively more of the X good) and they will tend to export the X good. In the lecture on Heckscher-Ohlin theory, we will make this new notion of comparative advantage more precise.

However, as the PPF in the Specific Factors Model has convexity, no price can induce *complete* specialization other than a terms-of-trade of 0 or infinity, something we rule out in practice.

In Figure 5.10, I draw an example of the Specific Factors Model trade equilibrium. Home is relatively more capital-abundant ($K/T > K^*/T^*$), and so trade incentivizes them to export the X good, while Foreign exports the Y good. There are aggregate gains from trade, because the trade allows the consumption possibilities frontier to expand.

5.5 Trade: winners and losers

In the Specific Factors Model, there are still aggregate gains to trade, as discussed in Section 5.4. However, there are now winners and losers.



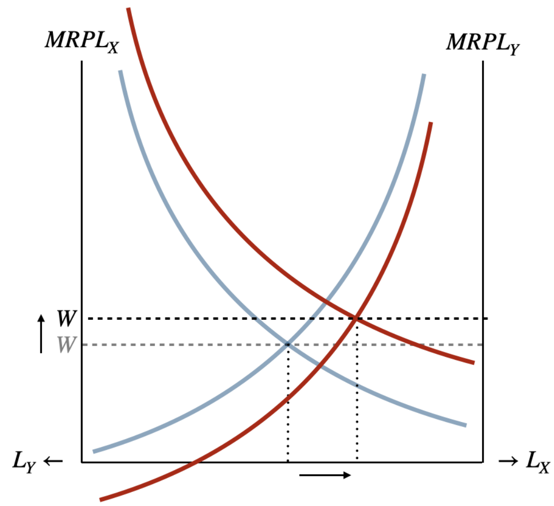


Figure 5.11: Labor market effects of trade in the Specific Factors Model. Left: moving from autarky to trade, effects on aggregates. Right: the labor market effects of the trade shock

The broad takeaway is that trade benefits the owners of the factor used in the export sector, and the owner of the factor in the import sector loses. In the model, these are capitalists and landlords. Whether trade benefits workers is ambiguous, and depends on what changes in the labor market look like. Workers can always gain if the profits earned by capitalists and landlords can be given back to them somehow. That is, if we can compensate the losers of trade with some of the earnings of the winners, trade can still constitute a Pareto improvement. However, that's not always the case in practice, and often the losers of trade are not compensated!

To see this, let's work through an example of what happens to wages and profits when we move from autarky to free trade using the graphical tools of the Specific Factors Model.

In Figure 5.11, I illustrate the effects of going from autarky to free trade in the Specific Factors Model. This is one example of what may occur in the model: I assume trade induces Home to export X and import Y, as the relative world prices are steeper than the autarky prices. This widens the consumption possibilities frontier and as a consequence aggregate welfare increases.

In the right panel, we can see who wins and loses from trade. In this example, I assume at world prices compared to autarky prices, the nominal price of X, P_X rises while the nominal price of Y, P_Y, falls (making P_X/P_Y steeper). This means each worker in X is more productive in revenue terms, shifting up the

X-industry $MRPL_X$ curve (i.e., X's labor demand curve), while Y's labor demand curve falls. Labor reallocates from the Y industry to the X industry – this must happen to produce more X and less Y – and the nominal wage rises. Consequently, nominal profits in the X industry rise and nominal profits in the Y industry fall.

In the Specific Factors Model, profits accrue to the owners of the specific factors. The X-industry uses capital, so in this example, capitalists and workers 'win' from opening to trade, while landlords 'lose.'

5.6 Technological change

Another feature of the Specific Factors model is its ability to analyze (biased) technological change. What I mean by 'technological change' is an improvement or increase in the amount of capital or land available to Home or Foreign. These technological changes are 'biased' because they only affect one industry.

Typically, countries do not get large changes in their endowments of capital or land. A country could benefit from changes in the global economy and receive a large amount of foreign direct investment (FDI) allowing it to invest more in installed capital, leading to an increase in $\bar{K}$. Countries like Singapore use land reclamation by infilling water with sand to increase $\bar{T}$.[3] However, what is more common is that the quality of capital or land changes, which can be thought of as an increase in the endowment in the Specific Factors Model. New farming technology like fertilizer can make land more productive – it's as if a country has more land. New technology or better management practices can put capital to new use, making it more productive – again, as if the stock $\bar{K}$ increased.

5.6.1 Export biased growth

Let's examine what happens with an increase in $\bar{K}$ when a country is already trading (i.e., an open economy) and is relatively small (i.e., doesn't affect the terms-of-trade it faces).

In Figure 5.12, I analyze what happens when there is export biased growth in a small open economy.

[3] Land can increase by bloodier means too – see Russia's current invasion of Ukraine.

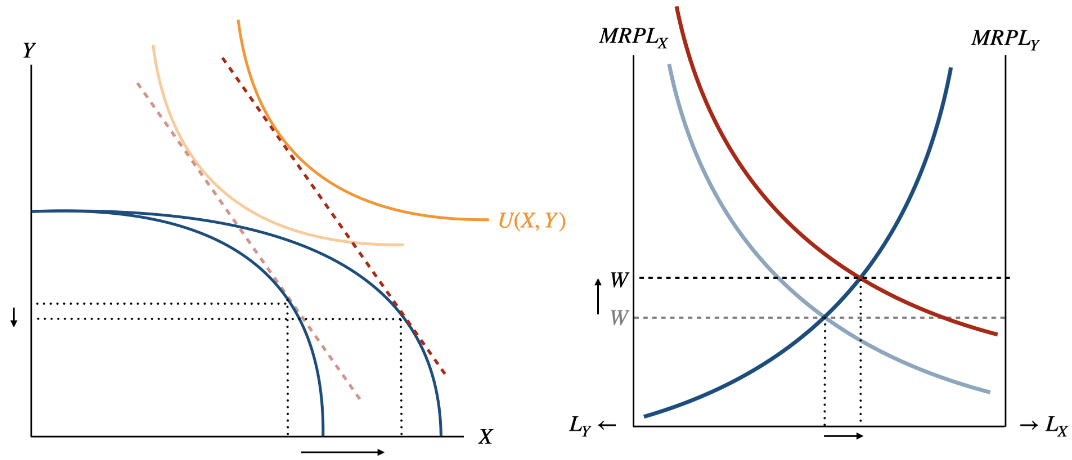


Figure 5.12: Export biased growth in a small open economy in the Specific Factors Model. Left: a change in $\bar{K}$ shifts the PPF out. Right: Labor productivity increases in the X-sector, driving up the equilibrium wage and reallocating workers from Y to X.

As capital is not used to produce Y, the Y-intercept doesn't change. However, more capital means the potential for the economy to produce more X, so the X-intercept and the entire PPF shifts out.

The terms-of-trade remain the same, but are tangent at a new point along the PPF: a point where Y-output is lower, as labor reallocates to produce more X. As the PPF shifts out, so does the consumption possibilities frontier, and aggregate welfare expands.

As there is more capital, each worker in the X-industry is more productive, and so the entire $MRPL_X$ schedule shifts up. To draw in more workers from the Y-industry, the equilibrium wage rises.

A key feature of this example is that the economy I analyzed was that the open economy was *small*. What I mean by this is that the change in X supplied on the world market did not affect the world price for X. However, this need not be the case. If a country is large, then changes in its export industry can affect the world price of that export good.

If a country is large, the change in X will shift the global supply curve of X to the right, driving down the price of X! The world market is flooded with more X. As we move down the demand curve, the marginal willingness-to-pay must also fall, because of diminishing marginal utility. In Figure 5.13, I show the world market for X. The rightward shift in supply drives down the price of X.

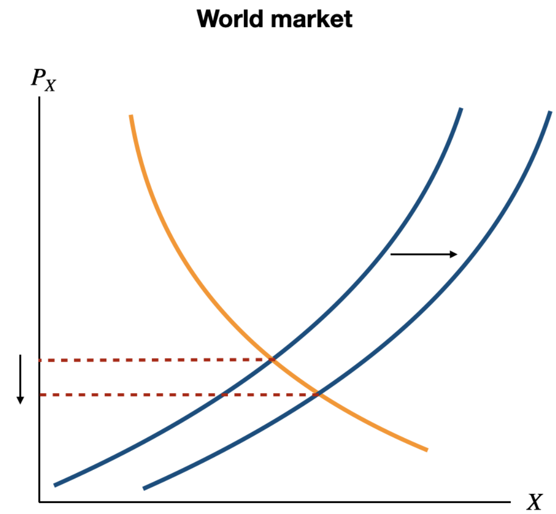


Figure 5.13: The world market when Home experiences an increase in $\bar{K}$.

As the price of X falls, the terms of trade change, too. Figure 5.12 no longer appropriately describes what happens at Home when they experience export-biased growth.

Figure 5.14 shows the effect of export biased growth in a large open economy, relative to a small one. The price adjustment for X due to changes in the world market partially flattens the terms-of-trade. The economy does not export as much X as it would have if the terms-of-trade did not adjust, but there are still gains to trade. As the nominal price of X fell, the $MRPL_X$ curve, relative to a small open economy, shifts down, tempering the wage increase and labor adjustment.

These effects are surprising, and lead to a few questions. For example, can the wage *fall* because of export biased growth? No – for the wage to fall from export biased growth, the terms of trade would need to adjust so much that Home exports less X than before the trade shock, but growth can't cause the supply curve to shift inwards.

Can export biased growth *lower* welfare? Yes – in theory it can, which is what we explore in the next section.

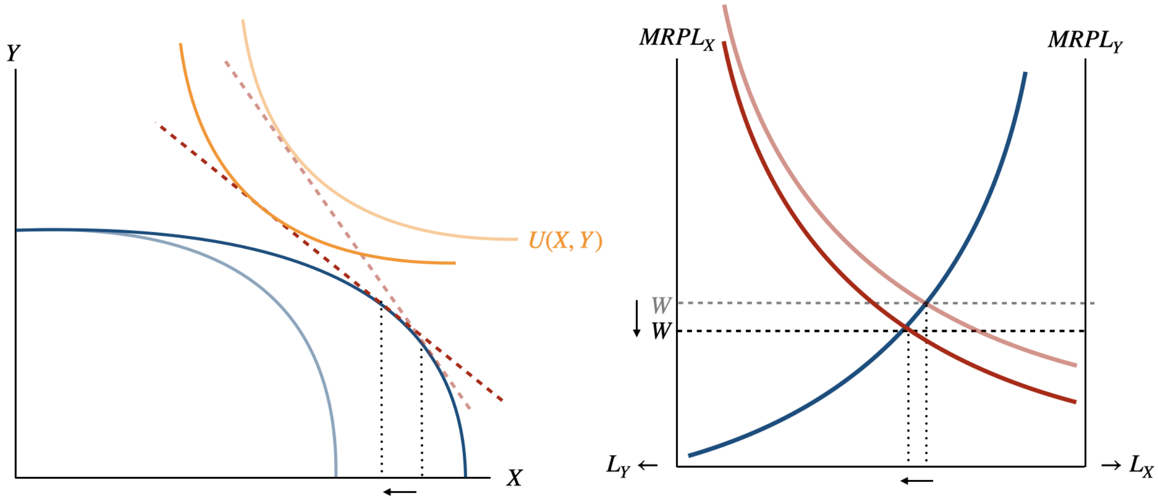


Figure 5.14: Export-biased growth in a large open economy – the terms of trade adjust, partially undoing the effect of the productivity shock.

5.6.2 Immiserizing growth

Famously, at age 24, Jagdish Bhagwati showed that it was possible for export biased growth to lower welfare, in Bhagwati (1958). Bhagwati would go on to advise the World Trade Organization (WTO) and the United Nations.

Three ingredients are needed: a country must exert significant effects on price in global markets (so it must be large in the market for its export good), global demand must be inelastic for the export good (so that increases in supply lead to large terms-of-trade adjustments) and the Home country needs to have inelastic demand for the import good (i.e., a country needs to be very reliant on imports).

Under these conditions, export-biased growth deteriorates the terms-of-trade so much that welfare decreases, as illustrated in Figure 5.15.

However, immiserizing growth is so far more of a theoretical possibility than something regularly observed in the data.

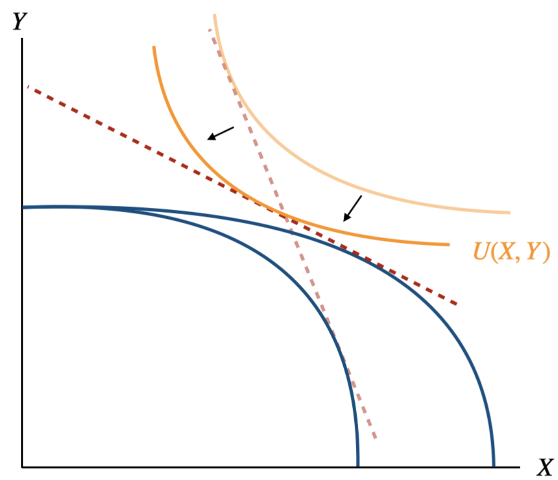


Figure 5.15: Immiserizing growth in the Specific Factors Model

Lecture 6 – Heckscher-Ohlin theory

The Heckscher-Ohlin model was, for a long time, the centerpiece of trade theory, and will serve as the end of our adventure into neoclassical theory.

The Heckscher-Ohlin (HO) takes the Specific Factors Model, and relaxes a few assumptions. In particular, the HO model allows capital and labor to be mobile across both industries. Instead of industries being described by their specific factors, they are described by their *factor intensity*: how much they relatively use one factor or the other.

Why the Heckscher-Ohlin model was so beloved for so long is that it produced quite a few 'theorems' that allow us to generalize our understanding of how trade causes specialization and affects different industries.

Another reason the HO model was so beloved is the miracle it performs. To get at what I mean, let me use Paul Krugman's fable about Samuelson's Angel (named after Paul Samuelson, the most important economist in neoclassical theory). We start from a world where there's capital and labor used two produce two goods consumed by a bunch of consumers. Then, one day, Samuelson's Angel appears and divides up the world into different 'countries:' borders within which capital and labor are mobile, but across which they are not! Is this a disaster? It turns out, in the HO model, that with free trade, the division of the world into countries doesn't hurt global welfare. It turns out that free trade can 'reproduce the integrated equilibrium.' What this means is that with free trade, the borders turn out to be irrelevant. Free trade in goods is a perfect substitute for free trade in factors.

In this section, we'll study how industries make production decisions when both factors are mobile, how industries 'trade' factors (labor and capital) within a country, using the Edgeworth box. Then we'll study the trading equilibrium in the HO model, and zoom through the key results.

6.1 HO model setup, expenditure minimization

The HO model is sometimes called the $2 \times 2 \times 2$ model, because there are two countries, two goods, and two factors: labor and capital. Because there are two factors, we also think of the economy as populated by two types of agents: workers and capitalists. One 'owns' their labor, and the other owns capital. Labor or capital can 'win' or 'lose' from trade.

Labor and capital are free to move across industries within each country, but can't move across countries. Each country has an endowment of both factors $\bar{L}, \bar{K}$ and $\bar{L}^*, \bar{K}^*$, which they both allocate to the production of either good. Countries differ in their endowments of labor and capital, and use the same technology to produce the same goods.

In the HO model, there are two industries that produce two goods, X and Y. They both use labor L and capital K to produce the goods, and have constant-returns technology, just as in the Specific Factors Model,

$$X = F(L, K), \quad Y = G(L, K).$$

The only difference now is that both industries can use K – it is not fixed – and we've done away with land.

Both industries pick the amount of labor and capital they want to maximize profits. One way to maximize profits is to minimize expenditure, or costs. The cost of using L units of labor to produce X or Y is wL, and for capital it's rK. That is, w is the wage and r is the rental rate on capital.

Let's consider the X-industry. To produce X units of output, there are lots of combinations of L and K that work. Their goal, to maximize profit, is to pick the least-cost combination of L and K to create output X.

We can think of all combinations of L and K that create X using an *isoquant*. Figure 6.1 shows this: because the production function has constant returns to *both* L and K, it has diminishing marginal returns to either. Therefore, just like a utility function, it should be bowed towards the origin.

When factor prices are w and r, industries follow a profit-maximizing logic that should be familiar by now.

Suppose an industry was using some combination of L and K to produce X. If they gave up a unit of L to

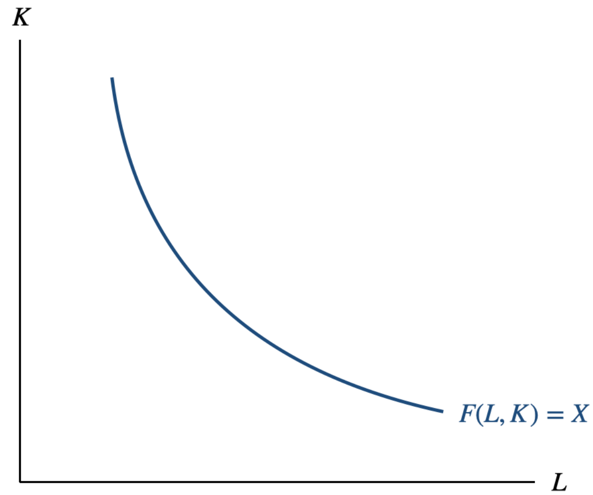


Figure 6.1: An isoquant for X

buy some K instead, they would make w dollars with which they could buy w/r units of K. The loss from giving up one unit of labor is the marginal revenue product of labor, $MRPL$. The gain in revenue from using one more unit of capital is its marginal revenue product, $MRPK$. As before,

$$MRPL < \frac{w}{r} MRPK \implies \text{use less labor, more capital}$$

and similarly,

$$MRPL > \frac{w}{r} MRPK \implies \text{trade capital for labor}$$

Therefore, the profit maximizing point occurs when,

$$\frac{MRPL}{MRPK} = \frac{w}{r}$$

We can display this graphically, too. If a firm is using L_0 units of labor and K_0 units of capital, the cost is $C_0 = wL_0 + rK_0$. To profit maximize, they pick a combination of L and K to minimize costs C while still producing $X = F(L, K)$ units of output. The combination of labor and capital that minimizes costs to produce X units of output is L_X, K_X, as shown in Figure 6.2. The profit maximizing, or expenditure minimizing nominal cost to produce X units is $C_X(X) = wL_X + rK_X$: note how the choices of L_X and K_X depend on X and the factor price ratio!

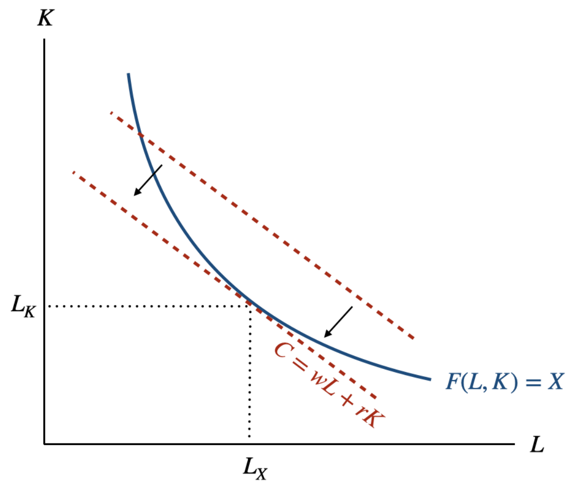


Figure 6.2: Expenditure minimization in the HO model

So, now we know given factor prices w/r, how industries choose the level of L and K they want to use.

6.2 Market clearing and the Edgeworth box

The last section should be unsatisfying – where to factor prices come from?

As in the endowment economy, factor prices are such that the factor market clears. In the goods market in the endowment economy, we found P_X/P_Y so that there were no shortages or surpluses using the Edgeworth box. The exact same thing is true here. In equilibrium, we need that,

$$L_X + L_Y = \bar{L}, \quad K_X + K_Y = \bar{K}.$$

These equations mean that all the endowed labor and capital have to be used up by either industry. Factor prices w/r adjust until there is no shortage or surplus of either factor.

We can again use the Edgeworth box to analyze how this works. If both industries set the ratio of their marginal revenue products to the factor price ratio, that means in equilibrium, their marginal product ratios are equal, too:

$$\frac{MRPL_X}{MRPK_X} = \frac{w}{r} = \frac{MRPL_Y}{MRPK_Y}.$$

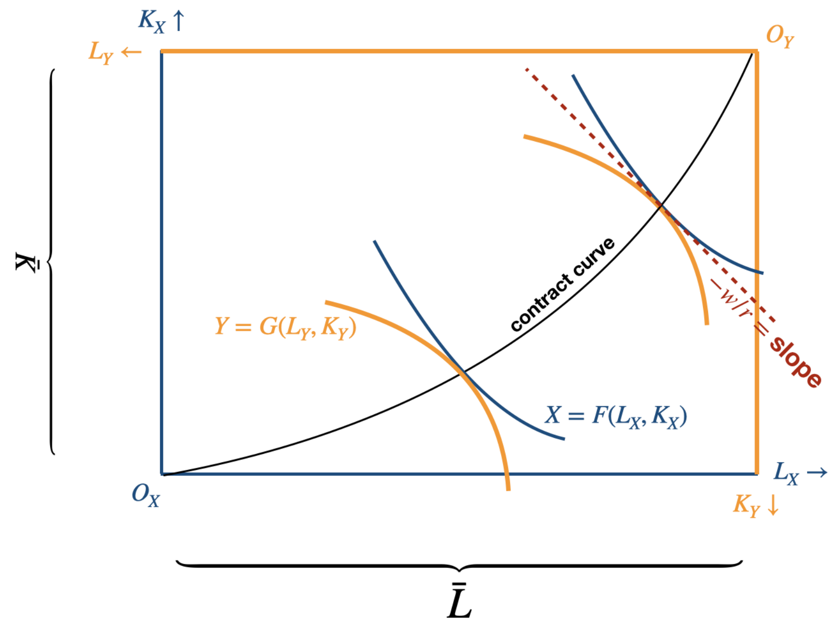


Figure 6.3: The Edgeworth box in the HO model

This is the same logic as profit-maximizing when the only mobile factor was labor. In the labor market, it was profit maximizing to set $MRPL = w$. Now, if firms set $MRPL = w$ and $MRPK = r$, then the above condition is also satisfied.

The *contract curve* in the HO economy is the set of all L, K combinations that split up $\bar{L}$ and $\bar{K}$ such that there is market clearing ($L_X + L_Y = \bar{L}, \quad K_X + K_Y = \bar{K}$) and industries profit-maximize ($\frac{MRPL_X}{MRPK_X} = \frac{MRPL_Y}{MRPK_Y}$).

This is displayed in the Edgeworth box for the allocation of factors in Figure 6.3. The length of the axes represents the total endowment of L and K in the economy, and so every point in the box represents a way to split up L and K between the two industries such that the full endowment is used. It is very similar to the Edgeworth box we used to analyze the endowment economy at the beginning of the course.

Every point along the contract curve represents a split of the factors L and K so that both are used up by both industries, and both industries are profit-maximizing, meaning that the ratio of the marginal products of labor and capital are equal.

Factor prices are simply the slope of the isoquants wherever we are along the contract curve. The factor price ratio adjusts along the contract curve, as the slope of the isoquants is different at different points, which is

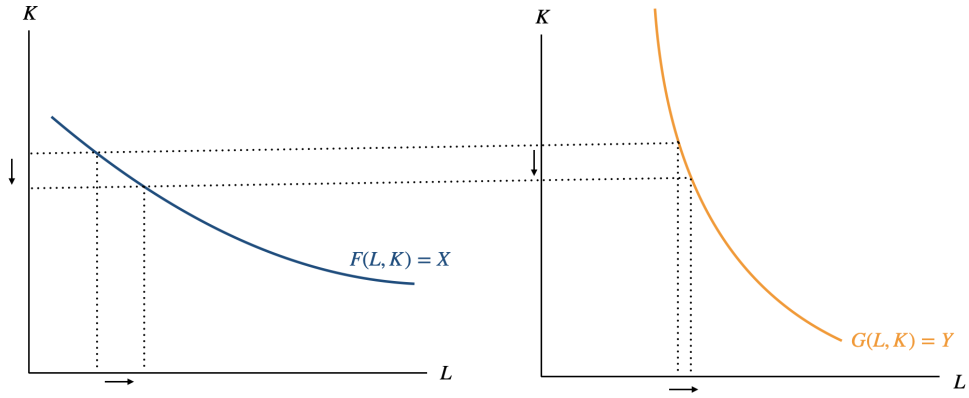


Figure 6.4: Reading factor intensity off isoquants

what we explore in the next section.

6.3 Factor intensity

What determines the slope of the contract curve and the shape of the isoquants? Factor intensity.

An industry's factor intensity is a relative concept: does one industry demand relatively more of one factor than the other?

Factor intensity can be understood using Figure 6.4. In the figure, to hold output constant, if the X-industry loses one unit of K, they need a lot of L to compensate. On the other hand, the Y-industry has a much steeper isoquant: a much smaller amount of L is needed to compensate for the loss of one unit of K. This means that the X-industry is relatively *capital-intensive*, i.e., relative to the Y-industry, which is more *labor-intensive*.

A common example in economics is a Cobb-Douglas production function that relates labor and capital to output in the following form,

$$X = (L)^{\alpha_X}(K)^{1-\alpha_X}.$$

We then can use the number α_X $(0 < \alpha_X < 1)$ to refer to an industry's labor intensity. If $\alpha_X > \alpha_Y$, we would say that the X-industry is more labor intensive.

The shape of the contract curve depends on which industry needs relatively more L or K. Consider two

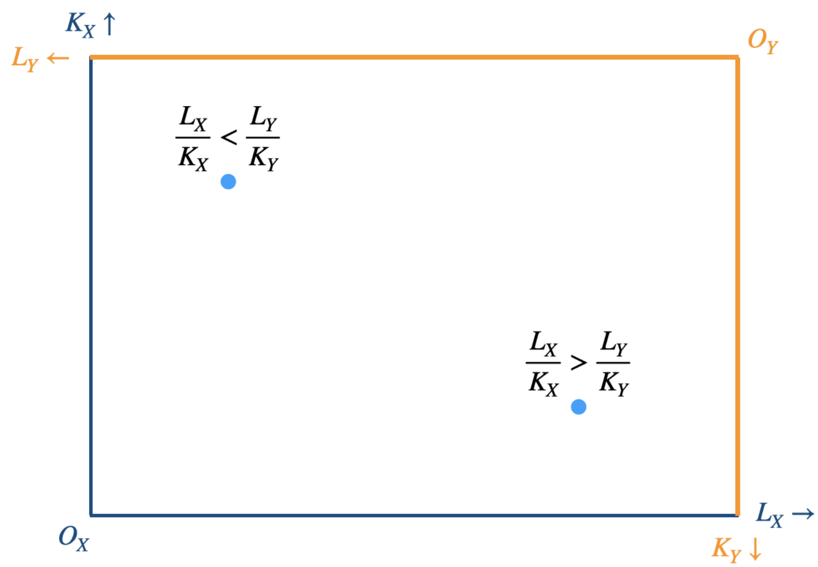


Figure 6.5: Determining factor intensity in the Edgeworth box

points in the Edgeworth box, as in Figure 6.5. At a point in the Southeast corner of the Edgeworth box, L/K is bigger in the X-industry than the Y industry. The opposite is true in the Northwest corner.

Just like in the Endowment economy, the allocation of factors in the HO model is 'efficient,' meaning that the price system splits up L and K between the two industries the same way a social planner would. As a consequence, if the X-industry is more labor-intensive, a social planner would want to make sure it always got relatively more labor, and so the contract curve would move through the southeast of the Edgeworth box. Conversely, if the X-industry was capital-intensive, then the contract curve would move through the northwest of the Edgeworth box.

6.4 The contract curve and the PPF

Recall that every point on the contract curve represents two things: one, it represents using up all the endowed factors for production; and two, that those factors are split up across the industries in an efficient way – i.e., in a profit-maximizing (or, equivalently, expenditure-minimizing) way. Each point also corresponds to an output of X or Y. Therefore, each point on the contract curve must be related to a point on the economy's PPF.

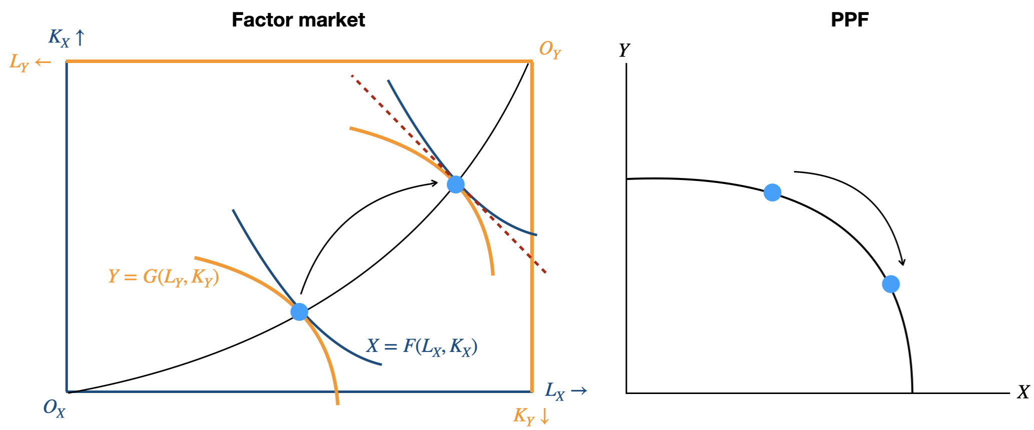


Figure 6.6: The PPF in the HO economy

Figure 6.6 shows the relationship between the PPF and the contract curve. For every point along the PPF, there is a way so that factor prices w/r can split up factors of production to product X and Y efficiently.

In the Specific Factors Model, movements along the PPF corresponded to movements in the labor market, and we used the labor market diagram to understand how labor moved between the two industries, and how the wage changed. Now, in the Heckscher-Ohlin model, there are two factors, labor and capital, and we use the Edgeworth box to study what happens to factor prices as we move along the PPF.

6.5 Factor intensity and factor prices

As we move along the PPF, labor and capital must reallocate from one industry to the other. What happens to factor prices depends on the relative factor intensity of that industry. Factor intensity also refers to the productivity of factors within an industry: a more labor intensive industry demands relatively more labor because labor is relatively more productive.

Consider the example in the last section. If the X-industry is more labor-intensive, as the economy produces more X and less Y, labor is reallocated to an industry where it is on the margin more productive, and capital is reallocated towards an industry where it is, on the margin less productive. This means that the factor price ratio should go up ($w/r \uparrow$), and that means the slope of the ratio of marginal products should become steeper!

If the X-industry was more capital-intensive, then the contract curve would move through the Northwest of

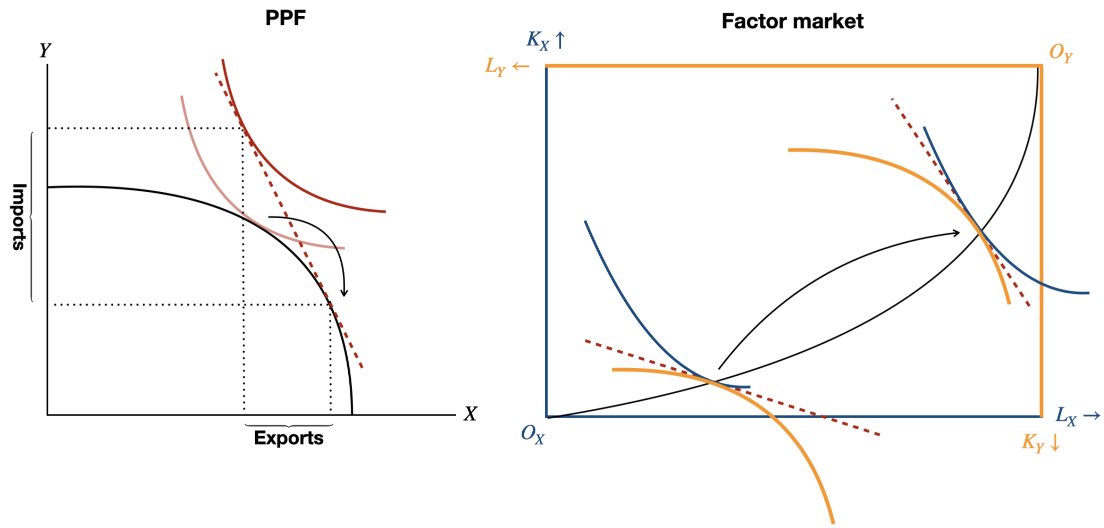


Figure 6.7: Opening to trade in the HO economy. X is more labor-intensive.

the Edgeworth box. Moving along the PPF as in Figure 6.6 would mean capital would need to reallocate to an industry where it is on the margin more productive, and labor needs to reallocate to an industry where it is on the margin less productive, meaning that the factor price ratio must get less steep ($w/r \downarrow$).

We now have the tools to analyze opening to trade in the HO economy.

6.6 Opening to trade in the HO economy

Opening to trade looks much like before. Suppose the Home economy opens to trade, and at Home, the X-industry is labor intensive.

In Figure 6.7, I illustrate opening to trade in the HO economy. I suppose world prices are steeper than autarky prices, incentivizing producers to produce more X and less Y. In the left panel, there are gains from trade, because the consumption possibilities frontier broadens. The economy finances its imports by selling exported X in the world market.

Domestically, factors reallocate from producing Y to producing X. As X is more labor-intensive, the marginal worker becomes more productive and the marginal unit of capital becomes relatively less productive; w/r rises.

In short, in this example, trade benefits workers because the economy exports the labor-intensive good. If X-were capital intensive, w/r would fall.

If w/r rises, it means that workers are the 'winners' from trade, while capitalists 'lose.' On the whole, aggregate welfare still goes up, which means the losers can be compensated – trade can constitute a Pareto improvement.

6.7 The pattern of trade in the HO economy

In the last section, I just supposed what world prices were. Of course, as in the Ricardian and Specific Factors Models, world prices depend on what Foreign does.

To know what world prices will be, and what whether trade will incentivize X or Y production at home, we need a sense of what comparative advantage is in the HO economy.

To do that, I will explain the main 'theorems' of the HO model. These will be offered without mathematical proof, as it's beyond the scope of this class.

The main theorems you need to know for this class are the HO theorem and the Stolper-Samuelson theorem.

There are two other key theorems in the HO model that we will not dedicate time to study. They are the Rybczynski theorem, and the Factor Price Equalization theorem. The first concerns what happens when factor endowments change (which we explored in the Specific Factors Model) and the latter claims that w and r must be equal in both countries. That is, if free international trade can perfectly replicate an integrated equilibrium (that trade in goods perfectly substitutes for trade in factors), then factor prices must equalize across nations. There is little empirical evidence for this, likely because the assumptions of the HO model that are critical to this result do not hold up in practice.

However, the other two HO theorems are more intuitive and have better empirical support. Moreover, they show us how to clarify our understanding of comparative advantage in economies that are more complicated than Ricardian economy we analyzed at the beginning of the course.

6.7.1 The HO theorem

The Heckscher-Ohlin theorem is,

An economy will export a good that uses its relatively abundant factor most intensively.

For example, if Home is endowed with relatively more capital than foreign, i.e., $\bar{K}/\bar{L} > \bar{K}^*/\bar{L}^*$, and the Y-industry is capital-intensive relative to the X-industry, Home will export Y.

Loosely, countries have a comparative advantage in the industry that uses their most abundant factor (relative to Foreign) most intensively (relative to their other domestic industry).

World prices are such that the global market for imports and exports clears. In doing so, they incentivize the (incomplete) specialization in countries 'comparative advantage' industries, in accordance with the HO theorem.

6.7.2 Stolper-Samuelson

The Stolper-Samuelson theorem is,

A increase in the relative price of a good will lead to an increase in the price of the factor that is used most intensively in its production, and a fall in the price of the other factor.

For example, if P_X/P_Y rises, the economy will produce more X. If X is labor-intensive, then w/r will rise.

This is what analyzed in the last section.

Together, these two theorems allow us to understand what happens when a country opens to trade: what will it export, and how will domestic factor markets respond? The HO theorem tells us what the pattern of trade will be, and Stolper-Samuelson tells us the winners and losers from trade!

Lecture 7 – Trade policy under perfect competition

Now we have all the pieces of the neoclassical model, we can sew them together to use the model to analyze the tools of trade policy. The different types of goals of trade policy and their names are given in Table 7.1.

	Increase	Decrease
Imports	Trade liberalization	Protectionism, Import substitution
Exports	Export promotion, (Export-oriented) industrial policy	Voluntary export restrictions

Table 7.1: The goals and names of different types of trade policy

In this lecture, we will study Protectionist and export-promotion policies, like tariffs and quotas which seek to insulate domestic industry from price competition in world markets, or subsidies, which are used to make domestic industries more competitive abroad.

7.1 Domestic and world supply and demand

First, let's recall where we can get world market supply and demand curves from our neoclassical theory, like the Heckscher-Ohlin model.

First, demand comes from consumers maximizing their utility subject to a budget constraint, as in Figure 2.6, which is reproduced as Figure 7.1. As the price of X falls, there are two effects: consumers are relatively richer and so they buy more X (income effect) and the price of X is relatively cheaper than Y, so they substitute towards it (substitution effect). For normal goods, on net these two forces compound so the quantity of X demanded rises as the price falls.

Supply comes from the concave PPF in the Specific Factors or Heckscher-Ohlin models as in Figure 5.4 (reproduced in Figure 7.2). As the relative price steepens, the optimal production mix of X and Y changes in factor markets reallocate factors from Y to the X industry, generating an upward sloping industry supply curve.

Global supply and global demand for a good come from simply adding Home and Foreign's supply and

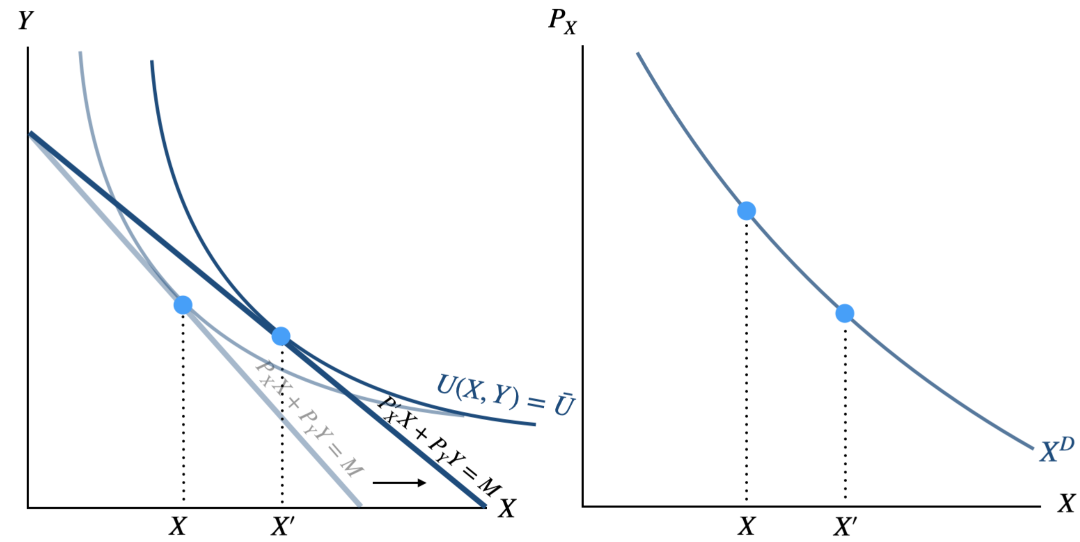


Figure 7.1: Left: Consumer optimization under a price change; $P'_X < P_X$. Right: demand curve.

demand curves together.

If Home is an exporter of X, it means that the world price of X must be higher than the autarky price (to incentivize exporting). If Home is an exporter of X, it means that Foreign must import it. Foreign would import X only if the world price is less than its autarky price. Therefore, as in the Ricardian model, the world price must be between the two autarky prices!

These ideas come together in Figure 7.3. In the left panel is domestic supply and demand, the middle is Foreign, and the world market is displayed in the final panel. World prices are between the two autarky prices, incentivizing domestic to export X and for Foreign to import X. The price is exactly such that the quantity of Home's exports equals the quantity of Foreign's imports – the world market clears; nothing is lost in the sea!

If we were derive domestic demand, $D(P)$ (which tells us the quantity demanded of a good by consumers at Home given a price P) and domestic supply $S(P)$ (which tells us the quantity of a good Home's industry would produce at price P), as well as the similar supply and demand schedules for Foreign, D^* and S^*, we can represent the autarky equilibrium as a set of two autarky prices (one for home, P, and one for Foreign,

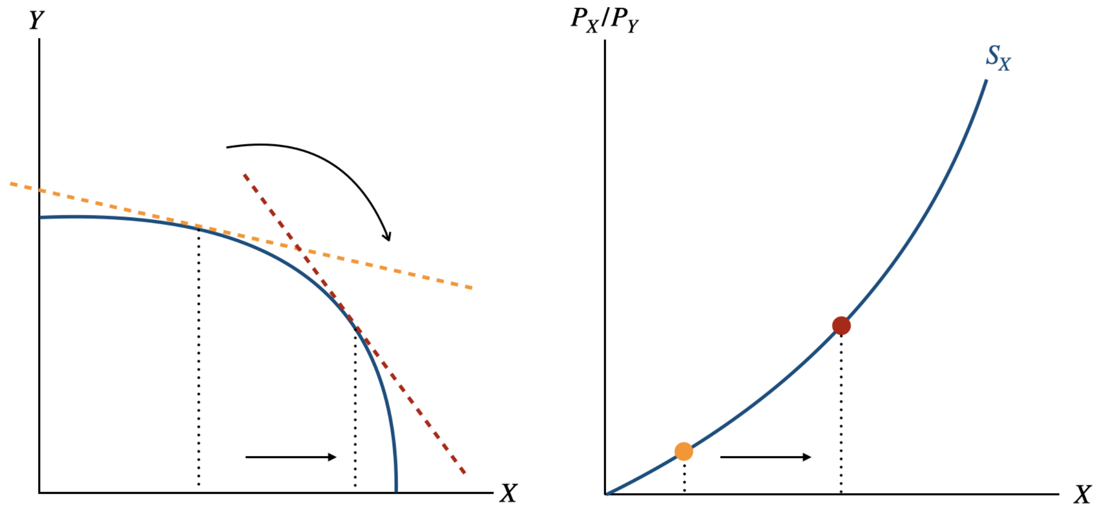


Figure 7.2: Domestic supply in the neoclassical model

P^*) that solve the equations,

$$\underbrace{D(P) = S(P),}_{\text{autarky at home}} \quad \text{and} \quad \underbrace{D^*(P^*) = S^*(P^*)}_{\text{autarky in Foreign}}.$$

As a matter of notation, if the good we are discussing is X, then $X^D = D(P)$ is the quantity demanded, while $X^S = S(P)$ is the quantity supplied.

Free trade occurs at a world price P^W that clears the global market for a good,

$$\underbrace{S(P^W) + S^*(P^W)}_{\text{world supply}} = \underbrace{D(P^W) + D^*(P^W)}_{\text{world demand}}$$

We can rearrange global market clearing (global supply $=$ global demand) to represent the global market clearing in terms of exports and imports, as below:

$$\underbrace{S(P^W) - D(P^W)}_{\text{Home's export supply}} = \underbrace{D^*(P^W) - S^*(P^W)}_{\text{Foreign's import demand}}.$$

As I've written it and illustrated in Figure 7.3, if $P^W > P$ then $S(P^W) > D(P^W)$, and there is a surplus produced at home, which is exported. In Foreign, $P^W < P^*$ means that $D^*(P^W) > S^*(P^W)$, and there is a shortage in Foreign which is met by buying imports in the world market. In the supply-and-demand diagram,

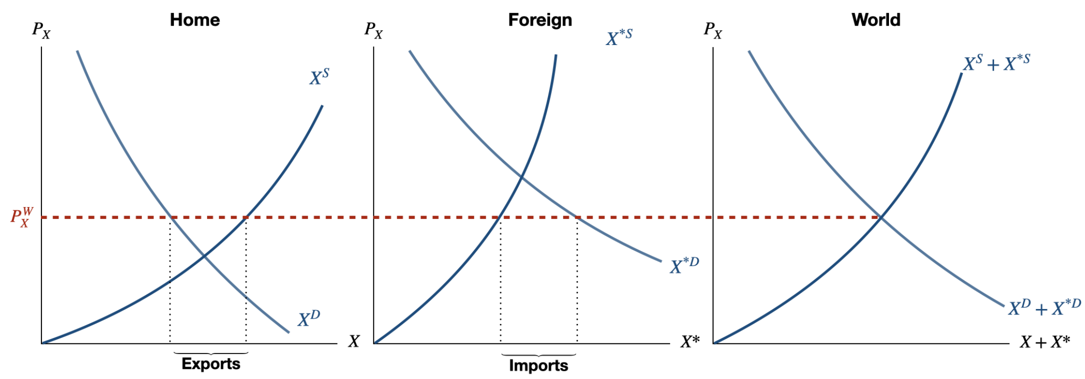


Figure 7.3: The world equilibrium in the supply-and-demand representation of the neoclassical trade model

we can read off the quantity of imports and exports as the difference between supply and demand at price P^W along the X-axis. The world price that equates world supply = world demand is the only P^W that also ensures that exports = imports. This can be shown graphically, as in Figure 7.4.

To demonstrate this using a different equilibrium, suppose Home imports X, while Foreign supplies it. Home has demand for imports $X^D - X^S$ whenever the world price falls below the autarky price. So we can trace out Home's *import demand curve*, $X^D(P) - X^S(P)$ by writing down the amount of X-imports demanded at home for P less than Home's autarky price. Likewise, whenever the world price exceeds Foreign's autarky price, Foreign has excess supply, which can be exported. Foreign's *export supply curve* starts at P equal to its autarky price, and $X^{*S}(P) - X^{*D}(P)$ grows as P rises. Where Home's import demand curve equals Foreign's export supply curve is the free trade equilibrium where imports = exports. From this representation of the world market, we can see why the world price is between the autarky pricess.

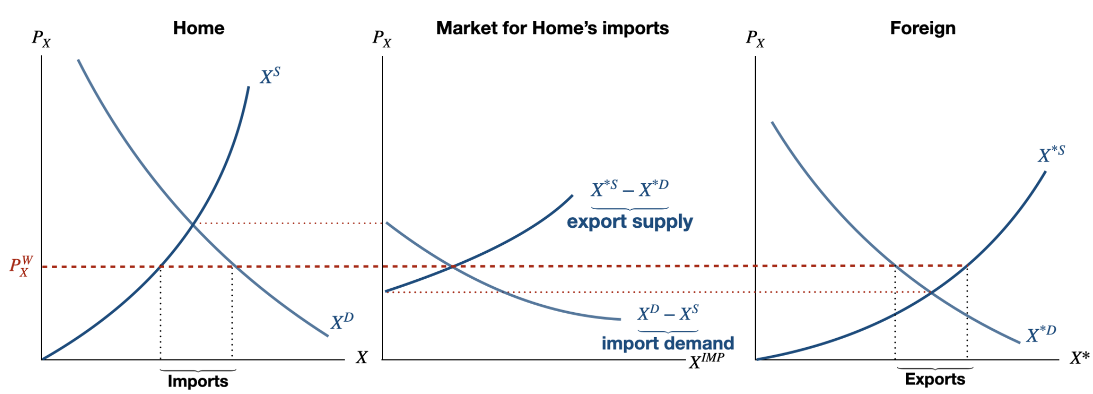


Figure 7.4: The market for Home's imports.

7.2 Tariffs

Tariffs are a tax on imports. Tariffs are called 'protectionist' policy, because they protect domestic industry from competition in world markets. A tax on imports will raise the price domestic consumers pay for a good, which reduces gains from trade (or extra consumer surplus) households receive from importing goods at lower prices in world markets. While this higher price hurts consumers, domestic producers can benefit as prices are higher. As they are shielded from Foreign competition, domestic producers will also produce more, too. Therefore, tariffs reduce imports both households demand less (a movement along the domestic demand curve) and because producers supply more (a movement along the domestic supply curve).

We will model tariffs as a *per unit tax* on each good imported. In a small open economy, a tariff of size T will cause imports to cost the world price plus T, $P^W + T$.

7.2.1 Tariffs in a small open economy

In a small open economy, we assume that when Home imposes import tariffs, it does not affect the world price. All of the analysis will carry thorugh if Foreign is importing and imposing tariffs on Home; you would just need to relabel Home and Foreign. Figure 7.5 demonstrates graphically the effect of an import tariff in a small open economy that is importing X. Before the tariff, the world price is lower than the autarky price and so Home imports X to make up for the gap in domestic supply.

When the tariff is imposed, the domestic industry finds it profit maximizing to expand at the new price. Facing higher prices, consumers demand less, and so the size of imports contracts.

Profits expand in the domestic industry, while consumer surplus shrinks. Some lost consumer surplus is a transfer to producers. This is displayed by region 1 on the Figure. Some lost consumer surplus is made up for from government revenue from the tariff. That is, the government makes back the lost consumer surplus in the form of revenue generated by the tariff. The amount of revenue they generate is equal to the size of the tariff, T times the number of imports, i.e., the area of rectangle 3 on the figure.

However, not all consumer surplus lost due to the tariff comes back in the form of tariff revenue or producer

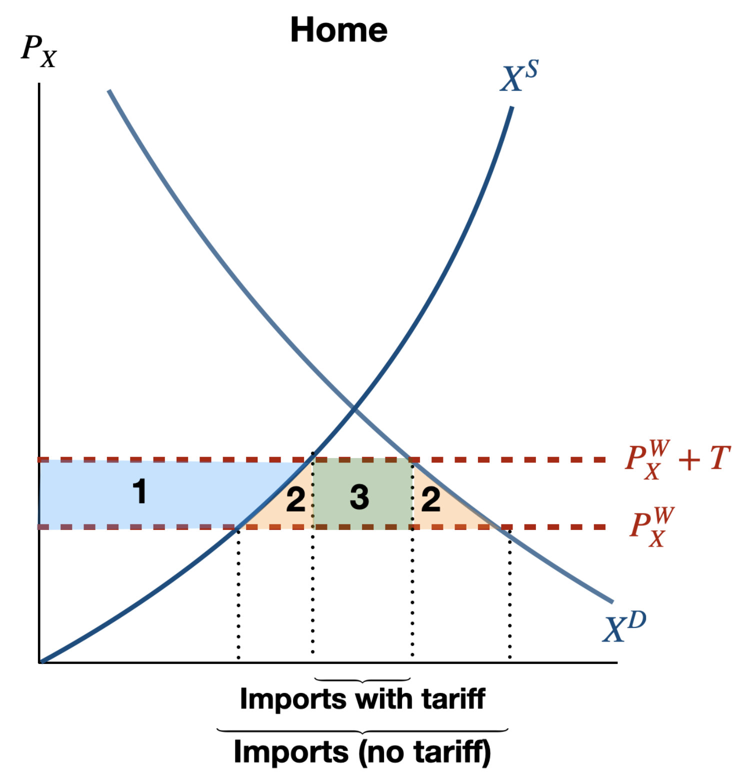


Figure 7.5: Tariffs in a small open economy

surplus – it is deadweight loss, an inefficiency that arises because prices no longer find a Pareto efficient allocation. The amount of deadweight loss is given by the two triangles on the figure, regions 2.

7.2.2 Tariffs in a small open economy on the PPF

Another way to view the welfare effects of tariffs is by studying their effects using a PPF diagram. In this section we will repeat the analysis of the last section, assuming a tariffs are placed on the imported good, X.

The tariff changes the terms-of-trade faced by producers. A tariff of T dollars per unit X makes the terms of trade steeper by affecting the nominal price of X,

$$\frac{P_X + T}{P_Y} > P_X/P_Y.$$

These steeper terms-of-trade will incentivize production of the X good, and the production point along the PPF will rotate towards X. The factors of production are reallocated out of the production of the Y good towards the production of X.

Consumers face a higher price too, and face terms-of-trade with the same slope as producers. However, the consumption possibilities frontier will no longer be represented by the line that runs through the production point with slope equal to the terms of trade. To derive it, we must understand what happens to the tariff revenue raised by the government. Per unit imported, the government makes T dollars, which we will assume are rebated back to households at Home in dollar terms and can be spent to finance the consumption of both X and Y.

Household income comes from production and the government revenue,

$$\text{Income} = \underbrace{(P_X + T)X^S + P_Y Y^S}_{\text{producer income}} + \underbrace{TX^I}_{\text{tariff revenue}}$$

where X^I represents the quantity of X imported, $X^I = X^D - X^S$. Household expenditure depends on their utility-maximizing choices X^D and Y^D,

$$\text{Expenditure} = (P_X + T)X^D + P_Y Y^D$$

and so their budget constraint, Expenditure $\leq$ Income, can be represented as,

$$(P_X + T)X^D + P_Y Y^D \leq (P_X + T)X^S + P_Y Y^S + TX^I$$

where households take TX^I, the transfer of tariff revenue back to households, as part of income and exogenous. However, *in equilibrium*, that transfer has to coincide with the choices households and firms make optimally (in particular, households' optimal choice of X^D given prices $P_X + T$ and P_Y). Plugging in the equilibrium definition of $X^I = X^D - X^S$ and rearranging the budget constraint,

$$(P_X + T)X^D + P_Y Y^D \leq (P_X + T)X^S + P_Y Y^S + T(X^D - X^S)$$

$$\text{implies in equilibrium,} \qquad P_X X^D + P_Y Y^D \leq P_X X^S + P_Y Y^S.$$

What this says is that the consumption choice of households, in equilibrium, must fall on a line that runs

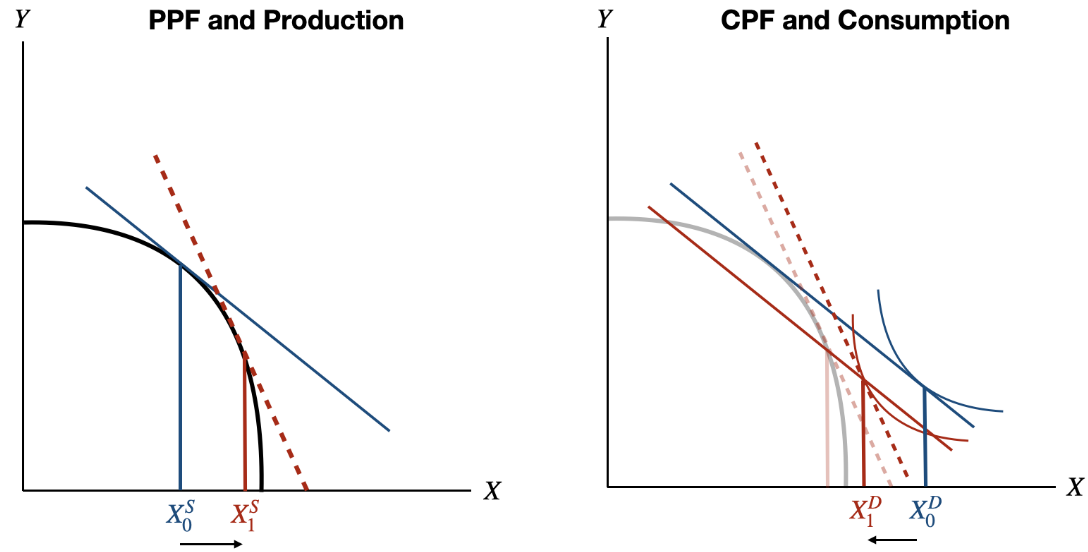


Figure 7.6: Tariffs in a PPF diagram. Tarifs rotate the terms-of-trade faced by producers, as they result in steeper prices. Production moves from X_0^S prior to the tariff (in blue) to X_1^S after the tariff, in red.

through the new production point, but has slope given by *the original terms-of-trade*. This is because the household is still trading in world markets at prices P_X/P_Y (the small open economy assumption). The tariff only affects the price of goods once they cross the border, and amount to a transfer from households to producers.

These effects are shown in Figure 7.6. In the left panel, because of the tariff on X, the terms-of-trade faced by producers is steeper, and production moves from X_0^S to X_1^S, an increase in the supply of X. In the right panel, we see that tariffs lower national income as the CPF running through (X_1^S, Y_1^S) at world prices is below the old CPF. As households pay a high price T, in equilibrium their perceived CPF has slope $(P_X + T)/P_Y$ and goes through a point above the production diagram because government tariff revenue is rebated back towards them. As national income is lower and the price of X faced by domestic households is higher, consumption after the tariff X_1^D is to the left of consumption prior to the tariff X_0^D.

So households' consumption point must be affordable in equillibrium (i.e., on the line that runs through X^S, Y^S with slope $-P_X/P_Y$) but consistent with utility maximization. That is, their indifference curve is tangent to a line with slope $(P_X + T)/P_Y$ that runs through a point above X^S, Y^S, because of the transfer of tariff revenue.

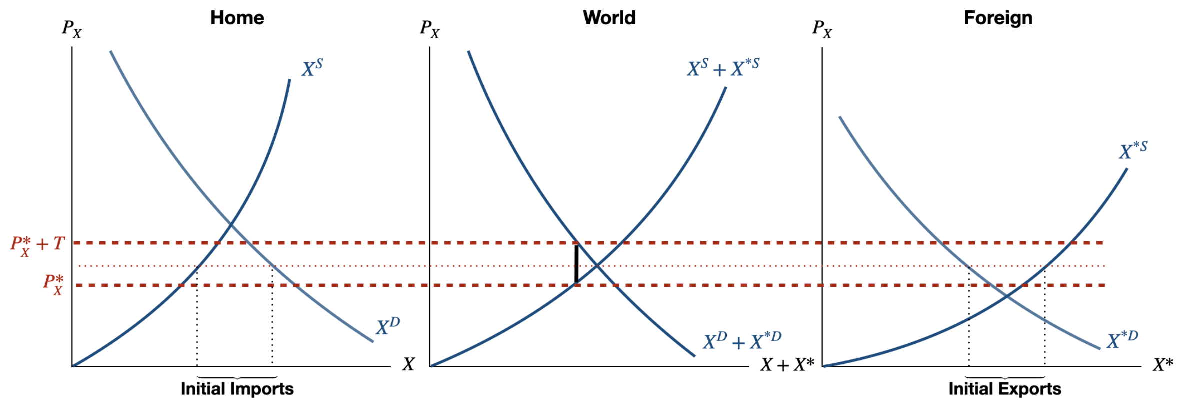


Figure 7.7: Tariffs in a large open economy – the world market. The tariff causes us to move up the world demand curve. The price faced by foreign must fall to reduce the surplus of X in the world market.

This diagram highlights from where the deadweight loss is coming. As shown in the supply-and-demand figure, the deadweight loss is coming from lost consumer surplus. Production is still efficient and occurs on the PPF. The distortion to prices and income *faced by households*, which are caused by the tariff, result in an aggregate welfare loss.

7.2.3 Tariffs in a large open economy

In a large open economy, we assume that the country that imposes tariffs affects the world price.

This happens because the domestic contraction in demand for imports is large enough that Foreign needs to supply *less* so that there is not a surplus in world markets. The tariff drives a wedge between the price faced at Home and in Foreign: Foreign must face a lower price to export less, while Home faces a higher price due to the tariff.

This is shown in Figure 7.7. Due to the contraction in world demand for exports (because of the tariff) Foreign faces a new price which reduces their exporting. The wedge between the price paid at Home and the price received in Foreign is exactly the size of the tariff, and is shown by the thick black bar in the figure.

Let's zoom in at see what happens at Home when they impose a tariff and affect the world price by examining Figure 7.8. When the tariff causes the world price to lower, it means that imports, prior to the imposition of the tariff are cheaper than they were before. The rise in domestic price is then less than T!

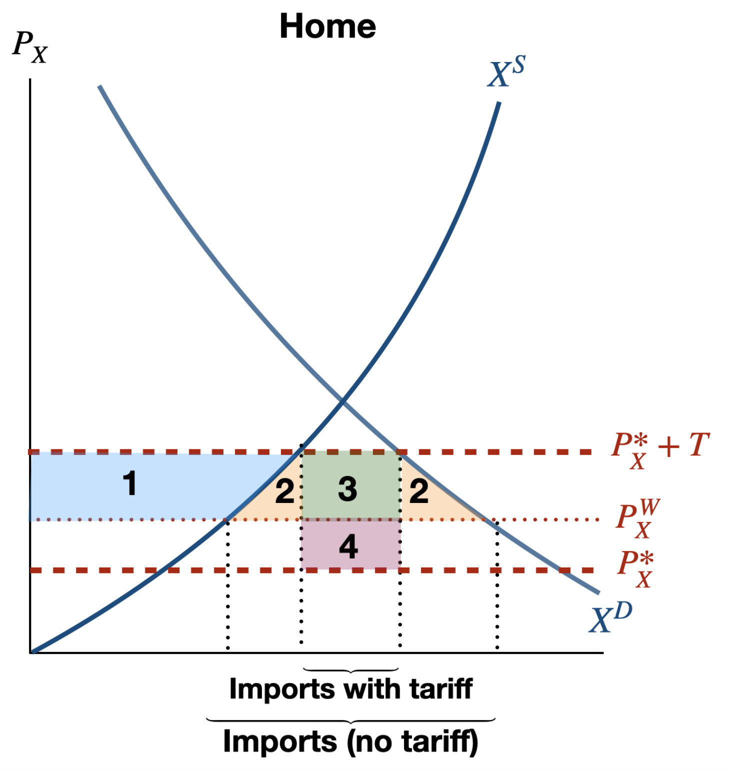


Figure 7.8: The 'terms-of-trade' gain from a tariff.

Therefore, not all of the government revenue is a transfer from consumer surplus: some of the incidence of the tariff falls on Foreign exporters! This additional component of revenue means that the revenue gain from imposing a tariff can actually outweigh the domestic deadweight loss from the tariff. This is called the 'terms-of-trade' gain (TOT gain), and is shown by the rectangle labeled 4 on Figure 7.8.

While it can be rational for a large country to exploit its monopoly power in global markets and manipulate terms-of-trade with tariffs, since the terms-of-trade gain can outweigh the deadweight loss, it is still inefficient from a global standpoint – the TOT gain is a transfer from Foreign to Home, and the deadweight loss still exists from a global perspective.

7.2.4 Optimal tariffs and trade wars

The existence of a terms-of-trade gain means that the total welfare effect of a tariff is ambiguous. Large countries in the global economy can manipulate the terms of trade to their benefit by imposing tariffs.

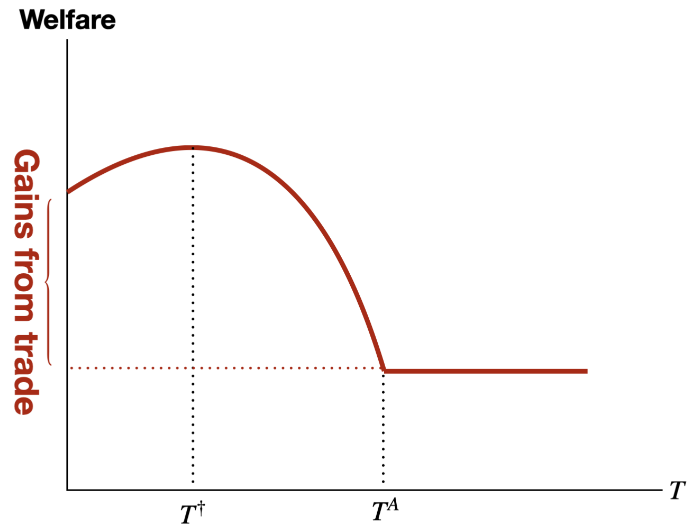


Figure 7.9: Welfare vs tariffs in a large open economy.

Of course, if a tariff is set too high – above the autarky price – the country would simply retreat into autarky. This means there is a nonlinear relationship between welfare and tariffs for large open economies. Small increases in tariffs may be good, due to the terms of trade effect, but eventually these gains are eroded as the tariff approaches the autarky price. This implies the existence of an optimal tariff, $T^\dagger$. As there are gains from trade, welfare under autarky must be lower than without the tariff at all. Let's call the tariff that makes a country indifferent to trade (a prohibitive tariff) T^A. The resulting relationship between the size of a tariff and welfare looks like Figure 7.9.

The existence of an optimal tariff relies on other countries not imposing tariffs on you. However, tariffs are still inefficient globally – the terms-of-trade gain comes at the expense of the exporting countries.

One mechanism countries have to discourage those purchasing their import goods from levying tariffs is the threat of retaliation. This can result in a *trade war*, in which countries mutually escalate tariffs against each other.

Let's consider this from a game theoretic standpoint. If Home optimally imposes a tariff, they gain, but Foreign loses. The reverse is also true, and if they both impose tariffs, they are both worse off than under free trade. A possible payoff matrix for this scenario is given in Table 7.2. Note that this resembles the Prisoner's Dilemma. Both countries have an incentive to impose tariffs optimally, but if they both 'defect' from the

	Free trade	Impose tariffs
Free trade	$H:10, F:10$	$H:12, F:7$
Impose tariffs	$H:7, F:12$	$H:5, F:5$

Table 7.2: A payoff matrix for a game of tariff-setting

free-trade equilibrium, everyone is worse off.

The existence of international entities like the World Trade Organization (WTO) and before it the General Agreement on Tariffs and Trade (GATT) promotes bilateral trade agreements in which both parties can agree to lower tariffs and 'cooperate' in the free trade equilibrium.

7.3 Quotas

Quotas are limits on the quantity of imports a country can receive. They are implemented through import licenses: giving a select set of importers the right to buy foreign exports and sell them in domestic markets, so long as those imports fall below the quota.

We will only study quotas in a small open economy. That is because, as will become clear in this section, for every quota, there exists a tariff that does the exact same thing.

An easy way to think about quotas is that a country remains in autarky, but domestic supply is a little different now. Goods can be supplied either by domestic industry, or by those who hold import licenses. The domestic supply curve then comes from profit-maximizing firms and those with the import licenses.

Let's think about what the domestic supply curve must look like with a quota by slowly raising the price and thinking about what production must look like. When the price of a good is low, it is profitable only to produce a small amount of goods by the domestic industry. As the price rises, it becomes profitable for domestic industry to supply more and more of it. Once the price hits the world price, suddenly, it is profitable for those who hold import licenses to supply, and the supply curve expands horizontally by the size of the quota. That is, the marginal cost of a good becomes the world price: instead of using domestic factors to produce the good and continue to run into diminishing returns (which give the supply curve its upwards slope), the cost of an additional unit is the cost to import it. To make an additional unit above the quota,

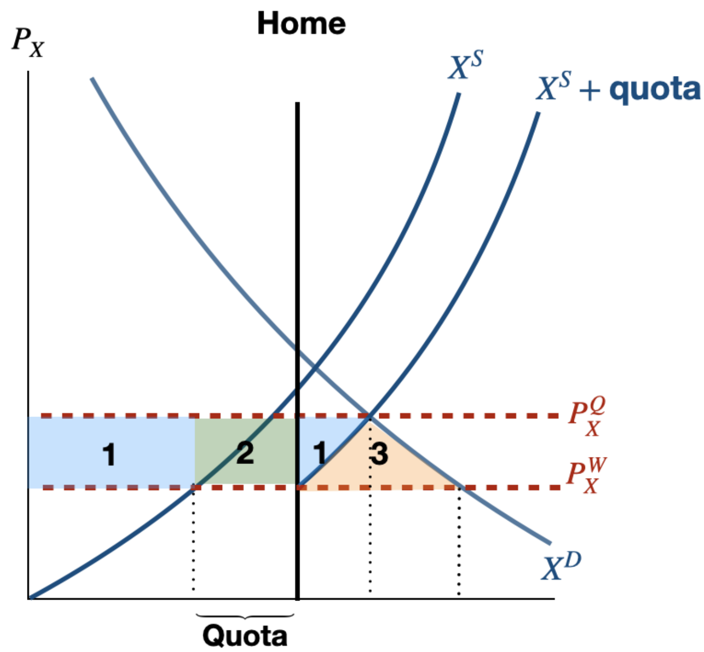


Figure 7.10: A quota in a small open economy

domestic industry has to again take over, and we again move up the domestic industry curve. This amounts to a very odd shift in supply, as shown in Figure 7.10.

As the quota causes an increase in domestic supply relative to the autarky price, the price falls relative to autarky. Conversely, the quota restricts the number of imports, so the relative to a free-trade equilibrium, the price must rise. As the price has risen, consumer surplus must have fallen. Some of this lost surplus is made up for in the form of increased producer surplus (as shown by regions 1 on the figure). Quotas, like tariffs, are protectionist – they shield domestic industry from competition abroad.

Those who hold import licenses get to buy at the world price and sell at the domestic price: their profits are the amount they can import times the change in price. These are called *quota rents*. They are exactly analogous to government revenue from a tariff and shown by rectangle 2 in the Figure. Finally, there is still deadweight loss from the policy, as shown by triangle 3. This is lost consumer surplus that is not transformed into producer surplus or quota rents.

Call the price under the quota P^Q. A tariff of size $T = P^Q - P^W$ would cause the exact same contraction in demand and increase in domestic supply, and therefore the same deadweight loss. The only difference

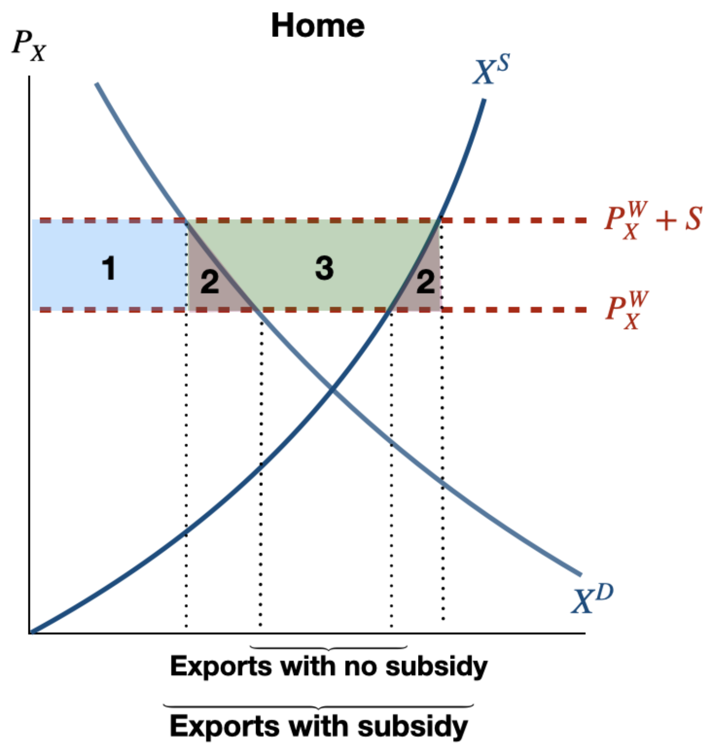


Figure 7.11: A subsidy in a small open economy

is whether the revenue from the policy goes to the government (under tariffs) or to those who hold import licenses (under quotas). The latter creates an opportunity for corruption and political favoritism!

7.4 Export subsidies

Subsidies are typically applied to export industries. The policy rationale varies – misguided notions of trying to encourage export-led growth, political reasons to promote goods and influence abroad, protecting industries that are affected by changes in international prices and exchange rates, or other political economy reasons. There are other motivations for trade policy that we will explore in later lectures.

We will model an export subsidy S as a per-unit subsidy that increases the price received by exporters at Home. Figure 7.11 demonstrates the effect of an export subsidy in a small open economy.

First, by raising the price producers face, the export industry finds it profitable to increase production. As the subsidy causes the export industry to run into diminishing returns, the marginal unit produced becomes more costly and the price faced by domestic consumers rises too. Thus we move up the demand curve. The increase in supply and decrease in domestic demand both cause exports to rise.

Some of the decrease in domestic consumer surplus is transformed into producer surplus, given by area 1. The total change in producer surplus is given by areas 1, 2 (left triangle) and 3. However, the government must pay S times the number of goods exported to finance the subsidy, given by the rectangle formed by areas 2 and 3. Therefore, the left triangle of 2 is consumer surplus transformed into producer surplus, then paid for by the government, while the right triangle 2 is government expenditure that is not transformed into producer surplus – thus, the sum of regions 2 represents deadweight loss!

Lecture 8 – Economic geography

We've been talking about 'countries' that trade, that differ in their population and factor endowments. Yet, what is a 'country,' economically speaking, and why is that the right unit of analysis to study trade?

Economically, countries are a mix of legal constructs that make it difficult for factors like labor to move between them, as well as culture and shared history that keep people tied to place. However, countries are also big, and have regions in them that might be very different.

When we look at the spatial distribution of economic activity, specialization doesn't seem to happen at the level of a country. Looking within countries, there are two striking features: 1. Economic activity is highly differentiated across space. 2. Economic activity is highly concentrated in space.

Trade theory helps us understand point 1. Factor endowments differ across space, and the law of comparative advantage causes specialization across places. This should apply as much within countries across them, as even within countries, there is a lot of trade. For example, in the United States, New York exports financial services, and Los Angeles exports movies, and both import agriculture from places as close their peripheries and as far as the great plains.

However, trade theory doesn't have answer to why New York and Los Angeles are big and dense and most of Nebraska is not. Factor endowments cannot be the story here – it's difficult to treat these different regions as countries because labor is highly mobile across regions in the United States. I could move to New York, if I wanted! What determines the differences in factor endowments across places?

In this lecture, we will study why economic activity is so spatially concentrated, especially when labor is mobile. The key idea is *agglomeration*: that there are economic reasons to co-locate. This will be balanced by *congestion*: economic forces that cause people to spread out. The balance of these forces create a *spatial equilibrium*. However, before we get into the economics of location, we need to disentangle the economic forces that shape human geography from the fundamental geographic forces that determine location.

8.1 First and second nature forces

First nature forces refer to features that vary across space that shape the distribution of economic activity. These might include access to waterways, good soil, nice weather, and so on. These are like specific factor endowments, making production easier in some places relative to others.

Second nature forces are man-made. These might include access to rail networks, a good labor market. There is circularity to this concept: how can the reason for people and economic activity to be somewhere be that there are people and economic activity there?

This circularity is the essence of 'agglomeration.' Agglomerative forces means economic forces that generate local *increasing returns to scale:* the incentives to be somewhere increase in the size of that place. Taken to its logical end, if there are returns to density, why don't we all live in the same place?

The answer to puzzle of agglomeration is *congestion*. Congestive forces are economic forces that increase local costs as a location grows. A classic congestion force is the scarcity of land: The more people you pack together, the less land there is to live on. Competition for a scarce resource drives up its cost, incentivizing dispersion.

The balance between agglomerative and congestive forces everywhere (on top of first nature features) explains the spatial distribution of economic activity.

8.2 Agglomeration, Marshallian trinity, and Jane Jacobs

Alfred Marshall, in his 1890 work *Principles of Economics* highlighted three features that generate agglomeration economies: (1) labor market pooling, (2) knowledge spillovers, and (3) input-output linkages.

Labor market pooling refers to the fact that in larger labor markets, with more employers and job-searchers, more productive matches can be formed. Both employers and employees are somewhat specialized. From a worker's perspective, it is easier to find an employer that is a good fit when there's more employers to choose from. From an employer's perspective, they can interview more applicants and find the best match in bigger markets.

Knowledge spillovers refers to two things. One, the spatially localized diffusion of production knowledge, and two, the innovative returns to interacting with many people.

Production knowledge is often tacit and embodied in people and capital goods that are slow to move across space. That is, innovations in how to do things often spread out very slowly. Imagine the first coffee shop to start doing counter instead of waited service. This is a 'process innovation' that turns out to save the coffee shop more money on staff and serve more customers. This idea would spread out slowly, and another coffee shop owner might only have the sense to try it out themselves if they were located nearby and could see the benefit and figure out how to do it themselves. Two, there are returns to having many people together. There's more social interaction in larger places: e.g., meeting people in cafes, bars, and at parties. A bad but useful model of social interaction in cities is one of a nuclear reactor, where social fission causes interpersonal collision, and the exchange and production of new ideas. Density heightens the reaction, and the intensity of social forces present in a particular place. We learn from our peers, and so more social interaction means the greater the chance of a new idea is born.

Input-output linkages refer to the fact that industries that co-locate and share inputs can economize on inputs by minimizing transportation costs between them.

There are other reasons cities exhibit increasing returns, as posited by Jane Jacobs in her books *The Death and Life of Great American Cities* and *The Economy of Cities*. Jacobs posits that the scale achieved in large markets leads to new combinatorial possibilities, like the ability to mix seeds together and form new crop varieties, or the ability to take an innovation in one industry and apply it in an other.

There is a circularity here. Agglomeration makes regions more attractive to move to *because of* their size – only making them bigger! This means that small accidents of history that shift the population can result in large, persistent effects. This is called *path dependence*.

In the paper 'Portage and Path Dependence' (Bleakley and Lin, 2012) that we will study in class, we see how a small historical accident led to the location of cities today. Historically, ships could only travel so far upstream before they had to unload (a process called portage). This process attracted business at portage sites, and the small increase in population at such places made them attractive, increasing the agglomeration forces until eventual such places were cities today. What constrains this process?

8.3 Congestion and spatial equilibrium

While the presence of increasing returns to scale suggests the possibility that everyone might want to just live in the same place, it's obvious from looking out at the world that we haven't collapsed into a black hole.

Instead, congestive forces balance out agglomeration, and the interplay between the two help explain the distribution of population across space.

A classic model of location choice is the *Rosen-Roback model*. The Rosen-Roback model is a model of a spatial equilibrium: it imagines that people are freely mobile across places. If people are freely mobile across places, and they choose where to live based on the utility they would receive by being in a particular place. If everyone has the same preferences, then in equilibrium, as people utility maximize, it must be that all places that people live in must give equal utility. This is a no-spatial-arbitrage condition.

Suppose regions are indexed by i, so $i = 1$ corresponds to New York, and $i = 2$ Omaha, and so on. The utility you get in a given place might be given by,

$$\underbrace{A_i}_{\text{amenities}} \times \underbrace{\frac{w_i}{P_i}}_{\text{wage in } i \text{ divided by the cost of living in } i} = \underbrace{U_i}_{\text{utility from living in place } i} = \underbrace{\bar{U}}_{\text{everywhere the same!}}.$$

Amenities might refer to features of a place like its good weather (first nature) or (dis-)amenities, like pollution (second nature). The real wage in each place depends on both the nominal wage and the local cost of living.

The spatial equilibrium concept is a good rough approximation of the world. Places with high wages and high amenities (like San Francisco) have high costs of living. Low-amenity places like Anchorage, Alaska, have high real wages.

Agglomeration forces can raise the local wage. What stops this from getting out of hand is that the local cost of living and dis-amenities can balance this out. Scarce land means that land prices get bid up very high in

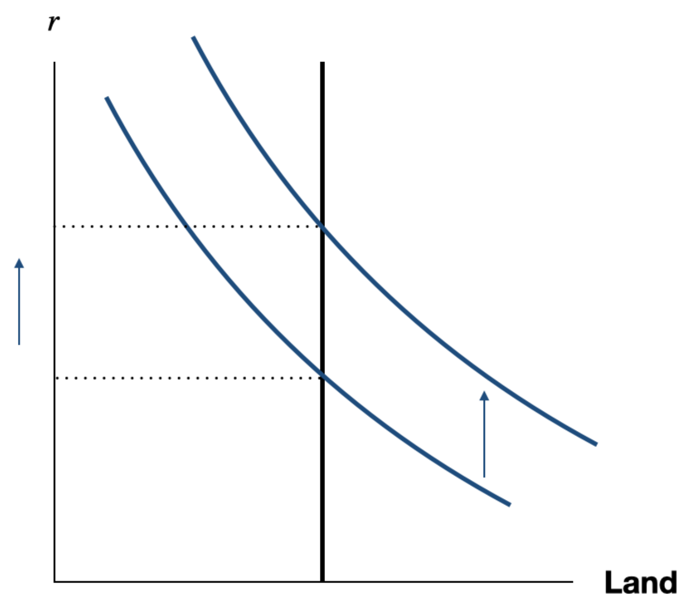


Figure 8.1: Supply and demand for a location. More demand for a location drives up the local cost of living by bidding up land prices

dense places (only some of this can be offset on the 'quantities margin:' by having smaller houses). Bigger places are louder, more polluted, and have more traffic. These increased costs to density temper the benefits of density.

8.4 Congestion forces and spatial equilibrium with the city

A classic congestion force is housing. To live somewhere, you must live in housing, which is built atop land. Land is in fixed supply, and totally immobile. This is what is displayed in Figure 8.1. More popular locations have more demand for housing, which bids up its price.

Another classic congestion force is commuting. Just like spatial equilibrium must hold for demand *across* locations, it must hold for demand for locations *within* them. The cost of housing may be higher in New York city, but there's no reason you need to live in Manhattan. Can't you benefit from the employment and amenity opportunities offered by Manhattan, but pay half the rent by living in Jersey City? Why should I pay Hyde Park rent, when instead I could live in South Shore? The answer, of course, is that there are costs to

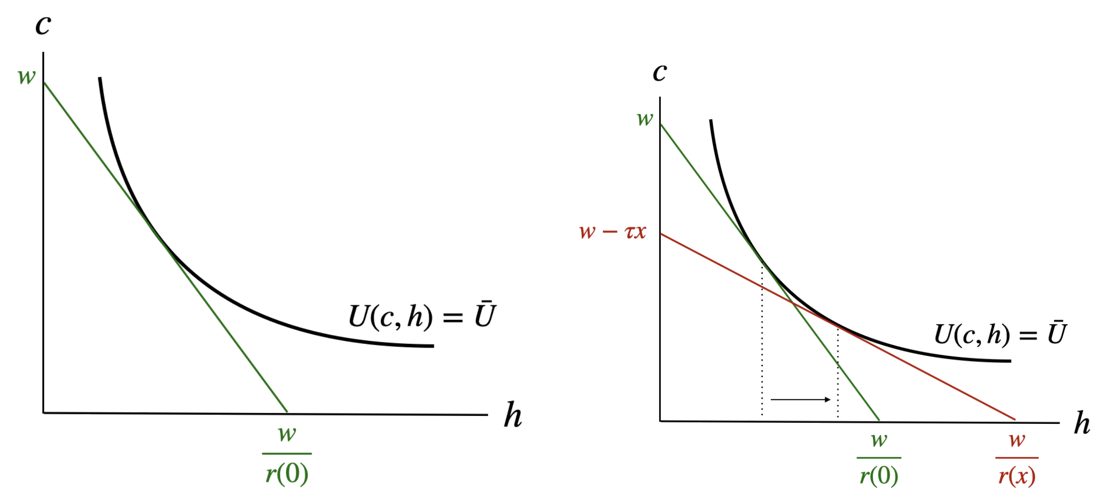


Figure 8.2: Left: Budget constraint at the CBD. Right: Utility maximization at $x = 0$ and x.

living far away from where you work and consume.

Another classic model of location is the Alonso-Muth-Mills model of a city.

The model setup is as follows. There is a downtown of a city, the 'central business district' (CBD – no, not like that), where people work and shop. They can choose to live at any distance from the CBD, denoted x. This means that the CBD is located at $x = 0$. However, if they choose to live x miles away from the CBD, they must pay τx dollars in gas and transport costs (yet alone Uber Eats fees).

When they choose to live at x, they consume $h(x)$ units of housing. The price of housing is $r(x)$. Everything else they consume – all consumption goods, c – do not depend on their location, and the price of these will be 1. They are produced in the CBD. One dollar gets you one unit of non-housing consumption c.

Households maximize their utility, $U(c, h)$ subject to the budget constraint $c + r(x)h(x) = w - \tau x$. Their budget constraint varies with x – it *depends on where they choose to live*!

In a spatial equilibrium, utility is everywhere the same. Let's call this level of utility $\bar{U}$, and suppose it is known. Consider someone who lives in the downtown, $x = 0$. Their budget constraint is simply, $c + r(0)h(0) = w$. The price of housing must adjust to achieve spatial equilibrium, so household utility maximization looks like the left panel of Figure 8.2. In the right panel, we move farther from the downtown. A household that wants to live x miles away will have to pay more in commuting, so their income is $w - \tau x$.

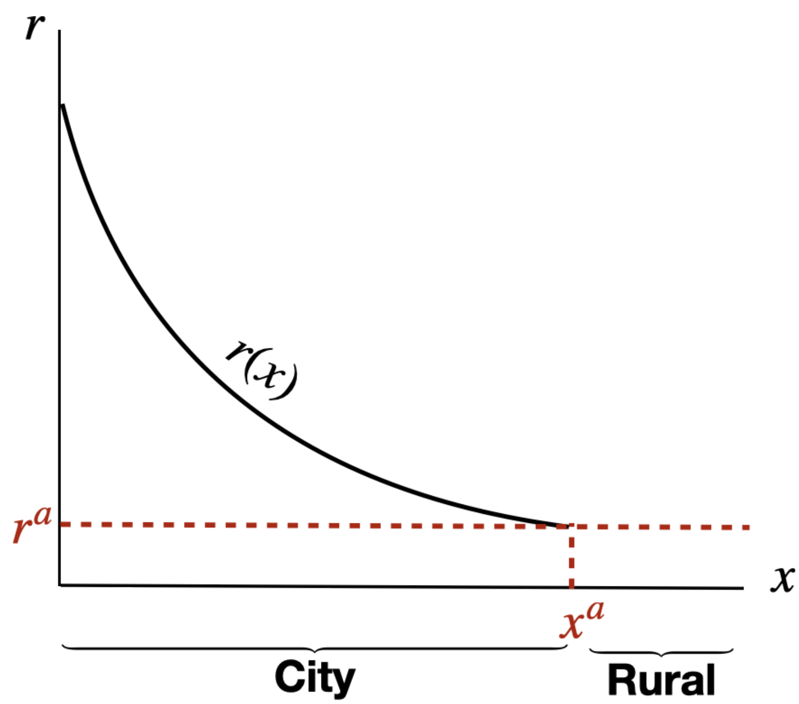


Figure 8.3: The division of land in the Alonso-Muth-Mills framework.

However, to maintain spatial equilibrium, housing supply *must adjust* to compensate them for living far away. The price of housing must *fall*! As it falls, there is a substitution effect: they will want to consume more housing

This looks a lot like the structure of cities. Households far away from the downtown pay less in rent, and have bigger houses, too. A high price of housing and limited land incentivizes construction firms in the center to build up, explaining skyscrapers and densely populated downtowns.

When does the city end? As you move farther and farther from the city, to incentivize someone to still work and shop downtown, the rent has to fall dramatically. Eventually, the amount a landlord could make leasing his land for housing will fall below the opportunity cost of that land: using it for, say, farming. The rental rate of land in farming is given by r^a. We will call the fringe of the city x^a. At $r(x^a) = r^a$, a landlord is indifferent to giving up his land for urban use or farming. As shown in Figure 8.3, As we move even farther away, $x > x^a$, all land will be farming or have the housing density equal to that at the fringe of the city.

Lecture 9 – Industrial policy

A common motivation for protectionist trade policy is the *Infant Industry argument*.[1]

The argument is that when an economy is open to trade, domestic industry that competes with imports doesn't have a chance to develop: competition with foreign imports keeps the industry in its 'infancy.' There is a lot of logic to this argument: if industries have *economies of scale*, then they benefit from increases in size. Industries that have to compete with foreign imports don't have the opportunity to grow, lower costs, and become competitive.

Looking out at the world, most rich nations export manufacturing goods, while poorer nations tend to export agricultural goods. Could trade not be a source of gains for these nations, but rather the cause of their poverty – that they have lacked the opportunity to develop their 'growth' industries? Marxist 'World Systems' scholars like Immanuel Wallerstein contend this is so: they divide the world into 'Core' and 'Periphery' nations. The 'Core' imports raw materials from the 'Periphery' to develop manufacturing goods. Peripheral nations depend on the core through trade, and lack the opportunity to grow.

Infant industry arguments, national security concerns, and anti-globalization sentiments from the political right and left all motivate protectionist trade policy.

Such policy rationale depends on stories that are *dynamic*: they are about growth, i.e., about short- versus long-run! The essence of these theories argues that when there are scale economies, comparative advantage is something you achieve through growth. To understand it, we'll need to develop some theory on short- versus long-run costs.

9.1 Producer theory – short versus long run costs

As we have developed, industry supply curves capture the *marginal cost* of production. When there are fixed factors, like in the Specific Factors Model, each additional unit requires more and more labor to produce,

[1] The infant industry argument is often attributed to Alexander Hamilton's *Report on Manufacturers*, but as some have pointed out, (here) Hamilton didn't really support protectionist policy, but rather using tariffs to raise money to fund subsidies.

due to diminishing returns, and thus is more costly. That is why the marginal cost curve slopes up. The marginal cost curve is the supply curve, because producers profit maximize when they set price = marginal cost. Therefore, the marginal cost curve tells us at a given price, how much output a firm would want to produce such that they maximize profits.

However, to think about short- versus long-run costs, we need to define some new objects.

$$\begin{aligned} C(X) &= \text{total cost of producing } X \text{ units} \\ MC(X) &= \text{marginal cost of the } X\text{th unit, the supply curve} \\ AC(X) &= \frac{C(X)}{X} = \text{the average cost of producing } X \text{units} \end{aligned}$$

The cost function is composed of two parts,

$$C(X) = \underbrace{VC(X)}_{\text{variable costs}} + \underbrace{FC}_{\text{fixed costs}}$$

Let's think about all these objects. The marginal cost, or supply curve, captures the cost of each additional unit, and therefore doesn't depend on fixed costs. Marginal cost could equivalently be thought of as 'marginal variable cost.'

Because fixed costs are fixed, their effect on average costs falls as output increases. However, because of the presence of fixed costs, for the first unit produced, $AC(1) > MC(1)$, so the AC curve must lay above the MC curve. As each additional unit is increasingly expensive – the MC curve slopes up. However, as long as the AC curve is above the MC curve, it has to slope down because of the falling contribution of fixed costs outweighs the rising variable costs. Once the MC curve crosses the AC curve, each additional unit starts to bring the average up, and so the average cost curve must start to slope up. This is displayed in the left panel of Figure 9.1.

We will assume that product markets are competitive – that there are no monopolists, and firms are price-takers. Per-unit profit is the difference between the price a good is sold at in competitive markets (its marginal

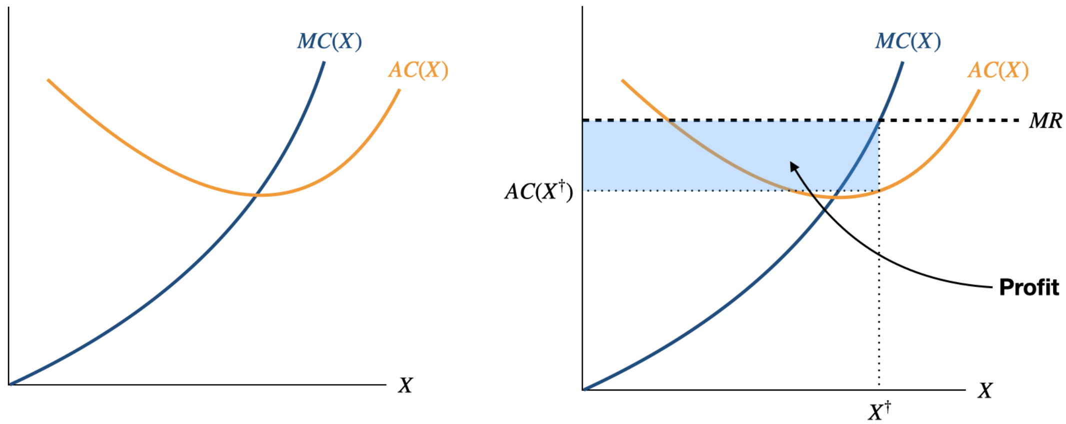


Figure 9.1: Left: Short-run average and marginal cost curves. Right: Addition of the MR curve (which is just the price, as firms are price-takers) and the profit rectangle.

revenue to the firm, MR) and its average cost. Profit multiplies this by output,

$$\text{Profit} = \underbrace{(P - AC(X))}_{\text{per-unit profit}} \times X.$$

Profit is maximized when firms produce at the level where marginal revenue equals marginal cost ($MR = MC$). Therefore, we can see profit graphically as the area of a rectangle with a base whose length is X and whose height is $P - AC(X)$. This rectangle is displayed in the right panel of Figure 9.1.

The cost curves described are short-run: they keep the level of technology fixed. In a standard microeconomics course, you might assume that these are short-run cost curves because they keep capital fixed. We'll conceptualize this more broadly than physical capital. In particular, we will conceptualize scale economies in two ways: one, static trade offs between fixed and variable costs; and two, from external economies of scale and from dynamic scale economies like learning-by-doing.

As output increases, firms can take advantage of new scale economies, and the average cost curve can shift down and out, as shown in Figure 9.2. This means the marginal cost will fall, too, and thus the industry supply curve shifts right.

In the long run, we can think of the 'Long-run average cost curve' ($LRAC(X)$) as the lower envelope of all the short-run average cost curves associated with each level of scale. What I mean by 'lower envelope' is

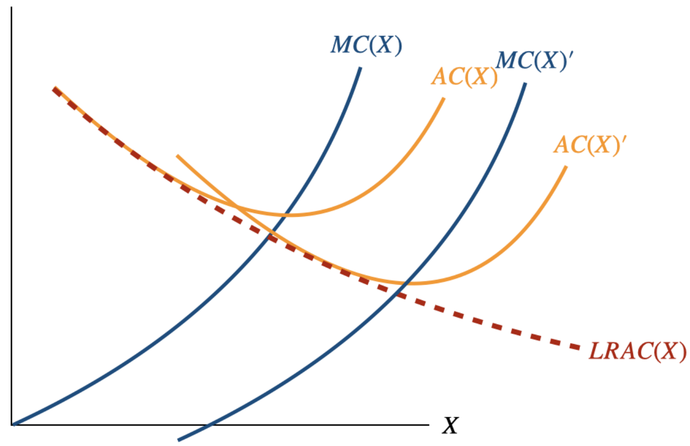


Figure 9.2: The long run average cost curve.

the following: suppose you drew average cost curves for every level of scale. Then, for each level of output, you found the minimum cost that it was possible to produce that at by looking across all the average cost curves, and marked that. If you did this for each level of output, you'd find the long-run average cost curve. Two things are of note: first, in the short-run, the minimum of the average cost curve is the most efficient way to produce a good. Second, each point on the long-run average cost curve does not correspond to a minimum of the short-run average cost curve. In fact, the minimum of any short-run average cost curve will always be *above* the long-run average cost curve! This is because if you are operating at a short-run minimum producing X goods, there is a cheaper way to produce the same quantity: be on a different, better short-run average cost curve that's associated with bigger scale economies.

Static scale economies may be internal to the firm, meaning they are not external, like agglomeration forces, but come from choices firms make. Firms may face a trade-off between fixed and variable costs. In some industries, a firm can pay a large fixed cost in order to have lower variable costs. For example, a firm can scale up production by building a larger factory, investing in R&D, and so on. These investments result in lower variable costs per unit, and more variable profit per unit. However, this requires a market large enough to make the additional profit-per-unit large enough to cover the fixed investment costs.

This can operate at the industry level too. As industries increase in size, there are *external economies of scale*: i.e., scale economies beyond those internal to the firm (like fixed costs). Larger markets have thicker labor markets (labor market pooling), can benefit of the diffusion of ideas across firms (knowledge diffusion) and can share inputs and minimize transportation costs between firms (input-output linkages). These are classic Marshallian agglomeration forces. There could also be dynamic forces, like *learning-by-doing*: the more a firm produces something, the better it gets at producing it, driving down costs. Firms can't take these scale economies into account when making production choices. They don't get to choose how much they learn from other firms nearby, for example. These are externalities, which means that there is scope for the government to subsidize these industries and increase welfare.

9.2 Promoting 'resilience'

A large portion of current global industrial policy promotes 'resilience,' or the idea that countries need to invest in the capacity to handle out-of-the-ordinary situations. Coming out of Covid, where supply chain disruptions and congestion and ports stymied global trade, this makes a lot of sense.

How can we think about 'resilience' in our model? Well, suppose there are two states of the world: business-as-usual, and a crisis, where suddenly production of a good needs to scale up. In the case of Covid, this might be the production of vaccines, or just textiles and other consumer goods demanded by people shopping from home.

With current technology, it made sense for business to achieve scale economies, and locate all production in one location, for example. This production structure means that during business-as-usual, the average cost of production is quite low. However, it becomes difficult to scale up production in the short run. An alternative production structure might be to diversify production across multiple locations. This raises average costs, but makes scaling up in the short run potentially less costly.

These two potential technologies are graphed in Figure 9.4. Also graphed is the amount of X the industry may need to produce in the business-as-usual or shock scenario.

The *expected* cost of production – averaging across the business-as-usual and crisis scenarios – must be an

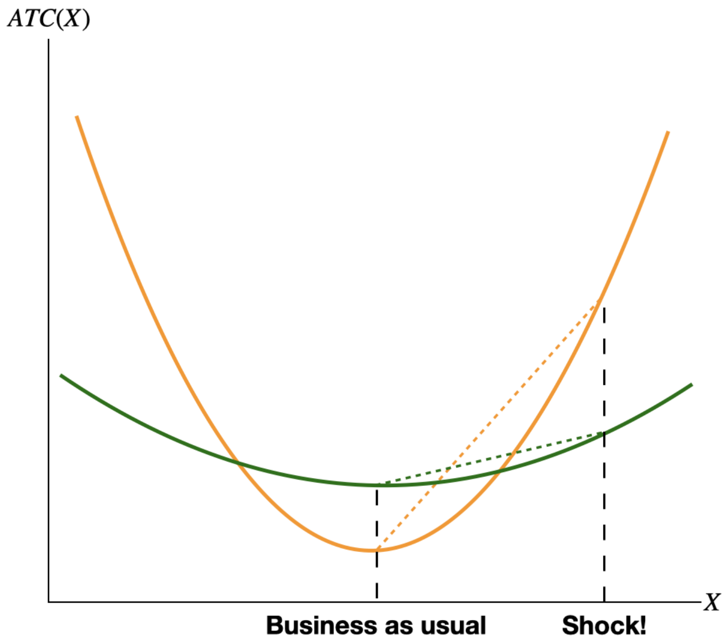


Figure 9.3: Two states of the world – which cost curve should we be on?

average of the two points on the average cost curve:

$$\text{expected average cost of production} = \text{probability of business as usual} \times ATC(X_0) + \text{probability of shock} \times ATC(X_1)$$

where X_0 is production during the business-as-usual scenario, and X_1 is production when there is a shock. This average must lie on the line that connects the two points on the ATC curves in either scenario.

Despite the cost structure of production being worse in business-as-usual on the flatter ATC curve, the expected average cost of production could be lower, if the shock happens with any regular occurrence.

Industrial policy aimed at changing the structure of industry – getting firms to diversify their supplier mix, operate in multiple locations, and so on – promotes 'resilience' by trading off costs today with costs in a crisis scenario. Such 'resilience' strategies are effectively a form of *insurance* against bad states of the world that may happen in the future.

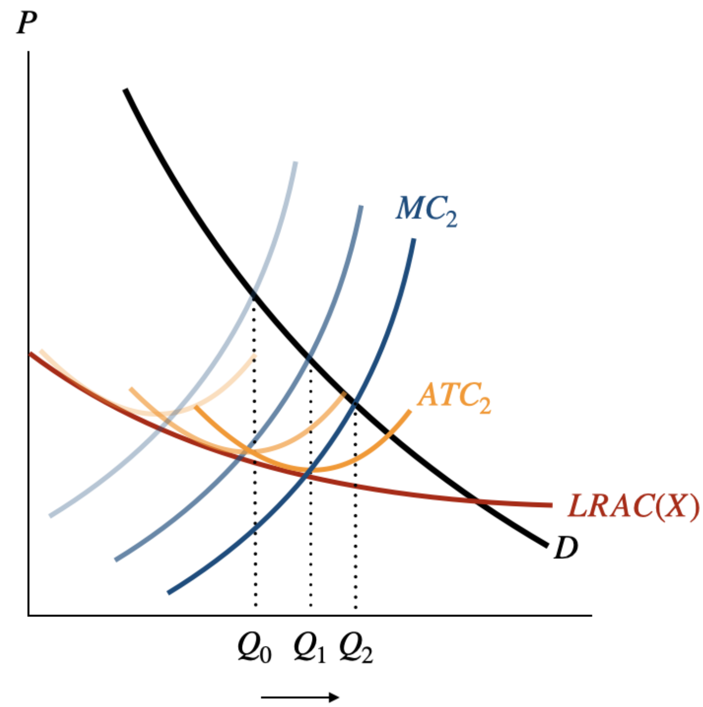


Figure 9.4: The long run average cost curve from a sequence of investments over time.

9.3 Forward falling supply curves

In the long run, an industry with scale economies has a *forward-falling* supply curve given by its long-run average cost curve.

To get intuition on why this ought to be thought of as the long run supply curve, we will consider how this operates in an industry making a trade-off between high fixed costs and low variable costs. Consider the following sequence of events to see what the long-run average cost curve is the long run supply curve. Suppose at time 0, firms produce until price = marginal cost, and firms produce some quantity, Q_0. Profit for producing Q_0 can be given by a rectangle with height between the equilibrium price and the short-run average cost curve, and a base Q_0. In the next year, all contracts need to be resigned, and the size of factories can be renegotiated, plants can be moved, and new physical capital can be installed. Taking the current price as given, to meet demand Q_0, it would be more profitable for the firm to operate with a different short-run average cost curve. In particular, they would be willing to pay a high fixed cost by investing in capital and research to lower their marginal cost and generate more profit. In particular, they will choose their technology, or scale, so as to operate on a short-run average cost curve whose minimum is at Q_0, as this would maximize profits when output is Q_0.

With the new technology in hand, they can actually produce more for less. Consumers would want in on this deal, and so in equilibrium, the firm scales up production until their new short-run supply curve (the marginal cost curve) intersects demand. That is, they simply profit maximize until marginal cost = marginal revenue. With the new technology, it would also be more profitable to expand production to Q_1, then, and meet demand. This process repeats until the long-run average cost curve intersects demand. This is the long-run equilibrium outcome!

9.3.1 External scale economies and dynamic scale economics

Another explanation for forward-falling supply curves is external scale economies.

To generate a forward-falling supply curve, all you need is that the average costs faced by firms falls as output rises. The reason for this does not need to be internal to the firm, as described above.

One set of external scale economies are agglomeration forces discussed in Lecture 8. As the size of an industry increases, firms can capitalize on nonexcludable inputs like knowledge, which spills over (i.e., diffuses) across firms, and thick (i.e., pooled) labor markets that allow workers and firms to find better employment relationships.

Another reason why costs may fall as output increases are dynamic scale economies. These arise when there is learning-by-doing in the production process. That is, when workers get better at their tasks the more they do them. This drives the per-unit and therefore average cost of production down over time. This means that average costs fall in cumulative output, total output in a given time period. This generates a 'learning curve' that relates costs and cumulative output. We will not model dynamics explicitly in this class, but this may be another reason that in the long run, costs fall with output.

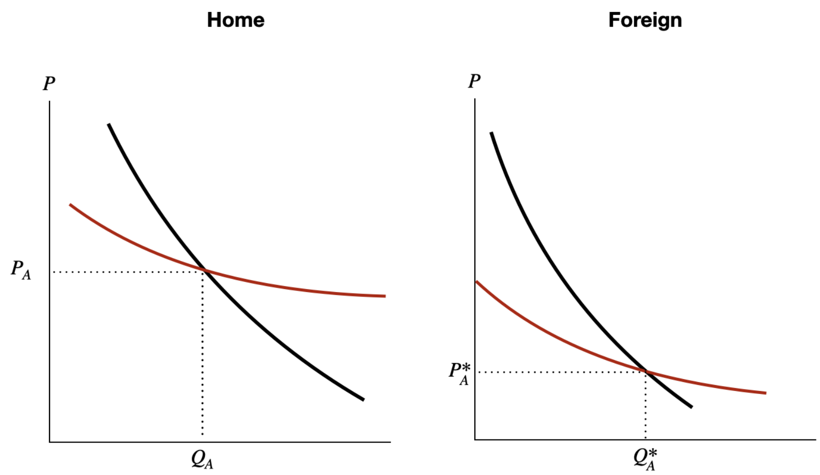


Figure 9.5: Autarky with forward-falling supply curves.

9.4 Trade liberalization with external scale economies

Now we will consider trade liberalization when there are forward-falling supply curves. This example comes from the Krugman, Obstfeld, and Melitz (2017) textbook.

Suppose there are two countries, Home and Foreign, who each can produce buttons with a production technology that features scale economies. At first, these industries are non-traded, and the supply must equal demand domestically in each country. Then, a trade deal is signed, which allows consumers in each country to buy buttons from wherever is cheapest. What will the trading equilibrium look like?

First, let's consider the autarky supply and demand curves as we did in Lecture 7. They now look like Figure 9.5. Supply curves now slope down, but otherwise this ought to look standard: autarky equilibrium is where supply equals demand.

In the figure, Home's autarky price is higher than Foreign's, so we aught to suspect that Home should import, while Foreign should export. It's the same arbitrage opportunity available to consumers in all the other models we studied so far, too. However, as Home imports from Foreign, Foreign's price *falls* because their production expands to meet both Foreign and Home demand. The equilibrium must be that Foreign does all production of buttons in the economy. Opening up to trade makes Home's button industry evaporate overnight. All production moves abroad!

This is shown in the panel of Figure 9.6. World demand is the sum of Foreign and Home demand. World supply is best achieved by concentrating production in one location and achieving scale economies, and only happens in Foreign.

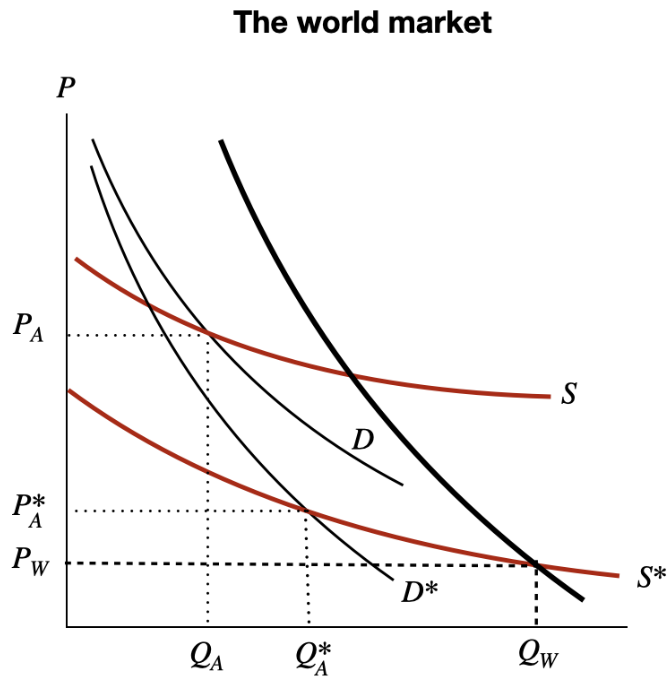


Figure 9.6: The world market for buttons

This means, like in the Ricardian model, there is complete specialization. The motivation, however, is very different. In the Ricardian model, there was complete specialization because the trade-off between goods was constant, within each country, no matter the scale of production. Countries simply specialized in their comparative advantage industries to maximize global welfare. Here, things are very different. Production in this model has *increasing* returns to scale, so a country's 'comparative advantage' in an industry rises the more of it they do. The need to concentrate production in one location is fairly economic and intuitive in this framework.

The rewards for concentrating production to the global economy are also quite clear: as output increases, the price *falls* – for both countries! We move down both countries' demand curves. There are huge increases to consumer surplus in both countries as a result of trade, and huge increases in producer surplus in Foreign, too. However, producers at Home might not be happy with this result: their producer surplus goes to zero as all production moves abroad.

9.4.1 Established advantage

This model creates *path dependencies* where history matters in determining who does what, and the location of production, just like how agglomeration forces (another increasing returns phenomenon) made cities remain located in places despite now-obsolete initial cost advantages. That initial cost advantages can lock-in production in one place is what we call 'established advantage' in this framework.

To understand this better, let's imagine another country enters the global marketplace. In Krugman, Obstfeld, and Melitz (2017), this country is imagined to be Vietnam. Let's suppose Vietnam's production technology is better than Foreign's. They have workers more skilled at button production, the land to build button factories on is cheaper, and maybe they have inventors that have come up with new labor-saving devices in the production of buttons. Overall, the forward-falling supply curve is below Foreign's.

However, because Foreign has established production first, they have achieved scale economies, and the world price is already low. Vietnam *could* produce all buttons for the world more cheaply, but its industry is in its 'infancy,' and the cost of the first button it produces is way higher than the world price. Their industry cannot 'mature' because the fixed costs to start production are too large. It is unprofitable to start production, even though if all production could relocate to Vietnam, there would be even larger gains to trade.

This is imagined in Figure 9.7. Vietnam's supply curve is below Foreign's, but the initial price Vietnamese producers would need to receive to make the first unit is below the world price, as Foreign has already achieved scale economies.

What would be best for Vietnamese producers, and the world? If the Vietnamese button production could be allowed to achieve scale economies!

How might this happen? How can Vietnam allow its button industry to take off? One answer is to go into total autarky. Another is to use tariffs. Either way, protectionist policy that raises the domestic price and protects domestic producers allows domestic production to scale up. Once the 'infant industry' has matured, the global economy can relocate all production to Vietnam.

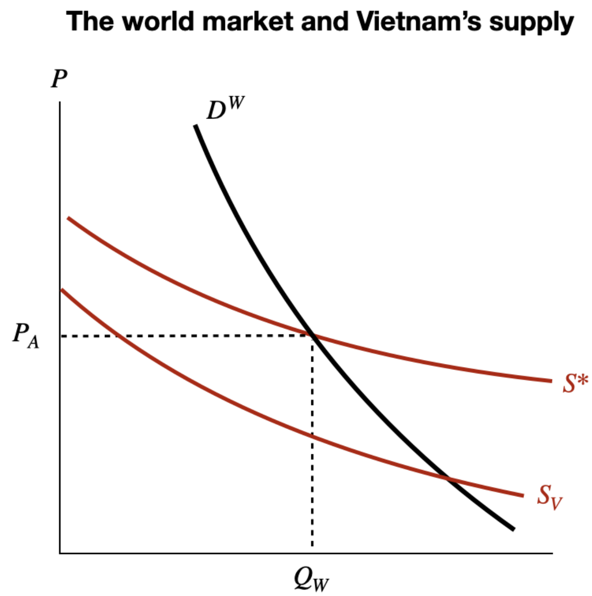


Figure 9.7: Established advantage keeps Vietnam out of the market

Of course, producers in Foreign would not like it if Vietnam did this. One, they would not get access to their markets, and not get to achieve the scale and profits they could otherwise. Two, if Vietnamese industrial policy is successful, it would steal all the business away from Foreign.

9.5 Tariffs with dynamic scale economies

The last example was a bit extreme, and all the production took place in the long run. Let's consider a shorter-run example, where there is learning-by-doing, meaning that as an industry's cumulative output increases, its cost fall, and its supply curve shifts right.

In this world, a policy-maker might still have a good rational to pursue protectionist policy. To see this, we'll start from a short-run free-trade equilibrium where the domestic import industry has learning-by-doing.

We'll suppose the technological know-how is fixed in the short-run and is described by the MC curve. Knowledge is a type of capital you can accumulate, but unlike physical capital, it's extremely hard to buy and sell in markets. It is often acquired through practice, which takes time. The short-run supply curve is the marginal cost curve, and holds the level of knowledge fixed.

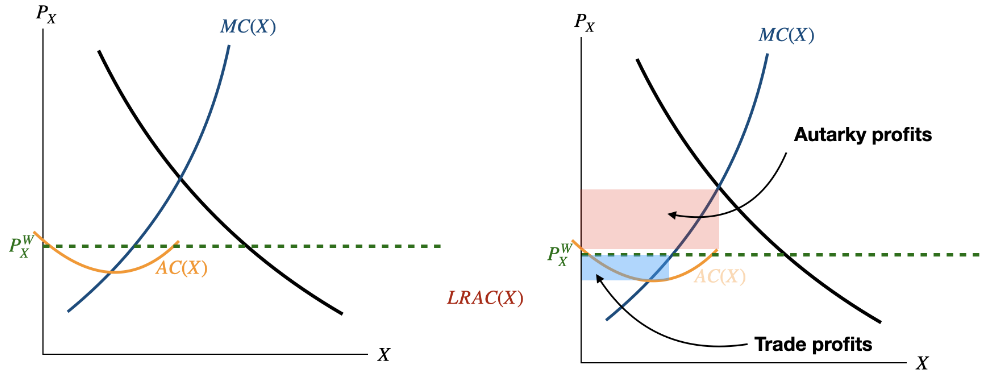


Figure 9.8: Left: A free trade equilibrium for an import industry with a downward-sloping LRAC. Right: Same equilibrium with profit rectangles shaded in.

When a firm is in autarky, profits and output are high, and they can accumulate a lot of knowledge capital. These profits and benefits from experience are eroded when trading at the world price P^W.

This baseline corresponds to Figure 9.8. Everything is the same as before, but I've added to the figure the short- and long-run average cost curves. The marginal cost, or supply curve, intersects the short-run $AC(X)$ curve at its minimum, which by assumption of a long run equilibrium is the world price, P_X^W. The short-run AC curve is tangent to the $LRAC$ curve, which slopes down due to increasing returns. The $LRAC$ falls below the minimum of the AC curve for the reasons discussed in the last section.

Consider what happens when an import tariff T is introduced. Imports now cost more, so in the short-run, we move up the marginal cost curve. Domestic production can better compete with imports, and produces more output. As there is learning-by-doing, as the domestic industry increases output, cumulative production increases, and with it, the tacit production knowledge or 'experience capital' that makes workers and capital more productive.

When the tariff is removed, Home imports less than they were at baseline, because the domestic industry expanded and became more competitive! All of these effects are visible in Figure 9.9.

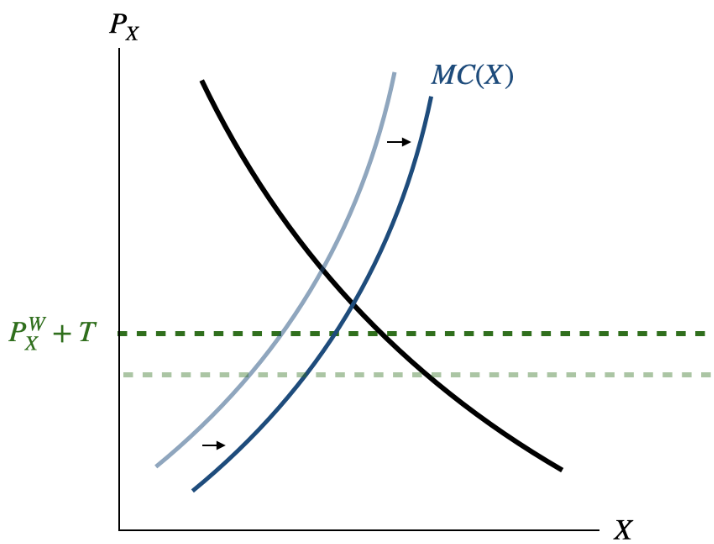


Figure 9.9: The effect of a tariff with long-run increasing returns to scale

Lecture 10 – Intra-industry trade and monopolistic competition

So far, this has been a course primarily about supply. We've tried to answer, when countries or regions trade, who produces what, with what technology, and why? Understanding the technology of production allowed us to analyze factor markets: the markets for labor and capital within each country.

These models have proved to be extremely useful! We came to understand that differences in technology and factor endowments led to gains from trade. Countries' consumption and production baskets differed, and the difference was made up in international markets. We even developed a theory of *why* different regions differed in factor endowments and technology when we studied economic geography. When we allowed those same increasing returns to exist at the industry level, trade theory showed us how economies of scale could lock-in production in different places ('established advantage') and why countries might be motivated to shield domestic industry from international competition.

However, this theory has been lacking on some serious and quantitatively relevant dimensions. Namely, we've reduced our analysis to talking about two goods and two industries producing them, without really clarifying what defines a product, or which firms produce them. That is, we've thrown away the notions of *product variety* and *market structure*. These are obviously first-order issues that trade theory ought to tackle. After all, isn't one gain to trade the ability to consume *new goods*?

It was not until after the 16th century that Europe got tomatoes through the Columbian Exchange. Venetian trade with the Ottoman Empire brought coffee into Europe. Following Japan's opening to trade, the Dutch brought European military technology into Japan.

There are 'new' goods are traded even when looking *within* broad industry classes. Today, through trade, consumers outside the U.S. gain access to American automobiles from companies like Ford, Buick, and GM, while U.S. consumers get to benefit, through trade, by importing British Aston Martins, German Audis, Italian Ferraris, Korean Hyundais, Japanese Mazdas, French Renaults, and Swedish Volvos. Each of these countries is both an importer and exporter of cars. What they trade are different brands, or 'varieties.' The trade theory we've developed so far is completely inadequate in describing this kind of *intra*-industry trade in products. Moreover, it fails to describe the effects of trade on the companies that produce these varied goods.

In this lecture, we will study the 'Krugman' model of intra-industry trade and monopolistic competition. This will allow us to think about *product differentiation* undertaken by different firms, and how these firms operate in the global economy.

Paul Krugman won the Nobel Prize in 2008 for his work in international trade and economic geography. Krugman's main contribution to economic theory is combining theories of monopolistic competition (among other things) with models of trade and economic geography to study intra-industry trade, regional specialization, and economic development. In this lecture, we will study a much-simplified version of the Krugman model of intra-industry trade (taken from his own textbook).

In studying this model, we will put on the table two new gains from trade: the *love-of-variety* effect, and the *pro-competitive* effect. The first describes how trade makes available new varieties of goods, and how a country can both import and export within the same industry. The latter refers to how opening to trade can *toughen competition* and lower prices for consumers at home.

To develop this model, we need to first study the mechanics of how a monopoly works. Then, we will take the theory of monopoly, and use it to study how firms *imperfectly* compete with one another in a *monopolistically competitive* marketplace. Then, we will open up this model to trade, to study how these new gains from trade work.

To use this model, it's best we develop the ideas with a little bit of algebra, which will help simplify and clarify all of the forces at work.

10.1 Why we need monopoly with scale economies

This model will rely on *internal economies of scale*. In particular, we will study firms that pay a fixed cost in addition to a marginal cost to produce their output. The presence of fixed costs means there are economies of scale in production: the average cost falls in the quantity produced. These scale economies are are internal to the firm, meaning that firms understand their average cost curves slope down. Internal scale economies are not the sort of scale economies that are at the core of agglomeration (which occur when many producers co-locate); those would be called 'external scale economies.'

If firms pay a fixed cost to operate and therefore operate with scale economies, price cannot equal marginal cost. If the marketplace for goods produced with this type of increasing-returns technology was competitive, forcing firms to price at marginal cost, the industry wouldn't exist at all. The reason is that profit would always be negative because of the fixed costs. No firm would choose to produce in such an environment. What is needed to sustain this industry is for firms to have some monopoly power and the ability to raise price above marginal cost, to ensure non-negative profit.

To see this, let's define the cost function for firms in this model.

$$C(X) = \underbrace{MC}_{\text{marginal cost}} \times X + \underbrace{F}_{\text{fixed cost}}$$

This implies that average costs are, $AC(X) = MC + F/X$, which falls as output increases. Recall that profit can be written as,

$$\text{Profit} = (P - AC(X)) \times X$$

If firms priced competitively, so that $P = MC$, then what would happen? If this occurred, profit would be,

$$\text{Profit} = (MC - \underbrace{(MC + F/X)}_{AC(X)}) \times X$$

Rearranging this,

$$\text{Profit} = -F < 0.$$

If this were true, no firm would choose to operate! Therefore, if the industry has fixed operating costs and constant marginal costs, then prices must be above marginal cost. For firms to price above marginal cost, they must have monopoly power.

10.2 Monopoly

When firms operate in a competitive environment and are 'price takers', the marginal revenue (MR) per unit produced is simply the price. Firms maximize profits when setting MR equal to MC, or marginal cost, and thus in equilibrium, $P = MR = MC$; price is marginal cost under perfect competition.

A monopolist is a firm that is the sole producer of their good. They face no competition to produce their specific variety, and they cannot be thought of as 'price takers.' Instead, under monopoly, we assume the firm faces a downward-sloping demand curve, D. What this means is that the firm knows that it can only sell more goods if the market price falls.

What is marginal revenue when firms face a downward-sloping demand curve? Marginal revenue is the additional unit revenue the firm gains from selling one more unit. For consumers to buy one more unit, the monopolist knows that the price must fall. Therefore, the marginal revenue curve must be falling in price. We can describe this curve even more. For a given output X, the marginal revenue of the Xth unit, $MR(X)$ must be below the price consumers would be willing to pay for X units. This is because to get consumers to buy one more unit, the price must fall not only for the additional unit, but also all the units sold previously. The monopolist only gets to charge one price! This reduction in price by increasing sales by one more unit means the revenue earned on all previous X units sold falls, too. This means the $MR(X)$ is steeper than the demand curve.

We also know that the difference between marginal revenue and price (i.e., demand) must depend on the slope of the demand curve. When the demand curve is flat (*elastic*) the monopolist can sell an additional unit with a very small cut in price. When the demand curve is steep (*inelastic*), then the monopolist must drop the price by a lot so that consumers will buy one more unit of their good. Therefore, the difference between the MR curve and demand will be large when demand is inelastic. This is demonstrated in Figure 10.1.

Let's now describe the demand curve. We will write,

$$X^D = X_0 - \sigma P.$$

In this representation, X^D is the X demanded, X_0 is the maximum amount of X consumer will demand

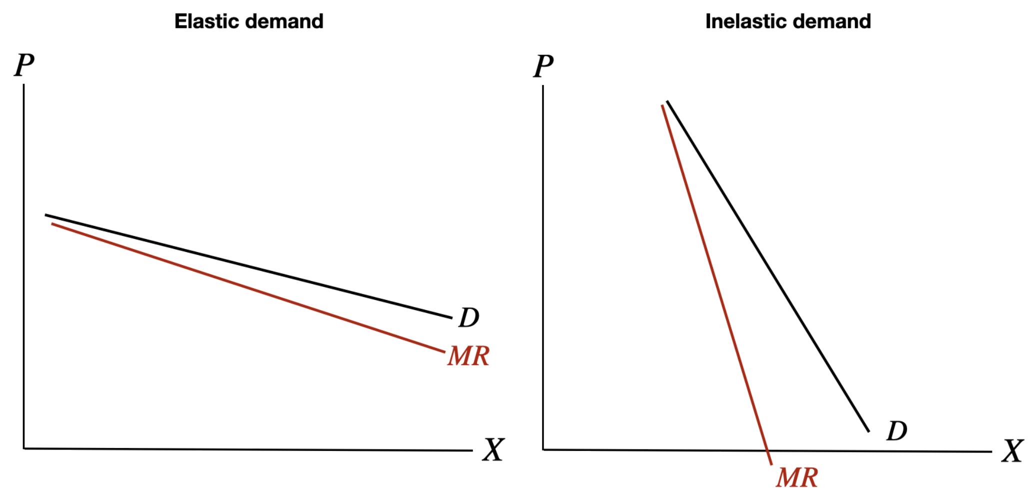


Figure 10.1: Demand and marginal revenue when demand is elastic, and inelastic.

(this will be irrelevant) and σ will describe the slope of the demand curve (this will be important). We can also describe the relationship between price and quantity demanded by inverting the demand curve,

$$P(X^D) = \frac{1}{\sigma}(X_0 - X^D)$$

The revenue earned by selling X units is,

$$R(X) = P(X) \times X = \frac{1}{\sigma}(X_0 - X)X$$

Therefore, the marginal revenue a firm would earn by selling one more unit is,

$$MR(X) = \Delta R(X) = \underbrace{P(X) \times \Delta X}_{\text{revenue from selling an additional unit}} + \underbrace{\Delta P(X) \times X}_{\text{change in revenue by lowering price on all previous units}}$$

The change in price by selling one more unit, $\Delta P(X)$ is $-1/\sigma$, and we are studying a marginal change in X, so $\Delta X = 1$. Plugging this in,

$$MR(X) = P - X/\sigma.$$

This describes all the economic logic before. The marginal revenue curve must be below the demand curve (hence is subtracted from price). The steeper the demand curve ($\sigma \downarrow$), the larger the gap $X/\sigma \uparrow$.

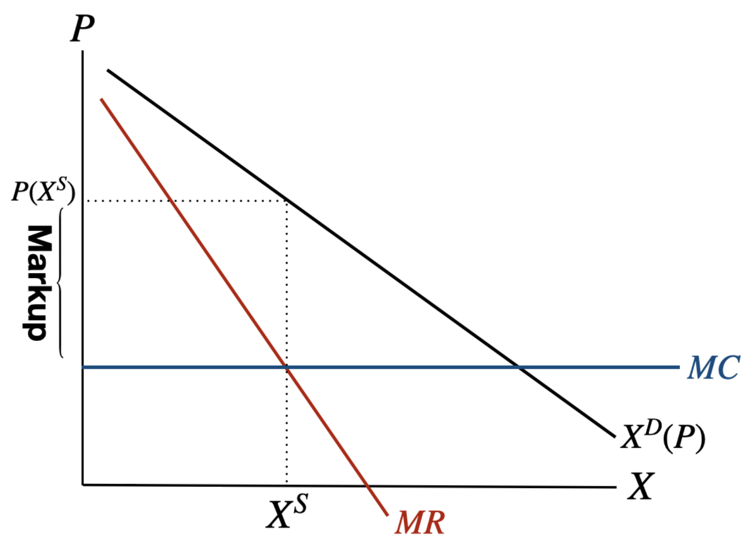


Figure 10.2: Monopoly pricing

Now, let's recall our monopolist's cost structure,

$$C(X) = MC \times X + F.$$

A monopolist that sets $MR = MC$ will charge a price above marginal cost. In particular the monopolist will choose to produce at the combination,

$$\begin{aligned} MC &= MR \\ \text{definition of MR,} \quad &= P - X/\sigma \\ \text{definition of } P \text{ from demand} \quad &= \frac{1}{\sigma}(X_0 - X) - X/\sigma \end{aligned}$$

Rearranging the above gives us,

$$X^S = \frac{1}{2}X_0 - \frac{\sigma}{2}MC.$$

this is the monopolists' profit-maximizing supply choice of X^S. They picked how much to produce by setting marginal revenue = marginal costs, knowing they faced a downward sloping demand curve. They under-provided (relatively to a competitive equilibrium) to keep prices high. That's what monopolists do! They restrict the quantity supplied in order to maintain high prices.

Using their profit-maximizing supply choice X^S, we can solve for the price,

$$\begin{aligned}
P(X^S) &= \frac{1}{\sigma}(X_0 - X^D) \\
&= \frac{1}{\sigma}(X_0 - (\frac{1}{2}X_0 - \frac{\sigma}{2}MC)) \\
&= \frac{1}{\sigma}(X_0 - \frac{1}{2}X_0 + \frac{\sigma}{2}MC) \\
&= MC + \underbrace{\frac{1}{2}(X_0/\sigma - MC)}_{\text{markup}}
\end{aligned}$$

What to clean from this is that price is above marginal cost. We call the difference between price and marginal cost the *markup*. The markup increases as the demand curve gets steeper $\sigma \downarrow$.

All of this can be demonstrated graphically! In Figure 10.2, I show demand X^D, marginal revenue MR, a constant marginal cost curve, MC, and the monopolist's optimal choice , X^S, as well as the markup over marginal cost, $P(X^S) - MC$.

10.3 Monopolistic competition

We will now use our model of monopoly to think about the setting of *monopolistic competition*. This arises when each good in the marketplace is owned by a unique monopolist, but goods are close substitutes.

In particular, we will assume that there are many goods in the marketplace, so that firms do not behave strategically, like under oligopoly. Firms will not *internalize* how their choices affect their competitors. Instead, the choices firms make only depend on those their competitors make in equilibrium.

This is a model of *product differentiation*: each firm owns a unique variety, and demand depends on how substitutable other goods are. Many markets behave this way: going to the grocery store, one is confronted with tons of varieties within the same type of good. For example, yogurts are differentiated on flavor, milk source, sugar content, branding, size, and so on. Yogurts are moderately substitutable: if Trader Joe's doubled the price of Icelandic Skyr, I would switch to Greek-style yogurt and not be too unhappy. However, there are many other yogurts available, too. The Skyr brand does not optimally price knowing how every possible brand would react to a change in their prices. There are too many competitors, and the effects are too small for Skyr to care. Instead, the yogurt industry, alongside craft beer, denim, toothpaste, socks, economics textbooks, and more, can all be described as monopolistically competitive.

10.3.1 Demand under monopolistic competition

I will now index goods by i. X_i^D describes the demand for the ith variety of X, so if X was Yogurt, $i = 1$ is Skyr, $i = 2$ is Greek, and so on.

We will describe demand using the following formula.

$$X_i^D = S \times \left(\frac{1}{N} - \sigma(P_i - \bar{P}) \right).$$

In this model, S = the total size of the industry (measured in output), N = the number of firms in the industry, σ describes the elasticity of demand, or how substitutable the products are ($\sigma \uparrow$ means goods are more substitutable and demand is flatter); P_i is the price charged by firm i, and $\bar{P}$ is the average price charged.

This is *one model* of demand for differentiated products, meant to illustrate all the economic forces at play.

First, note that as an individual good's price gets farther from the average good, $P_i > \bar{P}$, consumers will buy less of it, and if $P_i < \bar{P}$, consumers will buy relatively more of it.

Second, note that in a symmetric equilibrium, where $P_i = \bar{P}$ for all goods, then $X^D = S/n$: consumers split their total consumption of X-like goods over the number of firms in the industry.

For us, this demand curve is sufficient to capture all the relevant features of a monopolistically competitive market on the demand side.

10.3.2 Market equilibrium

To figure out what happens in equilibrium, we need a few more assumptions.

1. **Monopolistic Competition** Firms take as given the number of competitors, N and the average price of competitors $\bar{P}$ when choosing their price. They act as monopolists over their own variety of X, but face competition from other producers through the number of producers and the average price in the industry. In an oligopolistic setting, when there are only a few powerful firms, firms know they influence the average price and whether competitors enter or exit, and that requires game theory to study. That is not the context of our model!

2. **Symmetry** Firms are identical. While each product may be differentiated, the cost structure of firms is the same: they have the same marginal and fixed costs. Therefore, in equilibrium, all firms will charge the same price.

3. **Free entry** We will assume firms can freely enter and exit the industry. They will enter if profits are positive and exit if negative. Free entry results *zero profits*. Firms will enter until profits are competed away.

Using these assumptions, let's build up what the equilibrium looks like.

Monopolistic competition With no strategic interactions in the marketplace, we can rearrange firms' demand curves to look like the curves we saw in the monopolist's problem in the last section. You may convince yourself, then, that,

$$MR_i = P_i - \frac{X_i}{\sigma S}$$

Now, firms price optimally, so the $MR_i = MC_i$,

$$P_i = MC_i + \frac{X_i}{\sigma S}$$

Symmetry Now, let's impose symmetry. Studying the symmetric equilibrium, where $MC_i = MC$ for all i, and $X_i = S/N$,

$$P = MC + \frac{S/N}{\sigma S}$$

Cancelling S, $P = MC + \frac{1}{\sigma N}$. The markup is $\frac{1}{\sigma N}$. As consumers view goods as more substitutable $\sigma \uparrow$, or there are more substitutes, $N \uparrow$, the markup falls, and monopolists lose the ability squeeze consumers.

The competitive state of an industry can be described by the size of the markups. Competition is tougher when there are a lot of competitors (N high) or consumers are responsive (σ high). Even holding fixed σ, when more firms enter, markups fall, and competition toughens!

Free entry Now, let's impose free entry. We will determine the number of firms, N, endogenously. Firms will freely enter and exit until profits are zero.

Recall, profits can be written as,

$$\text{Profits} = (P - AC) \times X,$$

so when $P = AC$, profits are zero. Knowing that $X_i = S/N$ for all producers, AC can be written as $MC + F/(S/N)$, or, rearranging, the free entry condition can be written as,

$$P = MC + N \times F/S$$

Equilibrium The equilibrium of this model can be described by the *price level*, P (which is the same across all firms, due to symmetry), and the number of firms in the market N. Our two equilibrium equations are,

$$\begin{aligned} \text{Optimal pricing:} \quad & P = MC + \frac{1}{\sigma N} \\ \text{Free entry:} \quad & P = MC + N \times F/S \end{aligned}$$

Given costs MC and the size of a market, S, the price level and number of firms is given by the solution to the two equations above.

The optimal pricing equation says that firms will pick a markup over marginal cost that depends on the number of competitors they have. The higher N is, the tougher the competition is, and the lower the price will be.

The free entry condition, $P = AC$ is increasing in N. This says that when the price level is higher, more firms will enter – it will take more firms to dilute profits to zero when prices are high.

Of course, this can be demonstrated graphically! In Figure 10.3, I demonstrate what the equilibrium looks

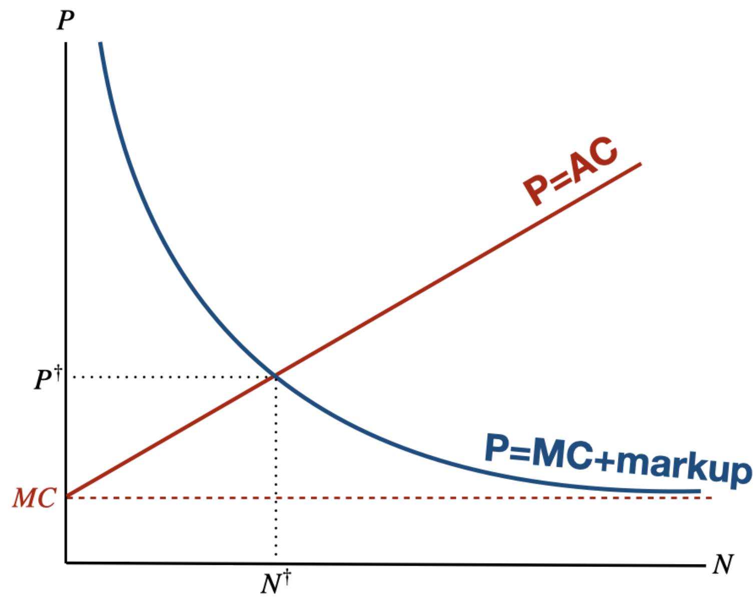


Figure 10.3: (Autarky) equilibrium in the Krugman model

like graphically. The free entry condition, $P = AC$ is upward sloping. It intersects the P-axis at MC. The downward-sloping curved line is the optimal pricing equation. It shows firms' monopoly pricing strategy as a function of the number of competitors they face. As N increases, their ability to raise price over marginal cost diminishes, until the optimal pricing equation asymptotes to MC.

This figure captures the entire Krugman model equilibrium.

There are two things to note here. One, we did a lot of algebraic heavy lifting to get to a diagram with two curves. This may feel disappointing! However, the algebra is meant to clarify the concepts underlying the graph. The most important thing is understanding these concepts, and what each line on the graph means.

The second thing to note is that there aren't a lot of things to play with in this graph. At the core of the model, there are fixed costs, F, marginal costs MC, substitutability of goods σ, and market size S.

What does trade have to do with any of these?

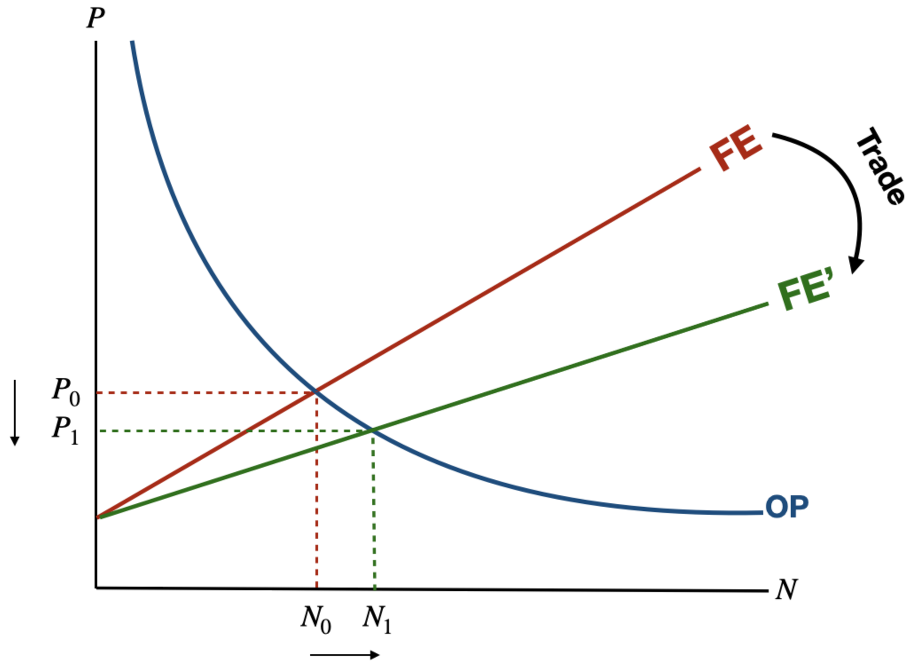


Figure 10.4: Trade in the Krugman model: an increase in market size

10.4 Trade in the Krugman model

Trade, in the Krugman model, appears by changing the size of a market, S. The more consumers demanding goods in a particular industry, the larger the demand will be for any given good in the industry. That's it!

Opening to trade in the Krugman model means $S \uparrow$.

How will this affect trade? S does not appear in the pricing equation. Firms' pricing decisions depend only marginal costs, σ, and N. S only shows up in the free entry condition. The larger the market, the more consumers you can spread your fixed costs over. More firms can afford to enter in larger markets. Said differently, the larger the market, the less fixed costs matter.

Recall the free entry condition,

$$P = MC + N \times \frac{F}{S}$$

As the market becomes larger in size, due to trade, S grows larger and the slope of the free entry condition flattens. At a given price P, in a larger market, more firms can afford to enter.

Thus, an increase in market size *rotates* the free entry curve! This is displayed in Figure 10.4. As the Free Entry cure (FE) rotates to FE', the price level falls from P_0 to P_1 and the number of firms in the market expands from N_0 to N_1.

These are the sources of our new gains from trade in this model!

As more firms enter, there are more varieties available in the market. American consumers can import Skyr and Greek yogurt. These are the *love-of-variety* gains from trade. Consumers value more varieties. Trade creates larger markets and allows more goods to exist in the marketplace.

The second gain from trade is the *pro-competitive* gains from trade. With more goods on store shelves, monopolistically competitive firms face more competition and have to lower prices. Consumers benefit from these lower prices! Trade toughens competition, limiting the ability for monopolistic firms to squeeze consumers.

10.4.1 Free trade with two countries

How should we think about free trade with two countries in this model? The answer is frustratingly simple: Trade integrates markets. Instead of two markets operating in autarky, trade creates one global market within an industry.

If there are two countries, Home and Foreign, with sizes S and S^* and prices and varieties P_A, N_A and P_A^*, N_A^*, when in autarky, what trade does is integrate the markets so that the new market size is $S^W = S+S^*$. From either country's perspective, the size of the market expands. That means the number of available varieties expands and the price level falls in *both* Home and Foreign, $N^W > N_A$ and $N^W > N_A^*$ while $P^W < P_A$ and $P^W < P_A^*$.

While trade integrates markets and makes more variety available for consumers, it will not be the case that $N^W > N_A + N_A^*$.

To see this, combine the equilibrium equations to solve for the optimal number of firms,

$$\underbrace{MC + \frac{1}{\sigma N}}_{\text{Optimal pricing}} = \underbrace{MC + N\frac{F}{S}}_{\text{Free entry}}$$

Rearranging this,

$$N = \sqrt{\frac{S}{\sigma F}}$$

This implies that,

$$\underbrace{\sqrt{\frac{S+S^*}{\sigma F}}}_{N^W} < \underbrace{\sqrt{\frac{S}{\sigma F}}}_{N_A} + \underbrace{\sqrt{\frac{S^*}{\sigma F}}}_{N_A^*}$$

This means that trade must create winners and losers among producers.

10.5 Winners and losers

The last section shows that some firms at Home and some firms in Foreign might have to exit, since the total number of surviving firms after market integration will be less than the sum of those that existed in autarky. Who wins and loses from trade in this model?

To make some progress on this question, we will do away with the **Symmetry** assumption, and allow firms to be *heterogeneous*. They will differ in their productivity, which I will define to be the inverse of marginal cost,

$$\underbrace{Z_i}_{\text{Productivity}} = 1/MC_i.$$

This notation will be useful in later lectures to understand which firms choose to trade at all.

With symmetry relaxed, firms will charge different prices and have different levels of profits, so therefore symmetry will kill zero profits, too. Firms will now have different profit levels.

Firms with lower marginal costs (high Z_i) will get to operate on a lower part of the demand curve and charge a higher markup. That is, more productive firms are more profitable because they charge higher markups.

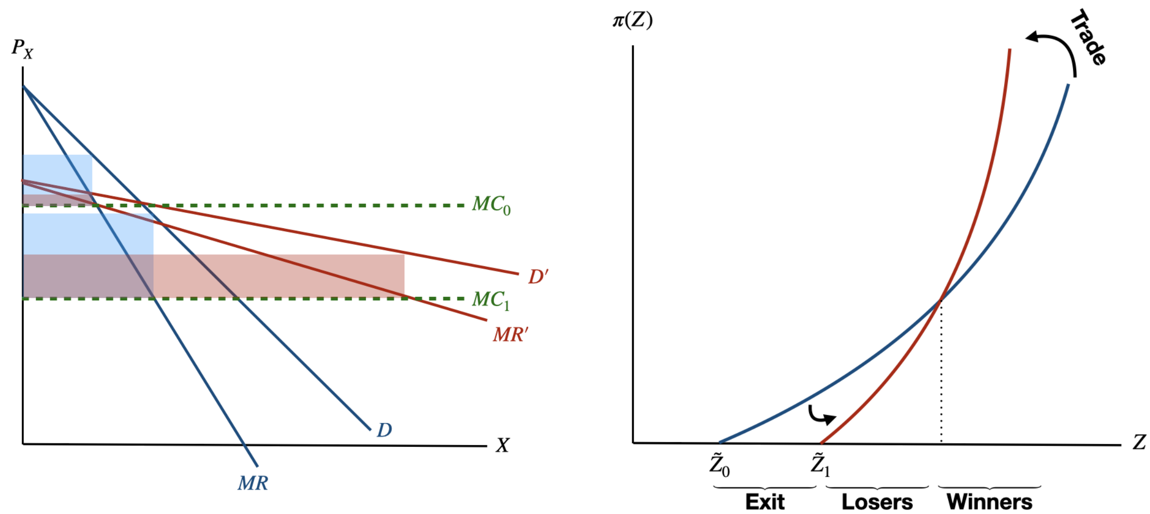


Figure 10.5: Left: change in profits for low- and high-productivity firms when opening to trade. Right: Winners and losers opening to trade in the Krugman model.

Profit can be written as,

$$\underbrace{\pi(Z_i)}_{\text{profit}} = \underbrace{\text{Markup}(Z_i) \times X_i}_{\text{operating profit}} - F$$

Let $\tilde{Z}$ describe the productivity of a firm such that their optimal markup creates operating profits exactly equal to the fixed cost. This firm will be indifferent to producing and earn zero profits, $\pi(\tilde{Z}) = 0$. Due to of the presence of fixed costs, some producers with productivity $Z_i < \tilde{Z}$ may not choose to produce at all.

What trade does, by increasing the market size, is toughen competition by pushing out low-productivity producers. With tougher competition, the profitability differences between the big and small firms are more exaggerated, and more productive firms get a larger market share than they did before. Why is this?

Recall from the Krugman model that the demand curve faced by a monopolisitcally competitive variety i can be written as,

$$X_i^D = \frac{S}{N} + \sigma S \bar{P} - S\sigma P_i$$

With trade, $S \uparrow$. In equilibrium, this results in $N \uparrow$ at rate $\sqrt{S}$ and $\bar{P} \downarrow$. The y-intercept on the demand curve is equal to, $\frac{1}{\sigma N} + \bar{P}$, so trade *lowers* the y-intercept on the demand curve. Moreover, as the slope of demand is $S\sigma$, trade also makes the demand curve *flatter*. Preferences are more elastic under trade and market shares are lower. The left panel of Figure 10.5 displays how this shift in demand from opening to trade affects a high- and low-productivity firm. The low-productivity firm has higher marginal costs than the high-productivity firm $MC_0 > MC_1$. By carefully drawing Figure 10.5, you can see that the change in profit from pre-trade (blue rectangle) to post-trade (red rectangle) is negative for the low-productivity firm. On the contrary, for the high productivity firm, their sales substantially increase, more so than the price declines. This means that profits for the high productivity firm rise. In sum, trade steepens the relationship between productivity and profit.

10.6 'Home market' effects

The theory so far has made no effort to explain *in which country* production occurs. We know some firms have to exit when a market integrates. When goods are differentiated, can we say which goods are produced where?

We would need to stretch the model we've developed in this lecture pretty far to answer this question. We definitely need to drop the Symmetry assumption. We would need more to make the location of industry matter. In this model, with free trade, the markets integrate, and firms do not care in which country they produce. We could modify this in several ways. We could assume firms cannot move, we could change the marginal cost of production when moving, or make firms pay additional costs. If there were external scale economies, like in Lecture 9, we would know that all firms would want to locate in the same place, but we still wouldn't know *which* place. The most realistic thing we could add to understand the geography of production would be trade costs. We will not model these formally.

Instead, the notion of 'established advantage' from Lecture 9 will still prove to be useful. With trade costs, firms would want to locate in the market where their sales are the largest, in order to minimize paying trade costs. Varieties that have relatively larger sales at Home (i.e., larger autarky market shares) will tend to be the goods that are exported, because firms will want to stay near their centers of demand. These are the goods that get to better achieve scale economies in autarky and will survive when markets integrate.

We call this the *Home market effect*: countries with high local demand for specific products also witness heightened sales of those products in foreign markets. Countries export the varieties that are most demanded in the Home market. *Comparative advantage* in producing a particular variety comes from having strong enough domestic demand to achieve large scale economies and drive costs down. Strong local demand 'establishes an advantage' for the production of a particular variety.

Lecture 11 – International Investment

In this lecture, we will think not just about goods flowing across countries, but also capital.

This lecture will contain some standalone content on national income accounting, and determinants of the current account. At the end, I will examine a model of foreign direct investment (FDI) that uses tools from the Krugman model.

11.1 National Income Accounting and the Balance of Payments

National income is the value of goods and services produced in a nation, GDP. A standard macroeconomics course begins with,

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

i.e., the value of all goods and services produced is either consumed, invested, used by the government, or spent on imports. Goods produced at home are either consumed at home, or exported abroad. In reality, goods are not the only component of national income that move between countries. There are many other payments between countries.

All such payments can be recorded in the Balance of Payments (BoP). The balance of payments is the difference between all money flowing into a country in one period minus all money flowing out. The balance of payments is composed of two pieces:

1. Current Account (CA)
    - Trade in goods (this course – commodities, intermediates, merchandise, and other final goods)
    - Trade in services (tourism, consulting, royalties)
    - (Net) factor payments (payments to labor and capital abroad – e.g., hiring a French worker at an American firm), NFP
    - Unilateral transfers (foreign aid and remittances)

2. Capital and Financial Account (KFA)

   - Trade in physical assets (ownership of art, machines, real estate)
   - Trade in financial assets (stocks, bonds, etc)
   - Trade in foreign reserves (currencies)

These are accounting devices for thinking about the flow of income between countries. They act like bank accounts or balance sheets, but they are fictious in that they aren't managed by any institution, but rather are the outcome of decisions of private agents acting in every country.

The current account is the mirror image of the capital and financial account. Anything credited to the current account is debited to the capital and financial account. In our models, trade has strictly been in goods: A country that exports X gets in return, Y at the 'real exchange rate' or terms-of-trade P_X/P_Y. In reality, trade is often financed on credit. Importing firms at Home pay exporting firms in Foreign in Home's currency, upfront, or in installments with interest, or via claims on streams of future income (like stocks).

This means that the value of the current account CA and the value of the capital and financial account KFA must sum to zero,

$$CA + KFA = 0$$

To keep track of this accounting, let's imagine a German tourist that enters the United States and spends \$100 on a hotel and dining in their first night in the United States. The U.S. current account rises by \$100 – this is an 'export' of goods and services (food and hotel lodging) to Germany. In exchange for the meal, bed, and housekeeping, and so on, the U.S. economy gets 100 dollars in return. To get these 100 dollars, the tourist would have to exchange, say, 80 euros at a bank. The bank would only make this exchange if they could do something with the euros. There are several things that can happen to those euros once they enter the bank. The bank could invest in a European firm, or buy European financial assets, which subtract exactly 80 euros / 100 dollars from the U.S. KFA. The bank could also use those euros to purchase European goods or services – which would lower the CA by again, exactly 100 dollars / 80 euros. Thus, it remains that $CA+KFA = 0$. This is true for all international transactions.

There is one more thing to note about this accounting: any country with a current account deficit must have a capital and financial account surplus.

11.1.1 Savings in the open economy

Let's leverage these definitions.

First, let's define GDP and GNP for Home. GDP (gross domestic product) is the value of all goods and services produced at Home, while GNP (gross national product) is the value of all goods and services produced by Home's factors. Some workers are Home get payments from Foreign firms (a U.S. worker at a French company) while some Foreign factors are paid by U.S. firms (Chinese workers who work at U.S. companies, rented machines owned by Canadians used in the production of Vermont maple syrup). The difference between GDP and GNP is the difference in these factor payments, which I will define as Net Factor Payments, NFP,

$$GNP - GDP = NFP$$

For now, ignoring government spending (assuming $G = 0$),

$$GDP = C + I + EX - IM$$

where EX is the value of exports and IM is the value of imports, and definining the current account CA as,

$$CA = EX - IM + NFP,$$

we can write GNP as,

$$GNP = C + I + CA$$

Thus, the current account reflects the net value of payments received by Home from Foreign.

Let's push this accounting further. Let's define national savings as,

$$S = GNP - C,$$

i.e., national savings is equal to the value of all national income that is not consumed. Using the definition of GNP, and rearranging, it must be such that,

$$CA = S - I.$$

In a closed economy, as in a traditional macroeconomics course, national savings = investment. However, in an open economy, with trade, national savings can be invested or spent abroad. A current account surplus or deficit reflects whether national investment outpaces or lags behind national savings.

11.1.2 Current account accounting

Consider our example of a German tourist in the U.S. When he exchanges euros for dollars, and spends those dollars in the U.S. economy. The CA rises, but the KFA falls as the bank invests those euros abroad. That investment back in Germany raises the German KFA, and that increase in German capital must then be matched by a reduction in the German current account. Put more simply, every action that credits a national account from Home's perspective must debit the account from Foreign's; everything has to add up. Therefore, put simply, if CA_i is country i's current account,

$$\sum_i CA_i = 0.$$

This is simply that globally, exports, imports, and net factor payments must add up. Thus, if Home is running a current account surplus, its excess savings must be matched *exactly* by a savings deficit abroad; Foreign must be running a current account deficit.

The current account is a flow variable. It measures the flow of savings between countries in a given unit of time (e.g., quarter, year, and so on). For every flow, there is a stock. For example, income to your bank account is a flow, and your wealth (measured by your bank account) is a stock.

The stock counterpart of the current account is the position of what's referred to as the Net Foreign Asset (NFA), which is the net value of claims on Foreign assets. For example, if $NFA < 0$, this may reflect that Home has been running a current account deficit for years, financed with the sale of domestic assets or interest payments from claims on Foreign assets. How the position of the Net Foreign Asset evolves depends on the interest rate r_t and the current account in a given year, t.

$$NFA_{t+1} = (1 + r_t)NFA_t + CA_t$$

In other words,

$$\underbrace{\Delta NFA_{t+1}}_{\text{changes in the net foreign asset position}} = \underbrace{r_t NFA_t}_{\text{valuation effect}} + CA_t$$

Note that a consequence of this definition is that persistent current account deficits at Home can be stable, in the sense that debt held by Foreign doesn't grow. For this to be true, we would need $NFA_{t+1} = NFA_t$. Let's call the level that achieves this NFA^s where the ss stands for 'steady state.' For it to exist, the interest rate must also be constant over time, r^{ss}. Then NFA^{ss} must solve,

$$NFA^{ss} = (1 + r^{ss})NFA^{ss} + CA^{ss}$$

i.e.,

$$-CA^{ss} = r^{ss}NFA^{ss}$$

or if the steady-state current account deficit is financed by the interest earned on its Foreign asset position.

11.1.3 Determinants of the current account

So far, we have done absolutely no economics. All previous parts of this lecture have been accounting. By repeatedly employing the fact that every dollar spent must end up somewhere, we derived that in an open economy, countries can run current account surpluses (and 'enjoy' excess savings) provided those savings inflows are exactly matched by a current account deficit somewhere else.

Current account deficits are by no means something bad or undesirable. They simply reflect differences in savings supply and demand globally. What determines the supply and demand of savings globally?

In this section, we will do some economics and develop a trade model of investment to see what determines international investment and current account surpluses and deficits.

In this model, there will be one good, and two different time periods: today and tomorrow. The good is produced with labor and capital, which is in fixed supply, $Y = F(L, K)$. Without investment, consumption would have to be equal to income, $C = Y$ every period. However, consumers can save, and augment the capital stock. In autarky, without international investment, the total amount of savings S is the total amount of investment capital available, $S = I$. Consumption and savings tomorrow, $C' + S'$ comes income tomorrow, which depends on investment.

$$C' + S' = Y' = F(L, K + I)$$

where, $I = S = Y - C$. For simplicity let's assume that there is only today and tomorrow, after which the world ends, so that households always choose $S' = 0$.

Households have a utility function $U(C, C')$ with associated marginal utility of consumption today and tomorrow MUC and MUC'.

The incentive to save comes from earning interest on savings, r. One dollar saved today earns $1 + r$ dollars tomorrow. Households take as given that $Y = F(L, K)$, and solve,

$$\max_{C,C'} U(C, C') \quad \text{subject to} \quad C + S = Y,\ C' = (1 + r)S + Y'$$

We can combine these equation to write the household's intertemporal budget constraint,

$$C' = Y' + (1 + r)(Y - S)$$

which rearranged can be expressed in present value terms,

$$C + \frac{C'}{1 + r} = Y + \frac{Y'}{1 + r}$$

where the interest rate is r, and $S = I$, $Y' = F(L, K + I)$.[1]

The maximum amount of consumption today is simply Y and the maximum amount of consumption afford-

[1] Careful macroeconomists should note that this formulation means that output cannot be "re-eaten" the next period, but rather costlessly converted into capital,

$$K' = K + Y - C.$$

However, capital is converted into output with diminishing returns. This is the neoclassical formulation of a consumption-savings problem.

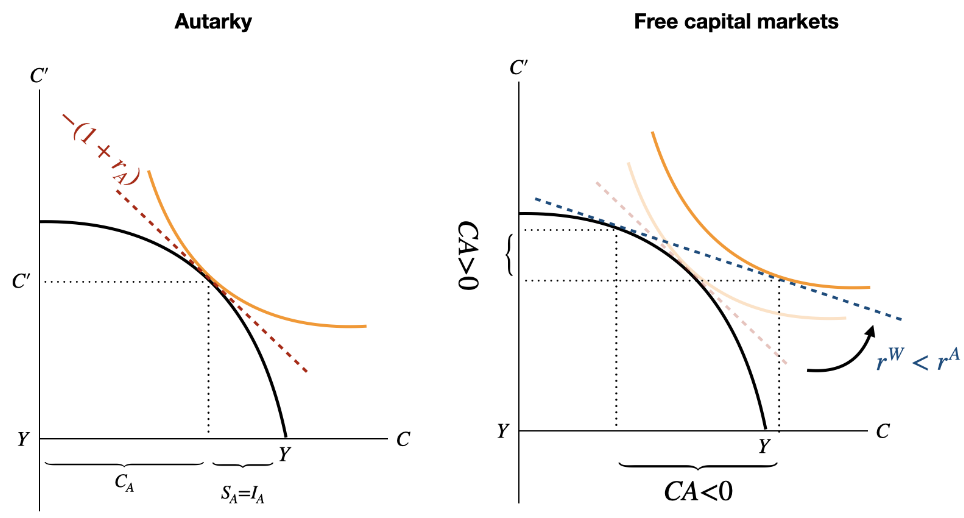


Figure 11.1: Inter-temporal trade as a model of current account deficits.

able tomorrow is $Y' + (1 + r)Y$. How much is saved $Y - C$ depends on the tradeoffs households face between utility today and utility tomorrow, which depends on the interest rate. Households maximize where $MUC/MUC' = (1 + r)$.

The interest rate, in this autarkic view of the world, comes from technology: it is the ability of the economy to transform one unit of savings into consumption tomorrow. It must therefore be equal to the marginal product of capital, MPK.

Absent other changes in the economy, investing I dollars means that consumption tomorrow is,

$$
\begin{aligned}
C' &= F(L, K + I) \\
&\approx F(L, K) + MPK \times I \\
\text{by definition,} \quad &= Y + MPK \times I \\
\text{using } Y = C + I, \quad &= (C + I) + MPK \times I \\
&= C + (1 + MPK) \times (Y - C)
\end{aligned}
$$

which shows us that $r = MPK$. Remember from the Heckscher-Ohlin lecture that this defines firms' optimal choice of capital. Thus, the equilibrium interest rate clears the capital market!

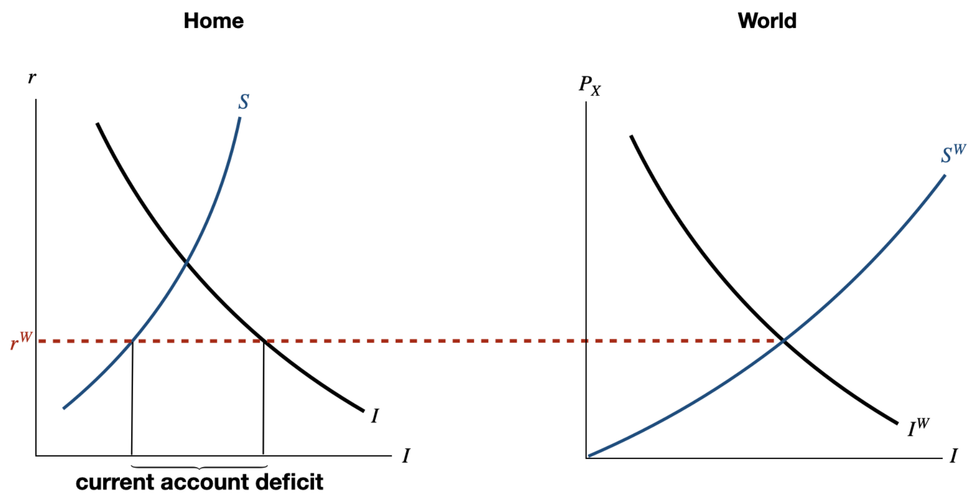


Figure 11.2: Home and world market for capital

Consider graphically an economy in autarky that then opens to trade in savings. Because there are diminishing returns to capital, saving one unit of C today has diminishing returns in creating more C' tomorrow. Households have indifference curves with all the usual properties. They optimally pick C and C' so that their $MRS = (1 + r^A)$ and firms demand K until $MPK = r^A$. The equilibrium is that some output today is saved to increase consumption tomorrow. This is graphed in the left panel of Figure 11.1.

In the right panel, we image that Home liberalizes its capital markets and can borrow at the world interest rate $r^W < r^A$. At a lower interest rate, Home wants to borrow more to consume today, and will pay its debts by lending to the world tomorrow.

This is one model of determining the current account.

Just like our SFM and HO models allowed us to derive world supply and demand curves for savings, we can use this model to derive an supply-and-demand open economy figure. I do this in Figure 11.2.

In Figure 11.2, world interest rates are low, because relative to Home, there is a global savings glut. Households abroad are saving more than firms need for investment, and that capital flows out of its domiciles to investment opportunities abroad, like those at Home. At these low interest rates, Home's investors demand more than its savers want to save, and so Home runs a current account deficit ($CA < 0$) as $I > S$. This investment capital comes from Foreign, who must run a current account surplus.

This is a very useful model of current account surpluses and deficits. It shows how the global financial market operates similarly to our trade model.

Consequently, all of the intuition we developed there is portable to this model, too. For example, interest rates are determined globally, so small countries that save don't put downwards pressure on interest rates. If households cut back on savings, the current account deficit will rise, unless the fall is met by a decrease in demand for capital by firms.

11.2 Foreign direct investment

One type of investment that firms undertake is foreign direct investment, or FDI.

FDI refers to a firm directly invests in a foreign country through the purchase of an ownership stake of a foreign business, or through the direct establishment of a business abroad. Examples of this include joint ventures (i.e., partnerships) between companies at Home and Foreign, or creating a local subsidiary of a business at Home in Foreign.

Why might firms want to do this? In this subsection, I will consider a firm that is grappling with the proximity-concentration trade-off.

If a firm is exporting abroad, they have to pay trade costs to do so. This may include fixed exporting costs, like hiring a trade lawyer, support staff, buying warehouse space near a port, etc, as well as fees to sell abroad, including certifications, health inspections for food processing, and so on. Firms also pay variable shipping costs to sell abroad too, like container shipping, the cost of shipping insurance, and so on, that vary with the volume of goods exported.

Firms would still choose to pay for all these things if selling abroad guaranteed a larger market over which to they could spread their fixed costs and earn more profit. A firm benefits from concentrating production at Home and achieving scale economies, but the cost of this is that they lose proximity to the markets they serve, and have to pay these trade costs to export.

An alternative to paying trade costs is taking their know-how and establishing a subsidiary abroad. Consider a potato chip company like Lays. Instead of growing and processing potatoes in Idaho to make chips and sell them all over the world, they could operate plants that make their chips locally with local potatoes, labor, and so on, in foreign markets. Then, per bag sold abroad, Lays no longer has to pay all the trade costs. However, the cost of opening a Foreign subsidiary could be quite large. In addition the opportunity cost of sacrificing the scale economies of concentrating all production in one place, establishing production abroad means paying a host of new fixed costs: new factories abroad, new overhead costs, and a much larger legal team.

Let's consider this through the heterogeneous-firm version of the Krugman model developed in Lecture 10.

We line up firms by their productivity, Z_i, which is the inverse of their marginal costs. Because more productive firms have lower marginal costs, they are more profitable.

Let's define a firm's operating profits from selling domestically, D, exporting X, or doing FDI by establishing a new plant in the following way,

$$\underbrace{\pi^D(Z) - F^D,}_{\text{profits selling at home}}, \quad \underbrace{\pi^X(Z) - F^X,}_{\text{profits exporting}} \qquad \underbrace{\pi^{FDI}(Z) - F^{FDI}}_{\text{profits doing FDI}}$$

where $F^{FDI} > F^X > F^D$. Let's assume that the markets sizes and the same at Home and at Foreign. Therefore, because of trade costs, $\pi^X(Z) = \frac{1}{t}\pi^D(Z)$ where $t > 1$ are trade costs, while $\pi^{FDI}(Z) = \pi^D(Z)$. To simplify notation, then, we can just define $\pi^D = \pi$.

Firms have three choices: sell only domestically, sell domestically and export, or sell domestically and do FDI. Their choice depends on their profits,

$$\begin{aligned} \text{domestic-only profit} &= \pi(Z) - F^D \\ \text{exporting profit} &= \pi(Z) - F^D + \frac{1}{t}\pi(Z) - F^X \\ \text{FDI profit} &= 2\pi(Z) - F^D - F^{FDI} \end{aligned}$$

Profit, as a function of productivity, can be displayed graphically. I do so in Figure 11.3. Firms select into

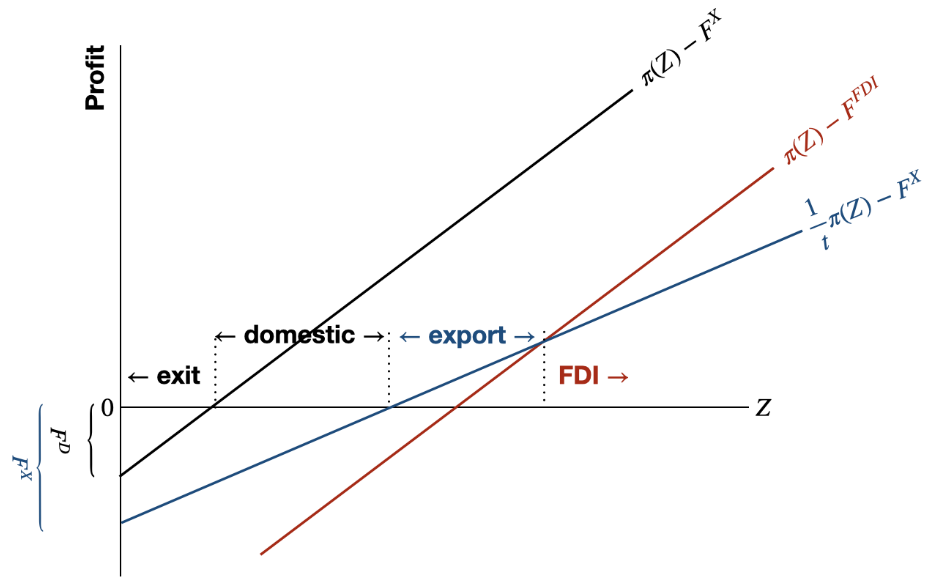


Figure 11.3: Profits under selling domestically, exporting, or doing FDI

exporting if,

$$\frac{1}{t}\pi(Z) - F^X > 0$$

and do FDI instead when,

$$\pi(Z) - F^{FDI} > \frac{1}{t}\pi(Z) - F^X > 0.$$

The first amounts to the condition that you export your operating profits can cover the product of your fixed and variable trade costs, $\pi(Z) > tF^X$, while the latter amounts to export if your operating profits can cover the additional fixed costs, adjusted for variable trade-cost savings, $\pi(Z) > \frac{t}{t-1}(F^{FDI} - F^X)$.

As more productive firms are more profitable, it is the case that the most productive firms choose to do FDI, while moderately productive firms select into exporting only. Low-productivity firms do not export at all, while the lowest productivity entrepreneurs do not operate in equilibrium because they cannot make enough profit to cover their fixed costs.

This model is sufficient to do some analysis. In Figure 11.4 I show a reduction in trade costs. A reduction in trade costs $t\downarrow$ means that more firms will select into exporting from operating domestically, but the marginal firm that begins to export is less productive.

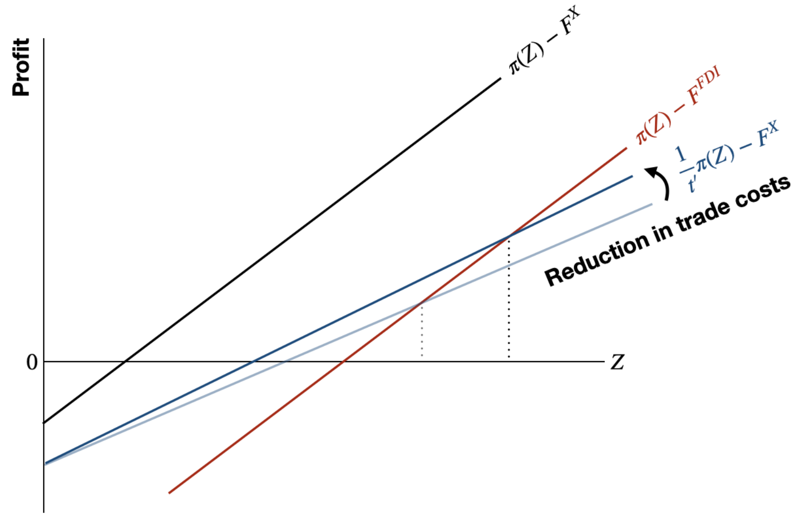


Figure 11.4: A reduction in trade costs

The opposite is true for FDI. With lower trade costs, the least productive FDI firms will stop doing FDI, and retrench to their domiciles and export. Thus the marginal firm going from FDI to exporting is more productive than the set of original exporters. The firms that remain doing FDI despite the lower trade costs are on average more productive than before.

Lecture 12 – Trade costs

So far, apart from tariffs, we have chiefly studied *free trade* in a world with two countries. What I mean by free trade is that absent tariffs, the prices faced by Home and Foreign are the same, and that it isn't costly to move goods produced at Home to Foreign, and vice versa.

We can add as much realism as we'd like to the models we've developed – realistic production, increasing returns, multiple factor markets, varieties of goods – but we'd still be stuck with the fundamental unrealistic feature of free trade and two countries.

In reality, of course, many countries participate in trade and interact with each other in global markets, and trade between these countries is not free! Japanese electronics exports bound for America would be loaded on a container ship in the Port of Tokyo. From Tokyo, it takes a little over two weeks to arrive in the closest U.S. port – Seattle. At minimum, it costs two weeks of fuel and longshoreman labor to make that trip, let alone insurance and port fees container ships must pay to make the trip and to dock in Seattle's berths upon arrival. If Seattle's port is congested, container ships drop anchor in the Puget Sound and can wait days – paying fuel and labor costs – before they can dock. Separately, an export of Canadian maple syrup bottled in Montreal can make it to Boston on a truck and return in the same day. Trade is not only not free, but its costs depend on who is trading with whom!

In this lecture, we will develop some tools for thinking about trade costs in a multi-country world.

We first will study *the gravity equation*. The gravity equation is a robust empirical relationship that describes trade flows across nations and is useful in predicting how changes in trade costs and national income will change the pattern of trade.

We will then turn to (a much-simplified version of) the 'Armington model,' which helps ground the gravity equation in some simple theory, and we will study the economic content of this model and the gravity equation.

Finally, we will look at what actually determines shipping prices, outside of fuel and labor costs, by briefly studying maritime shipping.

12.1 Gravity

The gravity equation is an *empirical* description of trade flows between countries.

Moving away from Home and Foreign, we'll now (finally!) allow there to be N countries, indexed by i or j. I will denote the value of trade (price $\times$ quantity) of goods originating in i but sold to j (i.e., i's exports to j) as X_{ij}. The GDP of each country is denote Y_i. The 'distance' between i and j is D_{ij}. The gravity equation is,

$$X_{ij} = \frac{Y_i \times Y_j}{D_{ij}}.$$

It says that the volume of trade between two countries is proportional to the size of both their GDPs, and inversely proportional to the distance between them.

Large countries export more goods (they are large and so they produce more goods) and large countries also import more goods (they have more people so they demand more goods). Holding fixed GDP, countries trade more with each other the 'closer' they get; trade declines in 'distance.'

I write 'distance' in quotes because there is a lot more than physical distance that separates countries. Countries may be distance because they are separated by many borders, even if they are geographically close. Crossing borders is costly: trucks, boats, and so on, queue at borders and pay fees to cross them. It is much easier to trade goods between New York and California, than London and Paris, even though the latter are much closer physically. Trade is stymied by much more than distance and time at borders. Translating documents and facilitating logistics between countries with dissimilar languages and cultures is costly, too. Thus, linguistic distance matters as well! Using the same currency may facilitate trade, too. In short, there are many bilateral shifters of trade.

12.2 Armington-lite

We ought to have some theory about what explains the gravity equation. It turns out, the demand system used in Lecture 10 does a very nice job.

First, let's suppose each country makes one type of good. There are American goods, Korean goods, Nigerian goods, and so on. Each country i has GDP (in dollar terms) P_iY_i, and bilateral trade is,

$$X_{ij} = P_jY_j \times \left(\frac{1}{N} - \sigma(P_i - \bar{P}_j + \underbrace{t_{ij}}_{\text{trade costs}}) \right).$$

Let's look a little at what this demand system says. First, if there were no trade costs, then a symmetric equilibrium would result in $X_{ij} = P_jY_j/N$, where N is the number of countries, and P_jY_j is the GDP of j's country, expressed as the dollar value of their endowment, Y_j. That is, country j spends a fraction $1/N$ of its GDP on imports from each country (including itself).

Trade costs t_{ij}, which depend on distance (and all sorts of things), make shipping from i to j more expensive, and reduce the volume of trade. Country i sells more to j when i's good is cheaper than the average price $\bar{P}_j$. Why does $\bar{P}_j$ vary now? It's because of trade costs! To see this, let's break open this term,

$$\bar{P}_j = \frac{1}{N}\sum_i (P_i + t_{ij}) = \underbrace{\bar{P}}_{\text{average trade-cost-free price}} + \underbrace{\bar{t}_j}_{\text{average cost of shipping to } j}$$

This is not an economic model yet. We've specified demand, but we haven't specified supply or equilibrium. Supply in this model is very simple. Unlike the rest of the course, we'll just assume Y_i is exogenous. Each country is endowed with some amount of its own good, and demands a bit of every good from around the world. How much it cares about variety is given by σ. Trade costs also shift demand.

In equilibrium, all markets clear, and imports have to equal exports. That is, the GDP of a given country is equal to the total value of all its production. The total value of all goods and services produced must equal all that's consumed and exported. This can be expressed as,

$$P_iY_i = \sum_j X_{ij}$$

This equation says supply $=$ demand for each country i's good, and reflects a system of N linear equations. This is the equilibrium of the Armington model.

To get an insight about what determines GDP in this model, we can use the equation for demand, and define global GDP as $Y^W = \sum_i P_i Y_i$,

$$P_i Y_i = \sum_j P_j Y_j \left(\frac{1}{N} - \sigma(P_i + t_{ij} - \bar{P} - \bar{t}_j) \right)$$

$$\text{using the definition of } Y^W, \quad P_i Y_i = Y^W \left(\frac{1}{N} - \sigma(P_i - \bar{P}) - \sigma \sum_j \frac{P_j Y_j}{Y^W}(t_{ij} - \bar{t}_j) \right)$$

This equation has some economic content! In particular, it is informative about the forces that drive countries relative sizes (shares of global GDP). First, if all countries were the same and there were no trade costs, each country would have a share $1/N$ of global GDP. However, because countries have different endowments Y_i, relative size depends on differences in prices, and a country's position in the global trading network, which I will call *market access*. To see this, rearrange the equilibrium equation to see,

$$\frac{P_i Y_i}{Y^W} = \frac{1}{N} - \sigma \underbrace{(P_i - \bar{P})}_{\text{export competitiveness}} - \sigma \underbrace{\sum_j \frac{P_j Y_j}{Y^W}(t_{ij} - \bar{t}_j)}_{\text{market access}}$$

What does market access tell us? Countries that are well-positioned in the global trade network are countries that are close to big countries – that for countries with big shares, of the global market ($\frac{P_i Y_i}{Y^W}$ big) that $t_{ij} < \bar{t}_j$. This equation tells us that geography matters!

To recap what we just did: using the demand system that countries want to import a bit from everyone (depending on how substitutable imports are, σ), but there are trade costs t_{ij} that hinder trade, and the fact that globally, imports = exports for every country, we got an insight in what the gravity equation means for a country's gains from trade! A country's gains from trade depend on how competitive its exports are, and how well-suited it is to export those goods.

12.3 Maritime shipping

Over 70% of the value of global trade moves between countries on boats. Most of this trade is carried on Ultra Large Container Vessels (ULCVs) or Panamax ships (i.e., ships that are of the maximum size able to fit through the Panama canal).

The cargo of these ships is either container shipping, or dry bulk cargo. Container shipping is a standardized unit that can move, via crane, from the backs of ships onto trucks. Most final goods (guitars, canned pickles, iPhones, Ikea cabinets, etc) are shipped via these containers. Commodities like wheat and oil are shipped in bulk and held in large containers in the hull of the ship. Speciality bulk carriers carry chemicals and oil.

While journeys on the open seas often have little traffic, and are only held up by changes in weather and sea currents, there is significant traffic at a handful of locations. Maritime bottlenecks form at a handful of locations: the Panama canal, the Suez canal, the Bosphorous (connecting the Black Sea to the Meditteranean in Istanbul), and the Straits of Malacca (between Indonesia and Malaysia). Once reaching a destination port, a ship may have to wait in anchorage, queueing for a berth so they can unload.

Due to shipping delays from weather, traffic, and congestion, much of commodities trade cannot be organized in advance. Instead, dry bulk carriers act as 'taxis of the seas,' and only contract with exporters in a destination upon arrival. Once cargo has been unloaded at its destination, ship captains look to contract with a local exporter and determine their next destination. In some locations, such contracts cannot be formed, or shipowners don't find them lucrative enough to sign. In this case, dry bulk carriers must ballast to another destination and hope to meet an exporter. This means they have to travel with junk (ballast) in their hull to provide adequate weight to make sailing safe.

This taxi-cab like organization of global shipping affects trade costs! How might it affect trade costs?

Let's consider a ship sailing from country i to country j. It has done so at a price t_{ij}. The value of this trip for a shipowner, V_{ij} is equal to the price net the costs, plus the value of arriving at the destination. If a trip sends a ship to an unprofitable destination, the shipowner will have to waste fuel traveling without cargo (i.e., ballasting) elsewhere. They would demand to be compensated for this cost upfront in the price of the journey. Similarly, a shipowner sent to a profitable destination may be willing to provide a discount on cargo, knowing that they will be able to find a profitable exporter once arriving at their destination.

Formally, this value is,

$$V_{ij} = \underbrace{t_{ij}}_{\text{shipping price}} - \underbrace{c_{ij}}_{\text{sailing costs}} + \underbrace{V_j}_{\text{value of arriving at } j}$$

where the value of arriving at j is,

$$V_j = \underbrace{-F_j}_{\text{port fees}} + \underbrace{\mathbb{E}_k[V_{jk}]}_{\text{continuation value}}$$

The continuation value is the expected value of meeting with an exporter at destination j. This may be high at busy ports, like at Singapore, in Hong Kong, which are global shipping hubs. However, if a boat is sent to Christchurch, New Zealand, it might not meet with a profitable exporter, and may have to travel to Sydney, Australia after their journey is complete. Shipowners will only sign contracts where $V_{ij} > 0$ and there are positive profits, and so they will demand a price that covers both the cost of the journey $-c_{ij}$ as well as the future potential cost or benefit of arriving at j. Thus, the price of shipping, t_{ij} depends not only on the fuel and labor costs of the journey, but on the attractiveness of the destination from the shipowner's point of view.

Lecture 13 – Conclusion

Thank you for studying international trade with me.

This course has been somewhat unusual. I have emphasized using graphical analysis and price theory throughout, and have tried to be sparing on the algebra. The set of topics covered has more narrowly focuses on the intellectual lineage of trade theory. We developed a core set of models that were rich enough to take to the data and use for policy analysis, but simple enough to explain on a whiteboard with little math. By structuring the course this way, we have actually covered *more* ground than is usual for a quarter-long international trade course, and we fit in a bunch of case studies too!

I tried my best to provide you with case studies to show you how to think about trade theory when understanding history, current affairs, changes in technology, the location of cities, and the organization of economic activity in the global economy. That is to say – case studies that were not all about what people think of when they think about 'international trade.' I wanted to show you how widely applicable the theory we were developing was.

My goal was to provide you a toolkit that you can apply not just to understanding current and past events, but one that you use in your future endeavors. Regardless of whether you will continue studying international economics, or move on to careers in business, finance, consulting, government, or whatever else, I hope that this course has provided you with something useful or intellectually engaging. Moreover, I hope I have done my role in your UChicago economics education in getting you to think like an economist.

Best wishes and thank you for a great quarter.

Kindly,

Jordan Rosenthal-Kay

Bibliography

Bhagwati, Jagdish (1958). “Immiserizing Growth: A Geometrical Note”. In: *The Review of Economic Studiess* 25.3, pp. 201–205.

Bleakley, Hoyt and Jeffrey Lin (2012). “Portage and path dependence”. In: *The quarterly journal of economics* 127.2, pp. 587–644.

Chen, Tuo, Chang-Tai Hsieh, and Zheng Michael Song (2022). “Non-Tariff Trade Barriers in the US-China Trade War”. In.

Feenstra, Robert C (2015). *Advanced international trade: theory and evidence*. Princeton university press.

Jaffe, Sonia, Robert Minton, Casey B Mulligan, and Kevin M Murphy (2019). *Chicago price theory*. Princeton University Press.

Krugman, Paul, Maurice Obstfeld, and Marc Melitz (2017). *International Economics: Theory and Policy, the latest edition*.

Sullivan, Jake (2023). “Remarks by national security advisor Jake Sullivan on renewing American economic leadership at the Brookings Institution”. In: *Speech, Brookings Institution, Washington, DC, April*.