Topics in International Trade

Tom Schmitz

Bocconi University

September 2020

Preface

This document contains lecture notes for the "Topics in International Trade" class in the Master of Science at Bocconi. I have drawn from a variety of sources, including Ventura (2005), Feenstra (2016), and lecture notes by Arnaud Costinot and Dave Donaldson at MIT (https://ocw.mit.edu/index.htm, , Course 14.581), Kim Ruhl at the University of Wisconsin-Madison (http://kimjruhl.com/adv-international-econ-2019f-econ-871) and Matthew Grant at Stanford.

I am grateful to Carlo Altomonte, Matthew Grant, Andrei Potlogea and Jaume Ventura for useful suggestions, to Carlo Cusumano, Filippo Deserti, Federica Ricci and other students for pointing out typos and mistakes, and to Eugenia Menaguale for preparing the graphs. All remaining mistakes are my own, and I would be glad about any additional feedback (to tom.schmitz@unibocconi.it).

Contents

Part I

General equilibrium with many goods

This first chapter a methodological one. It lays out a basic framework which is at the heart of (almost) all of the models we will study.

To analyse international trade, we need models with more than one good: if all countries of the world produced exactly the same thing, there would be no reason for exchanges across borders! However, general equilibrium models with many goods quickly become complicated.[1] Therefore, we will impose strong restrictions to get tractable models. These restrictions will be imposed on the demand side, that is, on consumers' preferences. This reflects the traditional focus of international trade theories: almost all of them emphasize supply-side reasons for trade (differences in productivities or resource endowments, economies of scale in production, etc.), and therefore deem it acceptable to simplify the demand side as much as possible.

1 Representative consumers and CES demand

1.1 Heterogeneity and the representative consumer

In all models covered in this class, we assume that all consumers in the world have the same preferences, and that the share of their income spent on different goods does not depend on their income level. These assumptions imply that the distribution of income across different consumers does not affect the demand for any good. Therefore, they offer a very convenient simplification: we can derive demand functions from the utility maximization problem of a "representative consumer", a fictitious individual which has the same preferences as everyone in the economy, and earns its entire income.[2]

Obviously, these assumptions do not hold in reality. Different people have different preferences. Furthermore, the German statistician Ernst Engel has shown already in the 1850s that the expenditure share of goods varies systemically with income: as people get richer, the share of food in their total expenditure falls, while the share of luxury goods and services rises. Therefore, redistribution from rich to poor consumers would clearly change demands, increasing the demand for food and lowering the demand for luxury goods. However, these effects are not crucial for the phenomena we are studying,

[1] For a textbook treatment of general equilibrium, see Mas-Colell, Whinston and Green (1995).

[2] For a proof and further discussion of this statement, see Acemoglu (2009), in particular Chapter 5.

and therefore we assume them away for simplicity.

Finally, note that the representative consumer assumptions do not mean that we are unable to study income inequality. Quite the contrary, the effects of international trade on income inequality will be an important focus of this class. The assumptions just ensure that income inequality does not matter for aggregate consumption: an economy with the same aggregate income will consume the same aggregate quantity of each good, no matter how that income is distributed.

1.2 The CES utility function

We now introduce a particular set of preferences which are compatible with the representative agent assumption, and on top of that, have a couple of other useful properties. Assume that there are I goods in the world, and that every consumer has a utility function

$$U\left(c_1, c_2, ..., c_I\right) = \left(\sum_{i=1}^{I} \left(c_i\right)^{\frac{\varepsilon-1}{\varepsilon}}\right)^{\frac{\varepsilon}{\varepsilon-1}}. \tag{1}$$

where ε is a positive real number. This utility function is called a constant elasticity of substitution (CES) utility function, and we will use it for 99% of the models covered in this class.

The representative consumer wants to choose her optimal consumption level for each of the I goods, taking her income $\mathcal{W}$ and the prices p_1, p_2, ... p_I as given. Therefore, she faces the budget constraint

$$\sum_{i=1}^{I} p_i c_i = \mathcal{W}. \tag{2}$$

Formally, the representative consumer's problem is therefore to choose c_1, c_2, ... c_I such as to maximize (1) under the constraint (2).[3] This problem has the Lagrange function

$$\mathcal{L} = \left(\sum_{i=1}^{I} \left(c_i\right)^{\frac{\varepsilon-1}{\varepsilon}}\right)^{\frac{\varepsilon}{\varepsilon-1}} - \lambda \left(\sum_{i=1}^{I} p_i c_i - \mathcal{W}\right), \tag{3}$$

where λ stands for the Lagrange multiplier. The first-order conditions of the problem are

$$\forall n \in \{1, 2, ...I\}, \quad c_n^{-\frac{1}{\varepsilon}} \left(\sum_{i=1}^{I} \left(c_i\right)^{\frac{\varepsilon-1}{\varepsilon}}\right)^{\frac{1}{\varepsilon-1}} = \lambda p_n. \tag{4}$$

[3] Of course, the actual constraint for the representative consumer is $\sum_{i=1}^{I} p_i c_i \leq \mathcal{W}$, but it is straightforward to show that this constraint must hold with equality at the optimum. Otherwise, the representative consumer could increase her utility by using the leftover income to increase the consumption of any good.

To illustrate one crucial property of CES utility, divide the first-order conditions for two goods n and n' by each other to get

$$\left(\frac{c_n}{c_{n'}}\right)^{-\frac{1}{\varepsilon}} = \frac{p_n}{p_{n'}} \Rightarrow \frac{c_n}{c_{n'}} = \left(\frac{p_n}{p_{n'}}\right)^{-\varepsilon}. \tag{5}$$

This shows that the elasticity of the relative consumption of two goods $(\frac{c_n}{c_{n'}})$ to their relative price $(\frac{p_n}{p_{n'}})$ is equal to $-\varepsilon$.[4] That is, the elasticity of substitution between two goods is constant (which explains the utility function's name).

1.3 Demands and the ideal price index

We have not yet fully solved for the representative consumer's choices: the set of first-order conditions is still a complicated system of equations. Instead of it, we would like to express the consumption of every good c_n as a function of only prices and income.

To do so, we need to use the budget constraint. First, note that we can rewrite the first-order condition for good n as $c_n = (\lambda p_n)^{-\varepsilon} U$. Replacing this into the budget constraint gives $\lambda^{-\varepsilon} = \frac{\mathcal{W}}{U}\left(\sum_{i=1}^{I} p_i^{1-\varepsilon}\right)^{-1}$, and replacing this back into the first-order condition yields

$$c_n = \frac{p_n^{-\varepsilon}\mathcal{W}}{\sum_{i=1}^{I} p_i^{1-\varepsilon}}. \tag{6}$$

Equation (6) is the demand function for good n. It depends on that good's price, the price of all other goods, and the consumer's income. Several things are worth noting.

First, the demand function of every good is linear in income, implying that the expenditure share of each good $\left(\frac{p_n c_n}{\mathcal{W}}\right)$ does not depend on income. Thus, if every consumer has the same CES preferences, we can indeed use the representative agent assumption. Mathematically, this result is due to the fact that CES utility is homothetic.[5]

Second, the way the demand for one good depends on the prices for other goods changes with the value of the parameter ε: when $\varepsilon < 1$, a price increase for some good i decreases the consumption of good n, while when $\varepsilon > 1$, an price increase for some good i increases the consumption of good n. So, when $\varepsilon < 1$, good are (gross) complements, and when $\varepsilon > 1$, goods are (gross) substitutes.

[4] Recall from your microeconomics classes that the elasticity of a variable x to a variable y equals $\frac{\partial x}{\partial y}\frac{y}{x}$ and gives the percentage change in x which occurs as a consequence of a 1% increase in y.

[5] This means that if the consumption of all goods is multiplied by a constant γ, the overall utility experienced by the consumer is multiplied by γ^{α}, where α is a positive constant (in this case, 1). Intuitively, this implies that consumers increase consumption in fixed proportions: if they get a little bit more income, they consume a little bit more of every good, in proportions such that spending shares do not change.

Finally, it turns out that we can introduce a very useful concept when manipulating Equation (6) a little. To see this, take both sides of the equation to the power $\frac{\varepsilon-1}{\varepsilon}$ and sum of from 1 to I to get

$$\sum_{1=1}^{I} c_i^{\frac{\varepsilon-1}{\varepsilon}} = \sum_{i=1}^{I} p_i^{1-\varepsilon} \left(\sum_{i=1}^{I} p_i^{1-\varepsilon}\right)^{\frac{1-\varepsilon}{\varepsilon}} \mathcal{W}^{\frac{\varepsilon-1}{\varepsilon}} = \left(\sum_{i=1}^{I} p_i^{1-\varepsilon}\right)^{\frac{1}{\varepsilon}} \mathcal{W}^{\frac{\varepsilon-1}{\varepsilon}}.$$

Now, take both sides to the power $\frac{\varepsilon}{\varepsilon-1}$ to finally get

$$U\left(\sum_{i=1}^{I} p_i^{1-\varepsilon}\right)^{\frac{1}{1-\varepsilon}} = \mathcal{W}. \tag{7}$$

In the following, we will denote

$$P \equiv \left(\sum_{i=1}^{I} p_i^{1-\varepsilon}\right)^{\frac{1}{1-\varepsilon}}, \tag{8}$$

and refer to P as the ideal CES price index (or more simply as the price index). Why do we do this? First, notice that P is indeed a price index, a (geometric) average of the prices of all goods on the market. It is called "ideal" because, as we have just shown, it holds the equality $UP = \mathcal{W}$, or $U = \frac{\mathcal{W}}{P}$. Thus, the utility of the representative consumer at her optimum equals her real income calculated using the price index P. This gives us an easy summary statistic for the welfare of a country, which we will use for instance to assess whether the country as a whole gains or loses from trade.[6]

1.4 The price elasticity of demand

Using the ideal price index, we can rewrite the demand function for any good n as

$$c_n = p_n^{-\varepsilon} P^{\varepsilon-1} \mathcal{W}. \tag{9}$$

Now, we can show another important property of the CES demand system, which holds if there are many goods, that is, if $I \to +\infty$. In this case, the price index P does not depend very much on the value of one individual price, and we can say that $\frac{\partial P}{\partial p_n} \approx 0$ for any good n. Using this approximation, it is easy to show that

$$\frac{\partial c_n}{\partial p_n} \frac{p_n}{c_n} = -\varepsilon,$$

[6]Strictly speaking, it is often hard to say whether a group of people is better or worse off after a change. That is because there is no way to compare utilities, even if everybody has the same preferences. If I have three utils less and you two utils more, it is not possible to say whether we are better or worse off as a group.

However, in our CES/representative-consumer world, a higher utility for the representative consumer does imply that a Pareto-improvement would be possible (that is, making some people happier without making anyone less happy). In that sense, the shortcut of saying "the country gains from trade" when its aggregate real income goes up is valid.

that is, that the price-elasticity of demand for every good is equal to $-\varepsilon$, and therefore, is a constant. This will be very helpful in many of the models we will consider.

2 General equilibrium

2.1 A simple production side

To close our model, we need to make assumptions about the production technology, competition, and the ownership of production factors. So far, we assume that there is only one country (and therefore no international trade).

We assume that every good i is produced with the production function

$$y_i = z_i l_i,$$

where l_i is the labour used for good i, and z_i the productivity with which the good can be produced. We also assume that firms operate under perfect competition.

Furthermore, assume that the representative consumer is endowed with L units of labour, which she supplies inelastically. That is, the consumer has no disutility from supplying labour, and therefore supplies L units no matter which wage she gets. This is also an assumption that we will use in all models we study. Denoting the wage by w, the income of the representative consumer is then $\mathcal{W} = wL$.

2.2 The equilibrium

Recall from your microeconomics classes the definition of an equilibrium in an economic model: an equilibrium is a set of prices such that

1. Consumers maximize their utility, taking prices as given.
2. Firms maximize their profits.[7]
3. Prices are such that all markets clear.

In our example, there are $I + 1$ prices, p_1, p_2, ..., p_I and w, the price of labour. There are also $I + 1$ markets, for all I goods (on each of which we must have $c_i = y_i$) and for labour (where the fixed labour supply L of households must equal the labour demanded by firms for production).

[7] In our example, there is perfect competition, so firms also take prices as given. In later chapters, we will also consider cases where firms have market power.

We have already derived the conditions for consumer optimization in the previous section. Firm optimization is easy: under perfect competition, the price of a good equals its marginal cost, so

$$\forall i \in \{1, 2, ...I\}\,, \;\; p_i = \frac{w}{z_i}.$$

Thus, goods with high productivity have low prices. From the relative demand function in Equation (5), we get

$$\frac{c_n}{c_{n'}} = \frac{y_n}{y_{n'}} = \left(\frac{z_n}{z_{n'}}\right)^{\varepsilon}. \tag{10}$$

This equation gives the relative consumption and production of each good as a function of relative productivities. With this, we easily solve for consumption and production levels. From Equation (10) and the production function, we get $\frac{l_n}{l_{n'}} = \frac{z_{n'}}{z_n}\frac{y_n}{y_{n'}} = \left(\frac{z_n}{z_{n'}}\right)^{\varepsilon-1}$. Then, expressing all labour demands as a function of l_1 and using the labour market clearing condition, we get

$$L = \sum_{i=1}^{I} \left(\frac{z_i}{z_1}\right)^{\varepsilon-1} l_1 \Rightarrow l_1 = \frac{z_1^{\varepsilon-1}}{\sum_{i=1}^{I} z_i^{\varepsilon-1}} L.$$

Thus, for every good n,

$$l_n = \frac{z_n^{\varepsilon-1}}{\sum_{i=1}^{I} z_i^{\varepsilon-1}} L, \quad y_n = c_n = \frac{z_n^{\varepsilon}}{\sum_{i=1}^{I} z_i^{\varepsilon-1}} L. \tag{11}$$

With this, we have solved for the quantities produced and consumed of every good. Note that the labour allocation depends on relative productivities. When goods are substitutes ($\varepsilon > 1$), the economy allocates more labour to more productive firms. For example, electric cars and non-electric cars are quite substitutable. Therefore, as firms get better at electric car production, we would expect more labour to be allocated to electric car production, and less to non-electric car production. However, when goods are complements, the economy allocates more labour to less productive firms ($\varepsilon < 1$). For example, cars and restaurant meals are not very substitutable. So, if car production becomes more productive, we would expect less labour to be allocated to car production, and more labour to be allocated to the restaurant industry.[8]

[8] This simple observation helps to understand the development process of modern economies. As first noted by Baumol (1967), productivity growth in agriculture and manufacturing is much faster than in services. As services and manufacturing goods are not very substitutable, the continued productivity growth of manufacturing means that more and more people will work in (low-productivity) services, and less and less people in (high-productivity) manufacturing.

To conclude, we can calculate the real income of the representative consumer.[9] To do so, replace our expression for prices into the definition of the ideal price index (Equation (8)). Then, we get that the representative consumer's real income is

$$U = \frac{wL}{P} = L\left(\sum_{i=1}^{I} z_i^{\varepsilon-1}\right)^{\frac{1}{\varepsilon-1}}. \tag{12}$$

Therefore, real income per unit of labour (equal to the real wage in this setting) is

$$\frac{U}{L} = \frac{w}{P} = \left(\sum_{i=1}^{I} z_i^{\varepsilon-1}\right)^{\frac{1}{\varepsilon-1}}. \tag{13}$$

Real income depends on the aggregate productivity of the economy: the more productive the economy is, the more goods there are to consume. Moreover, it also depends on the number of varieties I. Indeed, we can rewrite $\frac{w}{P} = \left(\sum_{i=1}^{I} z_i^{\varepsilon-1}\right)^{\frac{1}{\varepsilon-1}} = I^{\frac{1}{\varepsilon-1}} \cdot \left(\frac{1}{I}\sum_{i=1}^{I} z_i^{\varepsilon-1}\right)^{\frac{1}{\varepsilon-1}}$, showing that the real wage depends on the number of varieties (first term) and on the average productivity of each variety (second term). Note that when $\varepsilon > 1$ (that is, when goods are substitutes), the real income is increasing in the number of varieties I. That is, the consumer has a "love for variety": she is better off the more goods there are in the economy. This love for variety be important in models of trade and monopolistic competition.[10]

This simple model is the basic framework of the international trade theories that we will study. To conclude, the next section reviews three important extensions or special cases.

3 Extensions and special cases

3.1 Cobb-Douglas demand

The Cobb-Douglas utility function is a special case of CES utility, applying when the parameter ε tends to 1. It is possible to show that in this case, the utility function converges to $U = \prod_{i=1}^{I} c_i^{\frac{1}{I}}$, but that the first-order conditions are still the same as the ones of the general case, just replacing $\varepsilon = 1$.

[9] You may argue that we are not done yet, as we have not yet calculated the equilibrium price for every good. However, recall from your microeconomics classes that we cannot: there is always an infinity of price vectors which are compatible with equilibrium, and we can only determine relative prices, that is, the price of one good in terms of another. In practice, it is therefore common to normalize the price of one good, the wage, or the ideal price index P to 1. Obviously, neither of these choices affects the results.

[10] Mathematically, love for variety comes from the fact that when $\varepsilon > 1$, the function inside the sum of Equation (1) is concave, so the marginal utility of every good is decreasing with quantity. Thus, the consumer prefers to consume little quantities of a lot of goods rather than a large quantity of a few goods.

Therefore, from Equation (6), we get in the Cobb-Douglas case

$$p_n c_n = \frac{\mathcal{W}}{I}$$

for every good n. That is, with Cobb-Douglas preferences, the consumer's spending shares on each good are fixed: no matter what prices are, consumers always spend the same (in this case, an equal) percentage of their income on any given good. This is because the price elasticity of demand is equal to -1: if the price of a good increases by 1%, the quantity consumed decreases by 1%, and the total spending of the good remains exactly the same. In the general CES case, this is not true: if $\varepsilon < 1$, quantities decrease less (and the spending share of the good increases), whereas if $\varepsilon > 1$, quantities decrease more (and the spending share of the good falls).

Finally, in the Cobb-Douglas case, the ideal price index becomes

$$P = \prod_{i=1}^{I} (I p_i)^{\frac{1}{I}} .$$

Cobb-Douglas preferences make many calculations a lot easier, so we will often rely on them.

3.2 Utility weights

CES In the CES utility shown in Equation (6), all goods enter symmetrically. However, this is not necessary, and we could consider a more general utility function of the form

$$U(c_1, c_2, ..., c_I) = \left(\sum_{i=1}^{I} (\alpha_i c_i)^{\frac{\varepsilon-1}{\varepsilon}} \right)^{\frac{\varepsilon}{\varepsilon-1}},$$

where the α_is are utility weights, with a straightforward interpretation: a good with a high weight is more important to the consumer than a good with a low weight. As you will see in Problem Set 1, this generalization does not change the fundamental properties of CES utility.

Cobb-Douglas It is also possible to introduce weights in a Cobb-Douglas utility function. The generalized Cobb-Douglas utility function with weights is

$$U = \prod_{i=1}^{I} c_i^{\alpha_i}, \text{ with } \sum_{i=1}^{I} \alpha_i = 1.$$

Straightforward derivations show that then, the demand function for a given good is

$$p_n c_n = \alpha_n \mathcal{W}, \tag{14}$$

implying that the consumer always spends a fraction α_n of her income on good n, irrespective of prices. Finally, the ideal price index is now

$$P = \prod_{i=1}^{I} \left(\frac{p_i}{\alpha_i}\right)^{\alpha_i}.$$

3.3 A continuum of goods

In many cases, it will be mathematically convenient not to consider a finite number of goods I, but a continuum, an infinite and uncountable mass of goods. Fortunately, all the main properties of CES utility can be transposed to that case.

In particular, consider a set of goods indexed on the interval $[0, M]$. That is, every point of this interval represents a good, and the consumer needs to choose a consumption function c, defining for every point of the interval her desired consumption. In that case, the CES utility function becomes

$$U\left(c\left(i\right)_{i\in[0,M]}\right) = \left(\int_0^M c\left(i\right)^{\frac{\varepsilon-1}{\varepsilon}} di\right)^{\frac{\varepsilon}{\varepsilon-1}}. \tag{15}$$

Furthermore, the budget constraint becomes $\int_0^M p\left(i\right) c\left(i\right) di = \mathcal{W}$.

This problem is much more complicated than the ones we have studied before, because we need to solve for an optimal function c, and not just for a finite amount of numbers. Thus, we a priori need tools from a special branch of mathematics, optimal control theory.

However, it turns out that the solution of this problem is exactly analogous to the finite case. This is, loosely speaking, because an integral is always the limit of a (Riemann) sum, and here, some properties of that sum are carried over to the limit. In practice, therefore, one can just treat the integral in Equation (15) as if it where a sum, and derive it with respect to some particular variety n.[11] Doing this, one can repeat all the steps from sections 1.2 to 1.4, and see that nothing much changes.

[11] Mathematically speaking, this is nonsense: n is just one point in the integral, with measure 0, and therefore, $\frac{\partial U}{\partial c(n)} = 0$. However, even though this is nonsense, it works and gives correct results, for the reasons mentioned in the main text.

For any good n,

$$c(n) = p(n)^{-\varepsilon} P^{\varepsilon-1} \mathcal{W}, \tag{16}$$

where the ideal price index is now given by $P \equiv \left(\int_0^M p(i)^{1-\varepsilon} di \right)^{\frac{1}{1-\varepsilon}}$. The general equilibrium of the model also extends easily to the continuous case.

Finally, the Cobb-Douglas special case obviously exists as well for a continuum of goods. When ε tends to 1, the utility function converges to $\exp\left(\int_0^M \ln c(i) di \right)$, consumer demands are given by $c(n) p(n) = \frac{\mathcal{W}}{M}$, and the ideal price index is given by $P = \exp\left(\int_0^M \ln p(i) di \right)$.

3.4 Production functions

In this chapter, we have encountered Cobb-Douglas and CES functions as utility functions for consumers. However, it is worth noting that all of our results extend to production functions. For instance, consider a firm which produces a final output from combining a continuum of different inputs, indexed on the interval $[0, M]$, according to the production function

$$y = \left(\int_0^M m(i)^{\frac{\varepsilon-1}{\varepsilon}} di \right)^{\frac{\varepsilon}{\varepsilon-1}}, \tag{17}$$

where $m(i)$ is the amount of input i used in production. In order to behave optimally, the firm must solve a cost minimization problem: it must choose for any possible level of production y a combination of inputs which minimizes the total cost of production, $\int_0^M p(i) m(i) di$. This cost minimization problem (CMP) is extremely similar to the utility maximization problem (UMP) considered above. Indeed, we could have derived the consumer's demand functions by considering instead the CMP (the consumer's choice of the cheapest consumption bundle to achieve a given level of utility U) instead of the UMP. This would have yielded

$$c(n) = \left(\frac{p(n)}{P} \right)^{-\varepsilon} U,$$

where P is the ideal price index defined by $U = \frac{\mathcal{W}}{P}$. This is exactly Equation (16). Likewise, for the firm, the demand for a particular input n is given by

$$m(n) = \left(\frac{p(n)}{P} \right)^{-\varepsilon} y, \tag{18}$$

where $P \equiv \left(\int_0^M p(i)^{1-\varepsilon} di \right)^{\frac{1}{1-\varepsilon}}$ is the ideal price index for inputs. From this, we can see directly that the firm's input demand functions have price elasticity $-\varepsilon$. Furthermore, we can note another useful property. By the definition of the ideal price index, we have $y = \frac{\mathcal{C}}{P}$, where $\mathcal{C}$ is the total production cost of the firm. Thus, we have

$$\mathcal{C} = Py. \tag{19}$$

This shows that the CES production function has constant returns to scale (the cost function is linear in quantity). Furthermore, the marginal cost of production is given by the ideal price index P.

We know now how to solve a simple general equilibrium model with CES preferences, and with this, we are ready to start studying international trade theory. The next chapter starts this endeavour by considering the oldest and most famous theory of international trade, due to David Ricardo.

Part II

Ricardian Trade Theory and Gravity

1 Introduction

In the 18th century, economic thinking on international trade was dominated by mercantilism. Mercantilists argued that governments should promote exports and limit imports, in order to maximize their countries' holdings of gold and other precious metals.

This orthodoxy was attacked by Adam Smith in "*An Inquiry into the Nature and the Causes of the Wealth of Nations*" (Smith, 1776), the book that created modern economics. Smith underlined that "*the wealth of a country consists, not in its gold and silver only, but in its land, houses, and consumable goods of all different kinds*". The key to increase this quantity of consumable goods were specialization and the division of labour, nationally and internationally. Thus, mercantilism, which limited imports and international specialization, was inefficient. In Smith's words,

"*by means of glasses, hotbeds, and hotwalls, very good grapes can be raised in Scotland, and very good wine too can be made of them at about thirty times the expence for which at least equally good can be brought from foreign countries. Would it be a reasonable law to prohibit the importation of all foreign wines, merely to encourage the making of claret and burgundy in Scotland? [T]here would be a manifest absurdity in turning towards any employment, thirty times more of the capital and industry of the country, than would be necessary to purchase from foreign countries an equal quantity of the commodities wanted [...]*" (p. 425).

That is, as wine production costs are much higher in Scotland than in other countries, it makes no sense to use Scotland's limited resources (capital and "industry", i.e., labour) for wine. Instead, Scotland's resources should be used to produce other goods, and exchange these goods against cheap wine from foreign countries. This completely inverts the mercantilist position: Scotland's goal is not to export as much as possible, but to import the goods which could be produced only inefficiently at home. In the words of Krugman (1993), "*what a country gains from trade is the ability to import things it wants. Exports are not an objective in and of themselves: the need to export is a burden that a country must bear because its import suppliers are crass enough to demand payment.*"[12]

[12]Despite Smith's arguments, mercantilism continues to be influential. Its influence is built on arguments which are considered fallacious by most economists, such as the argument that exports create jobs and trade deficits are a signal

Smith's theory suggests that countries should import goods for which they are less productive (i.e., for which they need more resources) than the rest of the world, and pay with goods for which they are more productive. However, if this is so, what happens to countries which are more (or less) productive than the rest of the world for all goods? Do they have nothing to gain from trade? In 1817, David Ricardo picked up on this point and formulated an alternative theory (see Ricardo, 1817). Ricardo shared the basic intuition that countries should import high-cost goods and export low-cost goods, but argued that Smith was not thinking correctly about costs. Indeed, focusing on resource costs (i.e., the labour needed to produce a bottle of Scottish wine) is misleading. Scots do not consume labour: the only reason they care about using labour in wine production is because their labour resources are finite, and if they use them for wine, they cannot use them to produce other goods which they also want to consume! Thus, instead of thinking about resource costs, we need to think about opportunity costs. The relevant cost of wine is not the labour used to produce one bottle of wine, but the amount of other goods which could have been produced with this labour had it not been used for wine production. Ricardo showed that opportunity costs depend on relative productivities ("comparative advantage") rather than on absolute productivities ("absolute advantage"), and that when countries specialized according to comparative advantage, they could always gain from trade.

Ricardo's idea proved extremely influential, and shaped economists' thinking about international trade during the following two centuries. Nevertheless, Ricardian trade theory was strangely marginalized in economic teaching and research during most of the 20th century. Even though it was mentioned at the outset of any course in International Trade, it had become "*something like a family heirloom,*[13] *brought down from the attic to show a new generation of students, and then put back, allowing them to pursue more fruitful lines of study and research*" (Eaton and Kortum, 2012, p.65).

This state of affairs was due to the fact that Ricardo had just provided an example for a world with two goods and two countries. This conveyed a powerful intuition for the benefits of trade, but it was also hard to generalize to a world with many goods and many countries, and therefore, hard to apply to the real world. These limitations have eventually been overcome, due to two theoretical breakthroughs, by Dornbusch, Fischer and Samuelson (1977), and, most of all, by Eaton and Kortum (2002). Their models have lead to a boom of research in Ricardian trade theory, with researchers using the new models to test Ricardo's intuitions empirically and to quantify gains from trade in the real world. In

of bad economic health (see Krugman, 1997). However, this does not mean that economic theory implies free trade is optimal: as we will see later, countries can gain from using tariffs, and trade creates losers as well as winners.

[13] Heirloom: an object, perhaps an antique or some kind of jewellery, that has been passed down for generations through family members (Wikipedia).

this chapter, we will cover the journey of Ricardian trade theory in detail, going from David Ricardo's original intuition to the modern models and their applications.

2 David Ricardo's great intuition

2.1 Ricardo's example

Opportunity costs In his 1817 book "*On the Principles of Political Economy and Taxation*", Ricardo invited his readers to consider the following thought experiment. Imagine there are two countries (England and Portugal), producing and consuming two goods (wine and cloth).[14] These goods are homogeneous, and can be shipped across countries without any cost.

Goods are produced using labour, and there are productivity differences between countries. England can produce one liter of wine using the labour of 120 people, while Portugal can produce one liter of wine using the labour of 80 people. Furthermore, England can produce one unit of cloth using the labour of 100 people, and Portugal can produce one unit of cloth using the labour of 90 people. Thus, Portugal is more productive for both goods, which makes it impossible to apply Smith's logic: Portugal has lower costs for everything, and does not seem to need any British imports. Ricardo showed that this is incorrect: the relevant cost of wine in Portugal is not the units of labour needed to produce wine, but its opportunity cost in terms of the other good.

Precisely, the opportunity cost of wine in terms of cloth is the amount of cloth that has to be given up to produce one more liter of wine. In England, one liter of wine has an opportunity cost of 1.2 units of cloth: a liter of wine uses the labour of 120 people, which may have been used to produce $\frac{120}{100} = 1.2$ units of cloth. In Portugal, one liter of wine has an opportunity cost of 0.88 units of cloth (a liter of wine uses the labour of 80 people, which may have been used to produce $\frac{80}{90} \approx 0.88$ units of cloth instead). Likewise, a unit of cloth has an opportunity cost of $\frac{100}{120} \approx 0.83$ liters of wine in England, and an opportunity cost of $\frac{90}{80} = 1.125$ liters of wine in Portugal.

So, there are differences in opportunity costs between countries: wine is cheaper in Portugal and cloth is cheaper in England.[15] The good for which a country has a lower opportunity cost is called its comparative advantage good, and Ricardo shows that both countries gain if they exchange their comparative advantage good against the one of the other country.

[14] Ricardo may have picked Portugal because his family had Portuguese origins (hence his surname).

[15] When England has a lower opportunity cost than Portugal for good A, this directly implies that Portugal has a lower opportunity cost for good B, as these two opportunity costs are just the inverse of each other.

Gains from trade Ricardo's proof for his statement is the following: assume the Portuguese produce wine, the English produce cloth, and they agree to exchange 1 liter of wine for 1 unit of cloth on the international market.

Then, both countries are better off. Indeed, before, to get one unit of cloth, Portugal needed to use 90 units of labour. Now, it can get it cheaper: it is enough to use 80 units of labour, produce one liter of wine, and exchange it against one unit of cloth on the international market. Thus, Portugal can get wine as cheap as before, but has found a cheaper way to get cloth. Likewise, before, to get one liter of wine, England needed to use 120 units of labour. Now, it can get it for less: it is enough to use 100 units of labour, produce one unit of cloth, and exchange it against one liter of wine on the international market. Note that the gains from trade come from imports: countries gain because they can now import their comparative disadvantage good at a cost that is lower than the one of domestic production. This turns on its head the mercantilist notion that imports are bad (a sign of weakness) and exports are good (a sign of strength).

Ricardo's example provides an elegant and powerful principle, showing that countries can take advantage of productivity differences by specializing. Most importantly, it shows that we should not think about specialization in terms of absolute productivity differences, but in terms of relative differences, that is, in terms of opportunity costs. This principle extends beyond international trade, and is probably one of the most important insights of all time in economic theory.

However, Ricardo's example is also incomplete. For instance, he just assumes that the world market price of wine in terms of cloth is 1, without explaining where this price comes from! Instead, economic theory suggests prices should be determined by supply and demand. So, in the next section, we will transform Ricardo's example into a true general equilibrium model.

2.2 A simple model with two goods and two countries

2.2.1 Completing Ricardo's example

Ricardo's example is missing many elements of our simple general equilibrium models in Chapter I. First, it does not specify a demand side. So, let us assume that English and Portuguese consumers have Cobb-Douglas preferences with equal spending shares over wine and cloth, and inelastically supply L_E and L_P units of labour. Then, demand functions are given by

$$p_C c_{C,E} = p_W c_{W,E} = \frac{1}{2} w_E L_E$$

in England, and an analogous expression in Portugal. The supply side of Ricardo's example is already more complete, as we have production functions for wine and cloth in both countries. On top of that, we assume that there is perfect competition.

Finally, note that we continue to assume that goods can be traded without costs. Thus, the law of one price applies, and each good has a unique price, p_C or p_W, which applies in both countries. On the other hand, we assume that labour cannot move across countries. Therefore, English wages may be different from Portuguese ones. This structure is very common in international trade theory: in all models in this class, we will assume that goods can move across countries, but production factors cannot.[16] Now, we are ready to analyse the equilibrium outcome in Ricardo's example.

2.2.2 The equilibrium with international trade

It turns out that the best way to solve for the equilibrium is to try to figure out who produces what. Perfect competition implies that English and Portuguese producers sell at marginal cost. These marginal costs are pinned down by wages and productivities:

$$\begin{aligned} mc_{W,E} &= 120w_E \qquad & mc_{W,P} &= 80w_P \\ mc_{C,E} &= 100w_E \qquad & mc_{C,P} &= 90w_P \end{aligned}$$

On the world market, workers buy only from the cheapest producer. Accordingly, they buy their wine only from English producers when $120w_E < 80w_P$, only from Portuguese producers if $120w_E > 80w_P$, and they are indifferent between both if $120w_E = 80w_P$. The same obviously holds for cloth, and therefore, we can deduce what each country produces as a function of the relative wage, w_E/w_P.

With this, we can deduce the sets of goods produced in England and in Portugal, I_E and I_P:

$$I_E = \begin{cases} \{\} & \text{if } \frac{w_E}{w_P} > \frac{90}{100} \\ \{\text{Cloth}\} & \text{if } \frac{80}{120} < \frac{w_E}{w_P} \leq \frac{90}{100} \\ \{\text{Cloth, Wine}\} & \text{if } \frac{w_E}{w_P} \leq \frac{80}{120} \end{cases} \quad I_P = \begin{cases} \{\} & \text{if } \frac{w_E}{w_P} < \frac{80}{120} \\ \{\text{Wine}\} & \text{if } \frac{80}{120} \leq \frac{w_E}{w_P} < \frac{90}{100} \\ \{\text{Wine, Cloth}\} & \text{if } \frac{w_E}{w_P} \geq \frac{90}{100} \end{cases}$$

This simple analysis already provides several insights. First, it shows that there are three possible

[16] In more advanced models, we will consider transport costs and show how they lead to price differences across countries. Note furthermore that in Ricardo's example, the world would be even better off if workers could move between countries. Then, all English workers would move to Portugal, and all goods in the world would be produced with the superior Portuguese technologies. Therefore, in Ricardian models, trade is in general just a second-best solution with respect to a first-best world in which workers could move between countries.

equilibria:

1. An equilibrium with $\frac{80}{120} < \frac{w_E}{w_P} < \frac{90}{100}$, in which England produces only cloth, and Portugal produces only wine. This is called a complete specialization equilibrium.

2. An equilibrium with $\frac{w_E}{w_P} = \frac{90}{100}$, in which England produces only cloth, but Portugal produces both wine and cloth.

3. An equilibrium with $\frac{w_E}{w_P} = \frac{80}{120}$, in which England produces wine and cloth, but Portugal produces only wine. This equilibrium (just as the one in 2) is called an incomplete specialization equilibrium.

All other values of relative real wages cannot occur in equilibrium. Indeed, when $\frac{w_E}{w_P} > \frac{90}{100}$, England would not produce anything, but that cannot be: then, labour demand in England would be zero, and the English labour market would not clear (likewise, when $\frac{w_E}{w_P} < \frac{80}{120}$, the Portuguese labour market would not clear).

Now, which one of these three equilibria applies? This depends on consumers' preferences for wine and cloth, and on the relative number of workers in England and Portugal. Intuitively, complete specialization is more likely if the countries are more or less equally large and if consumers more or less equally like both goods. Indeed, you can see intuitively that if one country is very large, and the other country very small, complete specialization is impossible. Imagine, for example, that England has a population of 1 million workers, and Portugal only a population of one worker. It is obvious that this one worker, even if she produces a lot of wine, cannot satisfy the wine demands of 1 million English consumers: so, in this case, the equilibrium is likely to be the one of Case 3, in which England continues to produce some wine itself.

In order to determine which equilibrium applies, we therefore need to do case distinctions. That is, we would guess that we are in the complete specialization equilibrium, solve the model under this assumption, and then check whether the relative wage we found is indeed between $\frac{80}{120}$ and $\frac{90}{100}$. If it is, our initial guess was correct and we found the equilibrium, if not, we are in one of the incomplete specialization equilibria (and can determine which one by the same guess-and-verify procedure). You will do this in a Problem Set. However, we can still make a number of key observations here:

a. Countries specialize according to comparative advantage. In each of these three cases, England exports cloth to Portugal, and Portugal exports wine to England. So, the market equilibrium actually realizes the exchange possibility highlighted by Ricardo.

This result is obvious in the complete specialization case: each country produces only one good, but its consumers want to consume both of them. It is also clear under incomplete specialization: even in these equilibria, there is always one country which produces only one good. This country must import the good it does not produce, and to pay for these imports, export the good it does produce.

b. Countries gain from trade iff they fully specialize. In this simple example, one can show that a country gains from trade just by comparing the budget set of its workers in autarky and under free trade. For instance, consider an English worker in autarky: she earns an income w_E, and with one unit of that income, she can buy $1/120$ liters of wine (recall that the price of wine is equal to its marginal cost, which is in autarky $120w_E$) or $1/100$ units of cloth.

Now, consider that same worker under free trade, and in a case of complete specialization, where England produces only cloth. The worker can still buy $1/100$ units of cloth with one unit of her wage: the wage is w_E, and the price of cloth, produced in England, is $100w_E$. Wine, on the other hand, is now produced in Portugal, and its price is the Portuguese marginal cost, $80w_P$. Thus, the English worker can buy $\frac{w_E}{80w_P}$ liters of wine. However, in an equilibrium with complete specialization for England, $\frac{w_E}{w_P} > \frac{80}{120}$, so that $\frac{w_E}{80w_P} > \frac{1}{120}$. This proves that England is better off under free trade than under autarky, because the purchasing power of English workers (their real wage) has increased: with every unit of their wage, they can buy as much cloth as before, but more wine. As their budget set has become larger, they must achieve a higher utility.

Note that this argument holds for any consumer preferences. In your Problem Set, you can verify that for the Cobb-Douglas preferences specified above, every consumer indeed has a strictly higher utility under free trade iff her country completely specializes.

Finally, note that if a country does not completely specialize, then its consumers are indifferent between autarky and free trade. In the example above, England not completely specializing implies $\frac{w_E}{w_P} = \frac{80}{120}$, and its workers can buy exactly as much with their autarky and their free trade wages.

c. Relative wages reflect relative productivities. We have shown that in equilibrium, it must be that $\frac{80}{120} \leq \frac{w_E}{w_P} \leq \frac{90}{100}$, that is, that England's wages are lower than Portugal's. This is a necessary consequence of Portugal being more productive in both industries. If wages were the same in both countries, Portugal would produce everything: but this is clearly not an equilibrium, as it would leave all English labour unemployed. So, English wages have to be lower in order to make England competitive at least for one industry (and the first industry in which it will become competitive is precisely the one for which it has a comparative advantage, i.e. the one in which its productivity deficit with respect to Portugal is smallest).

Thus, the Ricardian model predicts that less productive countries compete through lower wages. However, in the model, this low-wage competition is bad neither for the "poor" low-wage country nor for the "rich" high-wage country: both are better off trading with each other than in autarky.

This analysis shows that it is possible to translate Ricardo's intuitions into a simple equilibrium model. However, it also shows that this is cumbersome: even for a simple case with two countries and two goods, three cases need to be distinguished.

Once the model is extended to more goods or more countries, the number of cases multiplies, and the model gets more and more difficult to solve. Moreover, the whole concept of Ricardian comparative advantage becomes unclear: Ricardo taught us that England has a comparative advantage in a good if its opportunity cost is lower than the one of Portugal. However, imagine there is a third country, France. Then, does Portugal have a comparative advantage if its opportunity cost is lower than the one of Portugal? The one of France? Some weighted average of Portugal and France?

These difficulties have reduced the practical usefulness of Ricardian models for many years. Economists believed that Ricardo's insights were very important, but they did not know how to apply them to real-world trade, with its many countries and many goods. These limitations were overcome when the model was successively extended to feature more than two goods, and more than two countries. The first step towards this was done in 1977 by Dornbusch, Fischer and Samuelson, who considered a model with two countries and an infinity of goods. The beauty of their model lies in the realization that the Ricardian model is difficult when there is a countable number of goods, but very simple if there is an uncountable number of them.

3 The Dornbusch-Fischer-Samuelson model

3.1 Assumptions

The model of Dornbusch, Fischer and Samuelson (1977) introduces only one change to our previous model: instead of two goods, it assumes that England and Portugal trade a continuum of goods, indexed on the interval $[0, 1]$. That is, every point of that interval represents an industry whose products can be traded. For industry i and each country j, there is a production function which takes the same form as before:

$$\forall i \in [0,1]\,,\ y_j\left(i\right) = z_j\left(i\right) l_j\left(i\right)$$

From this, we can define a function T as

$$\begin{aligned} T: \quad & [0,1] \to \mathbb{R}_+^* \\ & i \mapsto \tfrac{z_E(i)}{z_P(i)} \end{aligned}$$

T gives, for every industry, the relative productivity of England with respect to Portugal. Dornbusch, Fischer and Samuelson make two assumptions about this function:

- T is decreasing in i. This means that industry 0 is the one in which England has the largest relative productivity advantage over Portugal, and industry 1 the one where it has the largest relative disadvantage. This assumption is not restrictive: we can always order industries in such a way that T is decreasing.

- T is continuous on its domain (this is just a technical assumption to keep the analysis simple).

Note that it may be that $T\left(0\right) < 1$. In that case, Portugal is more productive than England in all industries, just as in Ricardo's original example.

Furthermore, we denote the number of workers in a country by L_j, and to keep demand functions as simple as possible, we assume all workers have identical Cobb-Douglas preferences. As shown in Chapter I, this implies that for any good i,

$$p\left(i\right) c_j\left(i\right) = w_j L_j.$$

3.2 The equilibrium

As before, we start by asking which goods are produced in which country, noting that a country produces a good if and only if its marginal cost is lower than the one of the other country. For example, a given good i is produced in England if and only $\frac{w_E}{z_E(i)} \leq \frac{w_P}{z_P(i)}$. Accordingly, the set of industries produced in England is characterized by

$$i \in I_E \Longleftrightarrow \frac{w_E}{w_P} \leq \frac{z_E\left(i\right)}{z_P\left(i\right)} = T\left(i\right)$$

This shows that countries specialize according to their relative productivities, summarized by the function T (and not according to absolute productivities). As T is strictly decreasing, good i being cheaper in England than in Portugal implies that England also has lower costs for all goods j with $j < i$. Therefore, for any given level of the relative wage which holds $\frac{w_E}{w_P} \in (T(1), T(0))$, there exists a cut-off industry $i^* \in (0,1)$ so that England produces all goods of the interval $[0, i^*]$ and Portugal produces all goods of the interval $[i^*, 1]$.[17] i^* is defined by

$$\frac{w_E}{w_P} = T(i^*) \tag{20}$$

Equation (20) is the first fundamental relationship of the Dornbusch-Fischer-Samuelson model. It gives the equilibrium relative English wage as an decreasing function of the cut-off industry i^*, expressing the fact that if England wants to expand to the number of industries it produces in, then its relative wage needs to fall to make it more competitive.

Note that it is indeterminate which country produces the cut-off industry i^*, for which costs are equal. However, this indeterminacy does not matter, as this industry has mass 0 (it is just a point in a continuum of industries). This is precisely what makes the model so simple: as there is a continuum of industries, we do not have to worry about cases where the two countries "share" an industry any more!

To close the model, we need to now turn to the demand side.

In this model, a country's national income in this model is just its labour income, as labour is the only factor of production and firms do not make profits. Furthermore, it must also be true that the total labour income is equal to the total sales of all the country's firms: as firms do not make profits, all their revenue must eventually be paid to their workers. Therefore, for England,

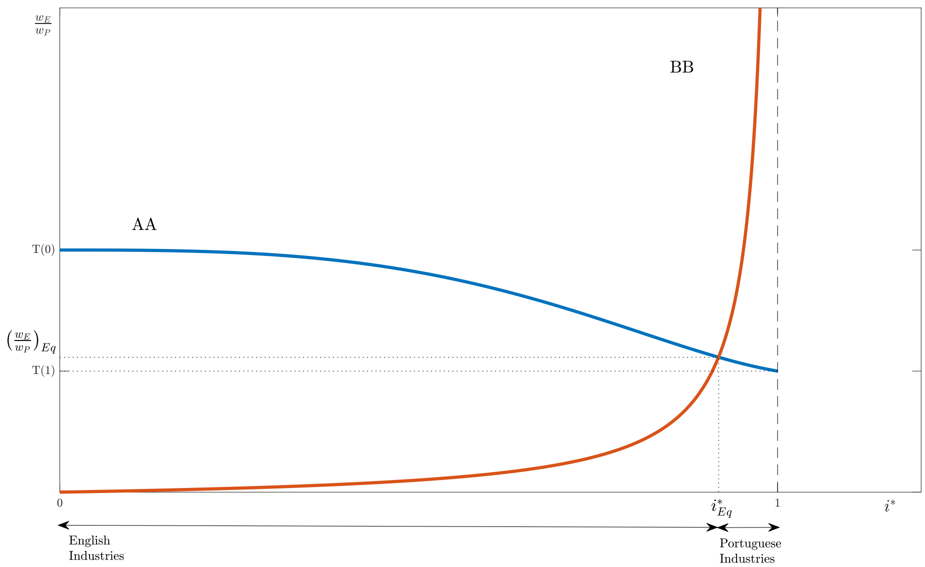
$$w_E L_E = \int_0^{i^*} p(i)\, c(i)\, di.$$

$c(i)$, the worldwide consumption of good i, is the sum of English and Portuguese consumption. Then, Cobb-Douglas demand implies that $p(i)\, c(i) = w_E L_E + w_P L_P$. Replacing this in the formula above, we get

$$w_E L_E = i^* (w_E L_E + w_P L_P). \tag{21}$$

[17] This is easy to prove: as T is continuous and decreasing on its domain, it takes every value between $T(0)$ and $T(1)$ exactly once. Therefore, Equation (20) has a unique solution on $(0,1)$.

Figure 1: Equilibrium in the Dornbusch-Fischer-Samuelson model



Repeating the same steps for Portugal gives

$$w_P L_P = (1 - i^*)\,(w_E L_E + w_P L_P)\,. \tag{22}$$

Finally, dividing Equation (21) by Equation (22) yields

$$\frac{w_E}{w_P} = \frac{i^*}{1 - i^*}\frac{L_P}{L_E} \tag{23}$$

Equation (23) is the second fundamental relationship of the model. It gives the equilibrium relative English wage as an increasing function of the cut-off industry i^*. This reflects that the fact that when England produces many industries, its firms collect a lot of revenue, and are able to pay their workers higher wages.

We have two equations, (20) and (23), in two unknowns, i^* and $\frac{w_E}{w_P}$: by putting them together, we can solve the model. Figure 1 illustrates the equilibrium of the Dornbusch-Fischer-Samuelson model, plotting the two relationships defined by Equations (20) and (23) in the $\left(i^*, \left(\frac{w_E}{w_P}\right)\right)$ space.

It is easy to prove that this model always has a unique equilibrium.[18] In this equilibrium, all basic

[18]As the BB curve takes value 0 at 0, and tends to positive infinity at 1, it must necessarily intersect the AA curve. Furthermore, as the AA curve is decreasing and the BB curve increasing, they intersect exactly once.

insights from Ricardo's two-country example still apply.

a. Countries specialize according to comparative advantage. By definition of the T-function, the goods that a country produces are the ones for which it has the highest relative productivities with respect to the other country (that is, the lowest opportunity costs), just as in the basic two-goods case. Thus, it's relative and not absolute productivities that matter for specialization.

b. Countries gain from trade. In the Dornbusch-Fischer-Samuelson model, all equilibria feature complete specialization (essentially, because the infinite number of goods now provides countries much more opportunities to specialize). Thus, in all equilibria, both countries gain from trade, which can be shown with the same simple method that we have used in the two-good model.

c. Relative wages reflect relative productivities. Thus, for instance, when $T(0) < 1$, meaning England is less productive in all industries, $w_E/w_P < 1$.

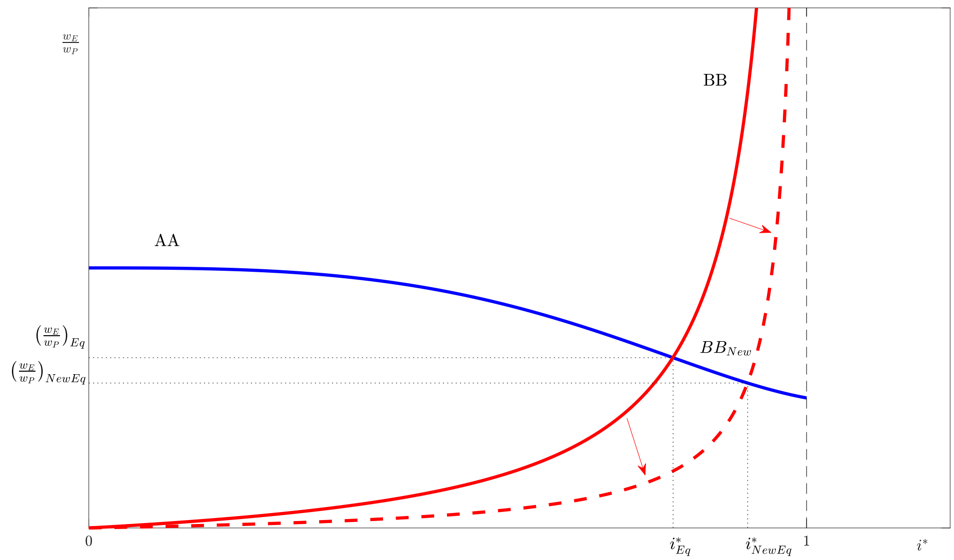
Note that we have not solved the model completely yet: to solve for the exact quantities produced in every country and for their real wages, we need to specify both countries' productivity distributions. You will do this in Problem Set 2. However, already like this, the model is much more convenient for analysis than the one from the previous section, mainly because we need not do case distinctions any more. Therefore, it can easily be used to analyse the effects of changes in parameters: how do, for example, population growth and changes in productivity affect trade and the welfare of trading countries?

3.3 Comparative statics

What happens when England gets larger? Suppose the English population increases, so that the English labour endowment increases from L_E to $2L_E$, while the Portuguese population remains unchanged. What is the effect of this on trade and specialization?

From the analysis above, it is straightforward to see that the AA curve does not move, as labour endowments do not appear in Equation (20). However, the BB curve shifts towards the right, as shown in Figure 2. Therefore, the equilibrium relative wage falls, and the equilibrium cutoff i^* increases, meaning that England now produces in more industries than before.

Figure 2: The effect of an increase in England's population



What is happening here? To analyse this, it is useful to decompose it into two effects that occur with extreme productivity distributions. This is shown in the two panels of Figure 3.

The upper panel shows a case in which productivity differences between England and Portugal are infinitely large: there is a set, say half, of industries for which England has productivity 0 (i.e., some industries in which England cannot produce), and for the other half, Portugal has productivity 0. Therefore, the AA curve is just a vertical line. In this situation of infinite productivity differences, an increase in the English population does not affect the specialization of the two countries, but only reduces the English relative wage. In technical terms, England does **factor deepening**: it just uses more labour for the production of the same goods it used before. As England produces more goods, but the production of Portuguese goods remains unchanged, the prices of English goods and therefore the wages of English workers have to fall. England's greater production deteriorates its terms of trade.

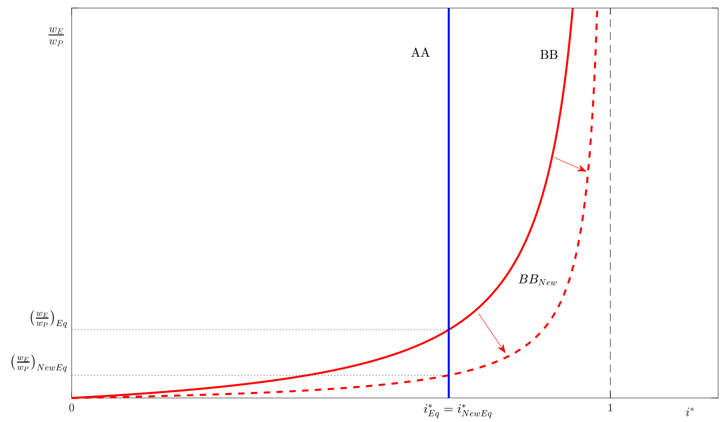
The lower panel shows the exact opposite case, in which there are no productivity differences between England and Portugal: the AA curve is just a horizontal line (crossing the y-axis at 1).[19] In this case, an increase in the English population just increases the mass of industries produced by England, but leaves relative wages unaffected. In technical terms, England now reacts to the population increase by **structural transformation**: it uses the new workers only to produce goods it has not produced

[19]Note that there are no gains from trade in this case, as there are no technological differences between countries.

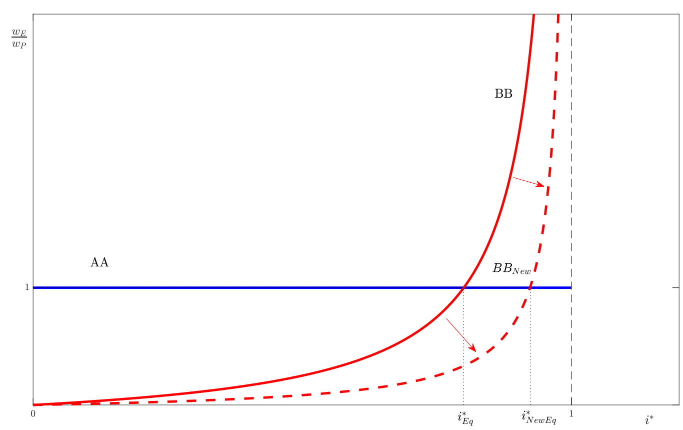
before. Indeed, the relative wage can never be different from one here (otherwise one country would not produce at all), so all the adjustment must come from a change in production patterns.

Figure 3: A decomposition: two extreme cases

Extreme productivity differences



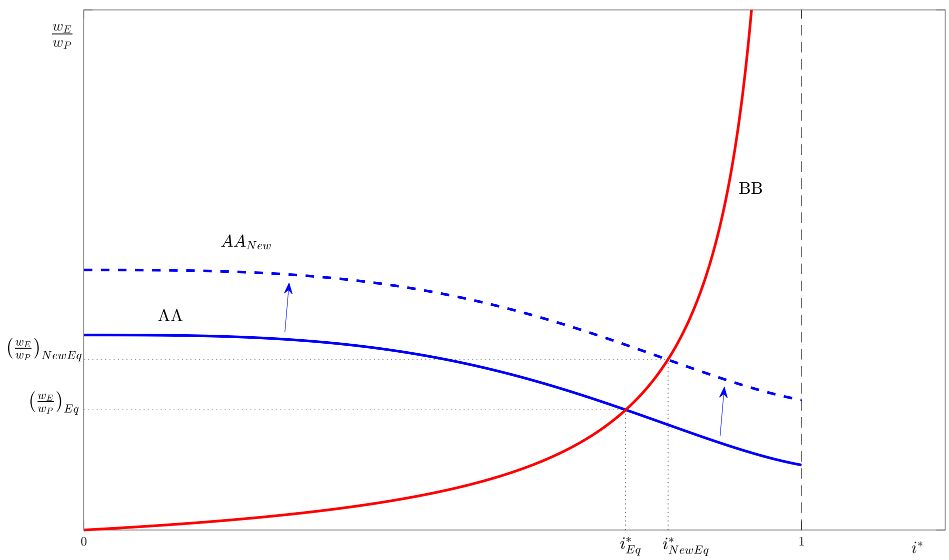
No productivity differences



In less extreme cases, the adjustment of England to its larger population is a combination of factor deepening and structural transformation: England both employs more labour in its "old" industries (which lowers its relative wages) and increases the range of goods it produces (and therefore its share in world income).

What happens when England gets more productive? Now, assume the English make some great invention, and are suddenly able to produce with productivity $2z_E\left(i\right)$ instead of $z_E\left(i\right)$ in every industry. The effect of this change is illustrated in Figure 4.

Figure 4: The effect of an English productivity increase in all industries



This change leaves the BB curve unaffected (indeed, productivities do not appear in Equation (23)). However, it shifts the AA line upwards, and as a result, both England's relative wage and the mass of industries it produces in increase. England is now able to do structural transformation without seeing a fall in relative wages: as it is now a little better in all industries, its wages have to rise relative to the Portuguese ones to restore equilibrium.

The Dornbusch-Fischer-Samuelson model is a very elegant and simple formalization of Ricardo's intuition, and allowed us to derive some interesting results. However, there are still only two countries in the model, which limits its applicability to the real world, and does not resolve the difficulty of thinking about comparative advantage in a world with many countries.

These difficulties were finally resolved by Eaton and Kortum (2002), who extended the Dornbusch-Fischer-Samuelson model to a case with many countries, triggering a boom for research in Ricardian trade theory. We know turn to study their model, before considering its many applications.

4 The Eaton-Kortum model

4.1 Aims and assumptions

In the Dornbusch-Fischer-Samuelson model, there were no restrictions on the productivity distributions z_E and z_P, which could be any real-valued functions. Eaton and Kortum were able to extent the model to a case with many countries by sacrificing this generality, assuming instead that productivity distributions take a very specific function form, the density function of a Fréchet distribution. This technical assumption had huge pay-offs: not only did it allow Eaton and Kortum to solve a Ricardian model with many countries quite easily, but it also yielded predictions for trade which were closely in line with the real world. This showed that the technical Fréchet assumption does not seem to be that much of a restriction.

The basic setup of the Eaton-Kortum model is the same as in Dornbusch-Fischer-Samuelson, except that there are now N countries instead of only two. These countries consume a continuum of goods indexed on $[0, 1]$, which are produced using labour. Firms in country j can produce good i with the production function

$$y_j(i) = z_j(i)\, l_j(i).$$

We also continue to assume that workers in all countries have identical Cobb-Douglas preferences,[20] so that the demand function for good i in country j is

$$p_j(i)\, c_j(i) = w_j L_j.$$

However, two things are new with respect to the Dornbusch-Fischer-Samuelson model.

1. First, we introduce trade costs: transport costs and tariffs. This is not an innovation of Eaton and Kortum (the original Dornbusch-Fischer-Samuelson model already has trade costs, we left them out in the previous section for the sake of simplicity). However, trade costs are crucial if the model is supposed to match real-world data, such as the fact that there is more trade between countries that are geographically close, or that countries with lower tariffs trade more.

   Trade costs are modelled as iceberg costs, a popular device invented by Samuelson (1954). That is, we assume that for delivering one unit of a good from country j to country n, it is necessary to ship d_{nj} units of the good, with $d_{nj} \geq 1$: the rest "melts away" during transit.

[20] Eaton and Kortum use the more general CES preferences. To simplify calculations, we stick with the easier Cobb-Douglas case here, but Section 4.2.4 and the Appendix discuss the model's solution with CES.

Trade costs are not necessarily symmetric, so d_{nj} may be different from d_{jn}. This is because trade costs do not only include transport costs (which in most cases should be symmetric), but also tariffs and other barriers to trade (which are not necessarily symmetric: for instance, country j may have higher import tariffs than country n). Furthermore, trade costs are independent of the good i that is shipped. This is clearly not true in reality: some goods are easier to transport than others, and tariffs differ across goods. However, this makes the model a little more tractable. There are more sophisticated versions of the Eaton-Kortum model which relax this assumption.

2. The main innovation of Eaton and Kortum, however, is to assume that the productivities $z_j(i)$ are, in every country, realisations of a random variable Z_j following a Fréchet distribution. This means that the cumulative distribution function F_j of each country's productivity is given by

$$\forall z \in \mathbb{R}_+^*,\ F_j(z) = \mathbb{P}(Z_j \leq z) = \exp\left(-T_j z^{-\theta}\right), \tag{24}$$

where T_j and θ are two positive parameters. To see the role of these parameters, it is helpful to consider the natural logarithm of the Fréchet distribution. The expectation of $\ln Z_j$ is $\ln \gamma + \frac{1}{\theta} \ln T_j$, and its standard deviation is $\frac{\pi}{\theta\sqrt{6}}$, with π being the circle constant (3.14159...) and γ being Euler's constant (0.577...). As there is a continuum of goods, this corresponds to the average productivity and the dispersion of productivities for every country. Thus, average productivity in country j is increasing in T_j. Furthermore, productivity dispersion is decreasing in θ: the higher is θ, the less productivities will be spread out across goods (and in the extreme case where $\theta \to +\infty$, all goods in country j would be produced with the same productivity).

   Therefore, technologies across the world can be described with a limited set of parameters: one T_j (capturing the average productivity level) for every country, and a unique θ for the worldwide productivity dispersion.

Note that as the productivity of an individual good is random, the basic Eaton-Kortum model does not predict whether a particular country should export or import a particular good. Instead, the model will only be able to make broader predictions about how much a country exports and imports in total. However, this already yields several insights about the patterns of trade and its welfare consequences. Furthermore, subsequent research has shown how the Eaton-Kortum model can be extended to bring back concrete predictions about exports and imports in particular industries.

4.2 Solving the model

As before, perfect competition implies that firms sell their goods at marginal cost. Therefore, for consumers in country n, the costs for importing good i from country j are

$$h_{nj}(i) = d_{nj}\frac{w_j}{z_j(i)}. \tag{25}$$

Equation (25) is also valid for goods produced at home, and we assume that $d_{nn} = 1$, that is, that there are no trade costs within countries. Note that trade costs now matter for consumer's decisions from where to buy: even if country j has low wages and a high productivity for good i, it may not be the cheapest source for consumers of country n if country n is far away from country j, or applies high tariffs to imports from country j.

4.2.1 Some results on cost distributions

h_{nj} is a random variable, i.i.d. across goods. Its cumulative distribution function, G_{nj}, is

$$\begin{aligned} \forall h \in \mathbb{R}_+^*,\ G_{nj}(h) &= \mathbb{P}\left(\tfrac{w_j d_{nj}}{Z_j} \leq h\right) \\ &= \mathbb{P}\left(Z_j \geq \tfrac{w_j d_{nj}}{h}\right) \\ &= 1 - \mathbb{P}\left(Z_j < \tfrac{w_j d_{nj}}{h}\right) \end{aligned}$$

$$\Rightarrow G_{nj}(h) = 1 - \exp\left(-T_j h^{\theta}(w_j d_{nj})^{-\theta}\right). \tag{26}$$

In the end, consumers from every country only buy a given good from the cheapest source. This cheapest source is not necessarily the same for all countries, because there are trade costs. For instance, maybe Mexico has the lowest wages and highest productivity for cars in the world, but Italians still buy their cars in France, which compensates its higher wages and lower productivity by lower trade costs, being closer to Italy and a EU member. This implies that, in contrast to the simple models studied before, there is no longer a unique world market price for every good.

The lowest cost for good i in country n is denoted $h_n(i)$, and is equal to the price of good i in country n. It holds, by definition, $p_n(i) = h_n(i) = \min_{j\in\{1,\ldots,N\}}(h_{nj}(i))$. Therefore, it also is an i.i.d. distributed random variable, and its cumulative distribution function is

$$
\begin{aligned}
\forall c \in \mathbb{R}_{+}^{*},\ G_n(h) &= \mathbb{P}\left(\left(\tfrac{w_1 d_{n1}}{Z_1} \leq h\right) \cup \left(\tfrac{w_2 d_{n2}}{Z_2} \leq h\right) \cup ... \cup \left(\tfrac{w_N d_{nN}}{Z_N} \leq h\right)\right) \\
&= 1 - \mathbb{P}\left(\overline{\left(\tfrac{w_1 d_{n1}}{Z_1} \leq h\right) \cup \left(\tfrac{w_2 d_{n2}}{Z_2} \leq h\right) \cup ... \cup \left(\tfrac{w_N d_{nN}}{Z_N} \leq h\right)}\right) \\
&= 1 - \mathbb{P}\left(\left(\tfrac{w_1 d_{n1}}{Z_1} > h\right) \cap \left(\tfrac{w_2 d_{n2}}{Z_2} > h\right) \cap ... \cap \left(\tfrac{w_N d_{nN}}{Z_N} > h\right)\right) \\
&= 1 - \textstyle\prod_{j=1}^{N} \exp\left(-T_j h^{\theta} \left(w_j d_{nj}\right)^{-\theta}\right).
\end{aligned},
$$

$$
\Rightarrow G_n(h) = 1 - \exp\left(-\Phi_n h^{\theta}\right), \tag{27}
$$

where $\Phi_n \equiv \sum_{j=1}^{N} T_j \left(w_j d_{nj}\right)^{-\theta}$.[21] This gives us the distribution of prices in every country (as a function of the wages of every country in the world, for which we still have to solve). We can now proceed to derive one of the most important formulas of the Eaton-Kortum model.

4.2.2 Import probabilities and spending shares

What is the probability that a country n actually buys a good i from another country j? This probability is independent of the good under consideration, and holds[22]

$$
\begin{aligned}
\pi_{nj} &= \mathbb{P}\left(\tfrac{w_j d_{nj}}{Z_j} \leq \min_{j' \neq j}\left(\tfrac{w_{j'} d_{nj'}}{Z_{j'}}\right)\right) \\
&= \int_0^{+\infty} \mathbb{P}\left(h \leq \min_{j' \neq j}\left(\frac{w_{j'} d_{nj'}}{Z_{j'}}\right)\right) g_{nj}(h)\, dh \\
&= \int_0^{+\infty} \exp\left(-\sum_{j' \neq j} T_{j'} h^{\theta} \left(w_{j'} d_{nj'}\right)^{-\theta}\right) \exp\left(-T_j h^{\theta} \left(w_j d_{nj}\right)^{-\theta}\right) T_j \theta h^{\theta-1} \left(w_j d_{nj}\right)^{-\theta} dh \\
&= T_j \theta \left(w_j d_{nj}\right)^{-\theta} \int_0^{+\infty} \exp\left(-h^{\theta} \Phi_n\right) h^{\theta-1} dh
\end{aligned},
$$

$$
\Rightarrow \pi_{nj} = \frac{T_j \left(w_j d_{nj}\right)^{-\theta}}{\Phi_n} \tag{28}
$$

Intuitively, π_{nj} is increasing in T_j (if country j is on average more productive, then country n will import a lot from j), but decreasing in w_j and d_{nj} (high wages and transport costs obviously make country j a less attractive source of imports).

Because of the law of large numbers, π_{nj} is exactly equal to the fraction of the overall set of goods that country n imports from country j (and π_{nn} is the fraction of goods which country n produces for its own consumption). As each country has Cobb-Douglas preferences, π_{nj} is also the share of country

[21] Above, we use De Morgan's law ($\overline{A \cup B} = \overline{A} \cap \overline{B}$) to pass from the second to the third line, and the fact that productivity draws across countries are i.i.d. to pass from the third to the fourth line.

[22] In the following derivations, g_{nj} is the density corresponding to the distribution function G_{nj}. To get from the first to the second line, we use the law of total probability for continuous random variables: for two random variables X, Y, $\mathbb{P}(X < Y) = \int \mathbb{P}(x < Y) f_X(x)\, dx$, where f_X is the density of X. To get from the second to the third line, we use the results from above.

n's national income which is spent on products from country j. Indeed, we have seen in Chapter 1 that with Cobb-Douglas preferences, consumers equally split their income across all goods. Thus, if consumers buy 5% of their goods from China, they also spend 5% of their income on Chinese goods. So, once we have solved for the level of wages around the world, Equation (28) gives us every country's trade patterns. This is as precise as the basic Eaton-Kortum model's trade predictions get: the model will tell us that 50% of Italy's GDP is spend on Italian goods, 5% on Chinese goods, 4% on French goods, etc., but not which exactly which goods Italy produces at home or imports from China. It is now time to proceed and solve for the wages.

4.2.3 Wages

To find wages, we can use the same reasoning as in the Dornbusch-Fischer-Samuelson model. Any country j's national income is equal both to its labour income, $w_j L_j$, and to the total sales of country j firms. Therefore, we get

$$\forall j, \quad w_j L_j = \sum_{n=1}^{N} \pi_{nj} w_n L_n. \tag{29}$$

This gives us a system of N equations (of which $N-1$ are independent) with N unknowns. By normalizing $w_N = 1$ (as usual, we can only solve for relative prices in General Equilibrium and may therefore normalize one price to 1), we therefore get an equation system which can be solved on a computer (see Problem Set 2). This gives us the relative wages and the trade patterns for every country.

There is only one element missing: to assess workers' welfare, we want to calculate the real wage, equal to per-capita utility. For this, we need to determine for each country the value of the ideal Cobb-Douglas price index P_n. In each country,

$$P_n = \exp\left(\int_0^1 \ln p_n(i) di\right)$$

As we know the price distribution in every country (given by Equation (27)), one can calculate this integral and show that it equals $P_n = \frac{1}{\gamma}\Phi_n^{-\frac{1}{\theta}}$, where γ stands again for Euler's constant.[23] From this, we get that the real wage in every country is

$$\frac{w_n}{P_n} = \gamma w_n \Phi_n^{\frac{1}{\theta}} = \gamma \left(\frac{T_n}{\pi_{nn}}\right)^{\frac{1}{\theta}} \tag{30}$$

[23]Appendix VIII provides a derivation for this formula.

Equation (30) shows that the real wage of a country can be expressed as a function of the parameters of the Fréchet distribution, and of the (endogenous) share of national income spent on home-produced goods. This simple formula will be crucial for much of what follows.

4.2.4 Generalization to CES preferences

For simplicity, we have assumed throughout that consumers have Cobb-Douglas preferences. However, results can easily be generalized to CES preferences.

We have used Cobb-Douglas preferences just twice in the above derivations. First, they implied that if consumers buy 5% of their goods from China, they also spend 5% of their income on Chinese goods. With CES, this is not true: for instance, if China charged very low prices, consumers would spend more than 5% of their income on Chinese goods (if the elasticity ε is larger than 1), or less than 5% (if the elasticity is smaller than 1). However, because of the special properties of the Fréchet distribution, it turns out that the price distribution of goods is the exactly the same for all source countries! Therefore, even in the CES case, π_{nj} as given by Equation (28) is equal to the share of country n income spent on country j goods. Second, we used Cobb-Douglas preferences to derive the ideal price index P_n. Here as well, results are very similar with CES: the ideal price index now equals $P_n = \left(\Gamma\left(1+\frac{1-\varepsilon}{\theta}\right)\right)^{\frac{1}{1-\varepsilon}} \Phi_n^{-\frac{1}{\theta}}$, where Γ stands for the Gamma function. Just like γ above, $\Gamma\left(1+\frac{1-\varepsilon}{\theta}\right)$ is a constant which drops out of all relevant results. Appendix VIII formally shows both of these results.

We have now solved the Eaton-Kortum model (or at least, shown how it can be solved on a computer, which you will actually do in Problem Set 2). In the next section, we will see how this model relates to a fundamental empirical concept, the gravity equation.

5 The gravity equation

5.1 Origins and "naive" gravity

In 1962, the Dutch economist Jan Tinbergen discovered a surprising result: international trade follows a law which is very similar to Newton's law of universal gravitation (Tinbergen, 1962). Newton's law says that the force of gravity between two objects is proportional to the product of the masses of the two objects, and inversely proportional to the square of the distance between them. Tinbergen hypothesized that a similar relationship might hold for international trade, with $X_{jn} \approx A\frac{Y_j^{\alpha_1} Y_n^{\alpha_2}}{d_{jn}^{\alpha_3}}$, where X_{jn} is a measure of trade between countries j and n (for instance, exports from j to n, or imports of j from n), Y is GDP, and d_{jn} the physical distance between j and n. In logarithms, this yielded an equation which was easy to estimate by OLS:

$$\ln(X_{jn}) = \ln(A) + \alpha_1 \ln(Y_j) + \alpha_2 \ln(Y_n) - \alpha_3 \ln(d_{jn}) + \nu_{jn}, \tag{31}$$

where ν_{jn} is an error term. Tinbergen's estimation yielded $\alpha_1 \approx \alpha_2 \approx \alpha_3 \approx 1$. Moreover, the R^2 of the regression was high, showing that country size and distance explain a lot of variation in trade flows. These results suggests a remarkable empirical regularity: trade between two countries is proportional to the product of their GDPs, and inversely proportional to the distance between them. Two examples can make this more salient. Consider two countries A and B which are at the same distance from Italy, with country A having a GDP that is twice as high as country B. Then, the gravity equation predicts that Italy trades twice as much with country A than with country B. Alternatively, consider two countries C and D which have the same GDP, with country C being twice as far from Italy than country D. Then, the gravity equation predicts that Italy trades half as much with country C than with country D.

Figure 5: Trade flows and GDP: the case of Japan

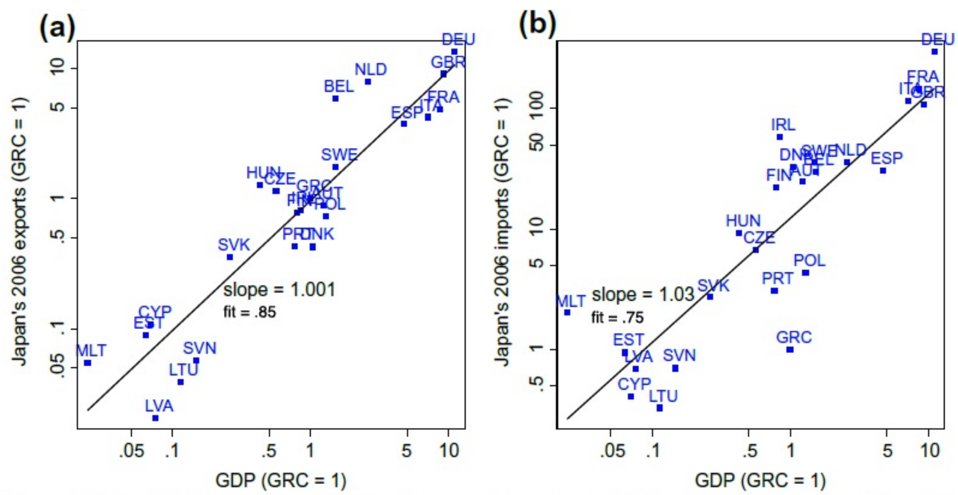


Figure 3.1 Trade is Proportional to Size; (a) Japan's Exports to EU, 2006; (b) Japan's Imports from EU, 2006. GRC: Greece

Source: Head and Mayer (2014).

Tinbergen's results have been reproduced many times since (although later researchers typically found a smaller coefficient on distance). For instance, Figure 5 plots the logarithm of Japan's imports and exports in 2006 against the logarithm of the GDP of its trading partners. The slope of both relationships is almost exactly 1, just as in Tinbergen's regressions.

Figure 6 shows the role of distance, by plotting the logarithm of France's 2006 imports and exports against a measure of geographical distance. There is a clear negative relationship, with a slope coefficient between 0.7 and 0.9.

Figure 6: Trade flows and distance: the case of France

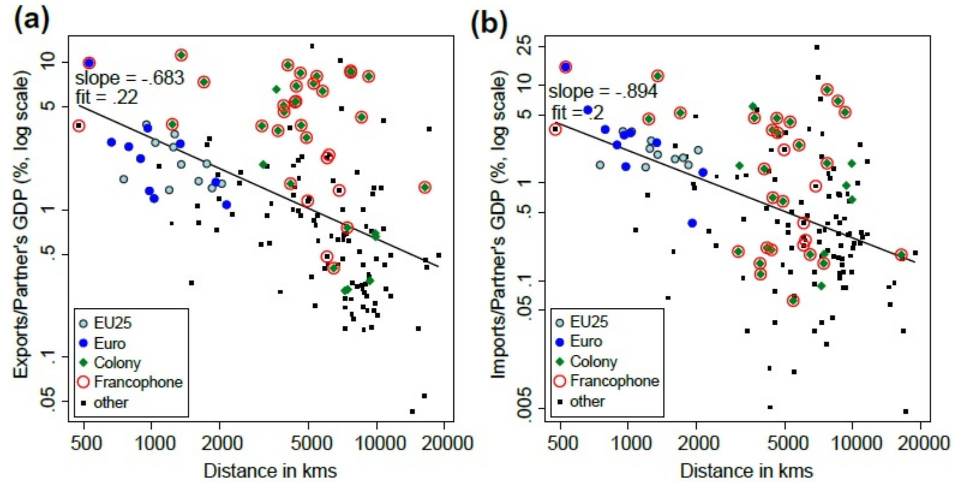


Figure 3.2 Trade is Inversely Proportional to Distance; (a) France's Exports (2006); (b) France's Imports (2006)

Source: Head and Mayer (2014).

While country size and distance matter for international trade, they are not the only factors. For instance, Figure 6 also shows that, controlling for distance, France trades more with French-speaking countries and with former colonies. To incorporate these factors, researchers after Tinbergen added more variables to Equation (31), such as dummies for countries sharing the same language, a colonial history, or having a free trade agreement. One famous study in this tradition is McCallum (1995). McCallum analysed trade between Canadian provinces and US States, and found that a Canadian province trades 22 times as much with another Canadian province than with a US State of the same GDP and at the same distance. He interpreted this as suggesting that even though both countries have a free trade agreement and (for most parts) a common language, there is a strong "border effect" discouraging trade.

Studies in the Tinbergen tradition have uncovered important regularities. However, they also have methodological problems. Indeed, Equation (31) is just a correlation, which is hard to interpret. For instance, suppose we introduced a dummy variable for free trade agreements, and estimated a positive coefficient for this variable. Does this mean that free trade agreements (FTAs) actually increase trade between countries? The answer is not necessarily: the positive coefficient may just be driven by an omitted variable. For instance, Italy trades more with Belgium than with Serbia. This could be because Belgium is a member of the European Union and Serbia is not, but also because Belgium is richer than Serbia, or because Belgium has historically closer ties to Italy. A researcher could potentially control for these issues by introducing, say, GDP per capita in the regression, but this is an imperfect fix: maybe richer countries are more likely to be members of FTAs in the first place. Then, the FTA dummy and GDP per capita are correlated, and coefficients are hard to interpret.

Moreover, doing comparative statics on the Tinbergen equation is impossible. That is, even if there is no omitted variable problem, a FTA is likely to affect GDP, and as GDP affects trade, the FTA will have both a direct and an indirect effect on trade. However, the gravity equation cannot tell us anything about the size of this indirect effect.[24]

These limitations have reduced the usefulness of the gravity equation for a long time: gravity was an interesting correlation, but hard to interpret or to use for analyzing actual trade policies. This changed in the 2000s, as we will discuss in the next section.

5.2 Multilateral resistance and relation to theory

The breakthrough in gravity research is due to Anderson and van Wincoop (2003). They point out that in an N-country Armington model, trade between two countries follows a gravity equation, and show how to estimate its parameters in a consistent way.[25]

The model of Anderson and van Wincoop is not the only one that delivers a gravity equation. In fact, their work made it clear that any trade model holding certain conditions delivers a gravity equation.[26] This is also the case for the Eaton-Kortum model, which we now use to derive the equation.

By definition, the share of the national income of country n spent on goods of country j is Y_{nj}/Y_n, where Y_{nj} stands for the spending of country n on goods from country j, and Y_n stands for the national income of country n (i.e., its GDP). As we have shown in Section 4, this share is given by Equation (28). Therefore, we get

$$Y_{nj} = Y_n \frac{T_j \left(w_j d_{nj}\right)^{-\theta}}{\Phi_n}. \tag{32}$$

[24] The gravity equation is also cross-sectional and cannot be used for comparisons over time: if Germany's and Italy's GDP double between 1950 and 2018, it does not follow that trade between Italy and Germany grows by a factor of $2 \cdot 2 = 4$. While the constant A is unchanged in the cross-section, there is no reason it should be unchanged over time.

[25] The Armington model is named after Armington (1969), and was introduced in Problem Set 1. It assumes that each country produces exactly one good, and that consumers want to consume goods from every country. Building on this model, Anderson provided a theorical foundation for the gravity equation 25 years before his paper with van Wincoop (see Anderson, 1979). However, this earlier paper did not focus on the econometric implications.

[26] The most important condition is that consumers have CES preferences. See Feenstra (2016) for details.

Summing this expression over n gives

$$Y_j = T_j w_j^{-\theta} \sum_{n=1}^{N} \frac{d_{nj}^{-\theta} Y_n}{\Phi_n}, \tag{33}$$

where we have used the fact that $\sum_{n=1}^{N} Y_{nj} = Y_j$, that is, that the national income of country j is equal to the sum of the spending of all countries in the world (including j itself) on country j goods. Then, replacing Equation (33) into Equation (32), we get

$$Y_{nj} = \frac{Y_n Y_j}{d_{nj}^{\theta}} \frac{1}{\Phi_n \chi_j}, \tag{34}$$

with $\chi_j = \sum_{n'=1}^{N} \frac{d_{n'j}^{-\theta} Y_{n'}}{\Phi_{n'}}$. This is a gravity equation: it tells us that trade between two countries is proportional to the product of their GDPs, and inversely proportional to the trade cost (taken to the power θ). Taking logarithms, Equation (34) becomes

$$\ln Y_{nj} = \ln Y_n + \ln Y_j - \theta \ln d_{nj} - \ln \Phi_n - \ln \chi_j. \tag{35}$$

This equation is similar to Tinbergen's naive gravity equation, but there are several key differences. First, the elasticity of trade to trade costs (a broader concept than the geographical distance considered by Tinbergen) is constant, but no longer equal to 1. Second, there are two extra terms, Φ_n and χ_j. Anderson and van Wincoop (2003) refer to these as "multilateral resistance terms". Intuitively, they absorb any factor which make a country export/import more to/from all other countries of the world. Indeed, in a general equilibrium model such as the Eaton-Kortum one, trade between two countries does not depend on the absolute value of trade costs or distance, but "*on the bilateral barrier between them relative to average trade barriers that both regions face with all their trading partners*". When a researcher does not account for this by controlling for multilateral resistance terms, her results will be biased.

To illustrate this, Anderson and van Wincoop (2003) go back to the McCallum (1995) study on border effects between the US and Canada. They show that McCallum's result is largely driven by the fact that he considers a "naive" gravity equation, without controlling for multilateral resistance. Revisiting McCallum's estimation, they find that borders reduce trade between the US and Canada by 44%, a much smaller effect than the one claimed in McCallum's paper.[27]

[27] To illustrate the bias, Anderson and van Wincoop present a model with international and intraregional trade, and

These insights help to clarify where the gravity equation comes from, and how it should be formulated. They also lend some empirical support to models such as Eaton-Kortum, by showing that they are at least consistent with some key correlations observed in the data (that is, the fact that Eaton-Kortum delivers a gravity equation can make us more confident that it is a "good" model).

However, just considering Equation (35) by itself does not yet allow us to make quantitative statements. Ultimately, we want to analyze how a change in trade costs d_{nj} affects trade between countries n and j, as well as the welfare of their citizens. To do so, the next section shows how we can consistently estimate the parameter θ. Then, Section 6 shows how we can combine these estimates with the Eaton-Kortum model for a quantitative analysis of trade policy (or other changes in trade costs).

5.3 Estimating the gravity equation

The key parameter in the gravity equation is θ, generally referred to as the **trade elasticity** (because it is the elasticity of bilateral trade to bilateral trade costs). Over time, researchers have proposed various ways to estimate this parameter.

Fixed effects The most straightforward way to estimate Equation (35) is to use OLS with fixed effects for the multilateral resistance terms (that is, two dummies for every country in the estimation sample). The first paper to do so - without yet using the theoretical justification of Anderson and van Wincoop - was Harrigan (1996).

This approach is easy to implement, but it faces one problem: there is no readily available measure of trade costs (d_{nj}), which should include tariffs, transport costs and other barriers to trade. The standard solution to this issue is to specify a reduced-form model for trade costs, assuming

$$d_{nj} = (1 + t_{nj}) \exp\left(\alpha_1 x_{1,nj} + \alpha_2 x_{2,nj} + ... + \alpha_K x_{K,nj}\right), \tag{36}$$

where t_{nj} stands for the tariff on exports from country j to country n, and $x_{1,nj}$, ..., $x_{K,nj}$ are other variables which affect the cost of exporting from j to n (such as distance, a common language, etc.).

compare a small country (Canada) and a large country (the United States). Assume both Canada and the United States face the same trade cost with respect to the rest of the world. This depresses international trade more for Canada, which because of its size would have had a high trade-to-GDP ratio under free trade (we will return to this point below). So, Canada needs to compensate more for the "missing" international trade by interregional trade. In other words, Canadian provinces trade more than US states, but this is not because of the US-Canada border. Formally, the multilateral resistance terms for US states are different than the ones for Canadian provinces (lower, because US states have access to a larger internal market), but McCallum's estimation had implicitly assumed they are the same.

Then, the gravity equation becomes

$$\ln Y_{nj} = \lambda_n + \lambda_j - \theta \ln (1 + t_{nj}) - \theta (\alpha_1 x_{1,nj} + \alpha_2 x_{2,nj} + ... + \alpha_K x_{K,nj}), \tag{37}$$

where λ_n and λ_j are importer and exporter fixed effects. The trade elasticity is identified by the coefficient on tariffs.[28] Note that GDP levels drop out, as they are absorbed by fixed effects.

This method is simple and easy to implement. Its potential drawback is that the level of bilateral tariffs (or other factors influencing trade costs) could be correlated with the fixed effects. Also, the model of how trade costs depend on tariffs and other factors is quite arbitrary.

An alternative dealing with the latter issue was proposed by Eaton and Kortum (2002), and refined by Simonovska and Waugh (2014). These papers use data on prices. Indeed, consider a good from country j which is exported to n. The Eaton-Kortum model indicates that the price of this good on its home market should be $p_j(i) = \frac{w_j}{z_j(i)}$, and the price on the export market should be $p_n(i) = \frac{d_{nj} w_j}{z_j(i)}$. That is, the trade cost can be backed out as $d_{nj} = \frac{p_n(i)}{p_j(i)}$. This method is more elegant than just assuming an ad-hoc model for trade costs. However, it relies on having high-quality price data for the same goods across different countries, which is very difficult to collect.

Ratio methods Starting with Head and Ries (1999), several researchers have developed so-called ratio methods. The idea behind these is to eliminate multilateral resistance terms by taking ratios, so that there is no longer a need for fixed effects. Part of the rationale for these methods was computational: standard softwares such as Stata just could not compute models with a very high number of fixed effects until recently.[29] However, the methods also have some econometric advantages.

To illustrate ratio methods, we consider here the approach due to Romalis (2007). Consider the Eaton-Kortum gravity equation given by Equation (34). Then, for two countries of origin i and j and two countries of destination n and k, we can write

$$\frac{Y_{ni}}{Y_{ki}} = \frac{Y_n}{Y_k} \left(\frac{d_{ki}}{d_{ni}}\right)^{\theta} \frac{\Phi_k}{\Phi_n} \quad \text{and} \quad \frac{Y_{nj}}{Y_{kj}} = \frac{Y_n}{Y_k} \left(\frac{d_{kj}}{d_{nj}}\right)^{\theta} \frac{\Phi_k}{\Phi_n}. \tag{38}$$

Taking the ratio between these two expressions, we get

$$\frac{Y_{ni}}{Y_{ki}} \frac{Y_{kj}}{Y_{nj}} = \left(\frac{d_{ki}}{d_{ni}} \frac{d_{nj}}{d_{kj}}\right)^{\theta}, \tag{39}$$

[28] We need the functional form for trade costs in Equation (36). If we instead included tariffs among the x variables, the trade elasticity θ would not be identified: there would be $K+1$ parameters, but only K estimated coefficients.

[29] This issue was resolved by Correia (2014).

an expression without any multilateral resistance terms. To estimate θ, we can use the same trade cost model as in (36). However, Romalis (2007) introduces a further twist. Suppose that countries i and n signed a FTA at a certain point in time, and that this FTA only affected tariffs. Then, taking logs and differences over time, Equation (39) becomes

$$(\Delta \ln (Y_{ni}) - \Delta \ln (Y_{ki})) - (\Delta \ln (Y_{nj}) - \Delta \ln (Y_{kj})) = -\theta \ln (1 + t_{ni}), \tag{40}$$

where $\Delta \ln (Y_{ni}) = \ln \left(Y_{ni}^{\text{After FTA}}\right) - \ln \left(Y_{ni}^{\text{Before FTA}}\right)$, and t_{ni} is the pre-FTA tariff for exports from i to n. As all other trade costs are not affected by the FTA, they drop out. Equation (40) is a double difference-in-differences specification: to assess the effect of export tariffs from i to n, we compare export growth from i to n (relative to export growth from i to another destination) to export growth from another country j to n (again, relative to export growth from j to another destination). This excludes that results are driven by any exporter or importer-specific unobserved factor.

Recently, Caliendo and Parro (2015) have developed another variation of this approach, relying on triple differences. While these approaches are elegant and easy to implement, they also have issues. Most importantly, using FTAs faces the issue that these are endogenous (i.e., a FTA may be signed because exports in a particular industry grow, not the other way around). The standard way to deal with this issue is to try to find instrumental variables for FTAs (see Baier and Bergstrand (2007) and Baier, Bergstrand, Egger and McLaughlin (2008) for some examples).

Taking stock Head and Mayer (2014) provide an overview of different methods to estimate gravity equations. Results vary across methodologies, with a median between 4 and 5. The next section explains why this number is important, and how it matters for questions we are actually interested in, such as the size of gains from trade, or the consequences of Brexit for British living standards.

6 Quantitative models: trade theory with numbers

6.1 The ACR formula

6.1.1 Gains from trade

What are the gain from trade in the Eaton-Kortum model? That is, what is the difference in consumer welfare between a world with trade and a world in which every country exists in autarky? The real wage of a country in the Eaton-Kortum model is given by Equation (30). In autarky, a country spends 100% of its income on domestic goods, and therefore

$$\left(\frac{w_n}{P_n}\right)_{\text{Autarky}} = \gamma T_n^{\frac{1}{\theta}}.$$

Therefore, the ratio of real wages with and without trade is just

$$\frac{\left({}^{w_n}/_{P_n}\right)_{\text{Trade}}}{\left({}^{w_n}/_{P_n}\right)_{\text{Autarky}}} = \left(\pi_{nn}\right)^{-\frac{1}{\theta}}. \tag{41}$$

Thus, gains from trade can be calculated with only two numbers: the share of national income spent on a country's own goods, and the value of the productivity dispersion parameter θ. As $\pi_{nn} \leq 1$, all countries gain from opening up to trade (as long as this leads to them doing some actual trade). Note also that if the parameter θ is very large (meaning productivity dispersion is low), gains from trade are almost zero. Indeed, low productivity dispersion within countries means that a country produces essentially all goods with its average productivity T_j. Then, there are no differences in opportunity costs across countries, hence no comparative advantage and no international specialization.

The share of national income spent on a country's own goods is very easy to observe in the data: it is just one minus the import share of GDP. θ, on the other hand is hard to observed directly. However, as we have seen in the previous section, θ is equal to the trade elasticity in a gravity equation. Therefore, we can use estimated gravity equations to set the value of this parameter.

The fourth column of Table 1 (taken from Eaton and Kortum, 2012) shows the gains from trade which are obtained with this approach for OECD countries.[30] The table shows that small countries win most from trade. This is a mechanical consequence of them having the highest import shares of GDP, and it is a very general feature of Ricardian models.[31] Intuitively, this is a revealed preference result: if a country trades a lot, it must be that it gets a lot of benefits from trading. In a Ricardian model, small countries can fully specialize in the production of very few goods in which they have very high relative productivity, while large countries need to continue producing a larger set of goods in order to provide all different goods in sufficient quantity to their population. Thus, small countries have better terms of trade, and gain more.

Table 1 also indicates that countries which are geographically isolated gain less than those which are

[30] The numbers shown in the table actually come from a slightly more complicated version of the Eaton-Kortum model, which has intermediate inputs and a non-tradable sector.

[31] Recall, for instance, that the Dornbusch-Fischer-Samuelson model predicts that the terms of trade turn against a country that is large or getting larger, as it needs a lower wage to employ its large population.

more central (compare Portugal and Switzerland, for instance). For an isolated country, the possibility to trade does not change much if trade costs are prohibitively high.

Table 1: Gains from trade in the Eaton and Kortum model

Table 2

The Home Share of Spending on Manufactures and Gains from Trade

Country	World GDP share (%) in 2006	*Home share of spending*		*Implied gains from trade*	
		Level in 2006 (%)	*Change since 1996 (percentage points)*	*Level in 2006 (%)*	*Change since 1996 (percentage points)*
Austria	0.66	31.4	–16.2	21.3	8.1
Canada	2.60	49.1	–1.5	12.6	0.6
Czech Republic	0.29	42.6	–14.7	15.3	5.5
Denmark	0.56	25.6	–18.1	25.5	10.7
Estonia	0.03	2.5	–19.6	85.4	56.7
Finland	0.42	58.2	–7.3	9.4	2.1
France	4.60	56.9	–10.3	9.9	3.0
Germany	5.94	53.7	–16.4	10.9	4.8
Greece	0.54	52.7	–11.6	11.3	3.6
Hungary	0.23	26.0	–34.5	25.1	16.4
Iceland	0.03	27.9	–10.0	23.7	6.2
Ireland	0.46	39.6	9.9	16.7	–5.7
Italy	3.80	68.9	–7.1	6.4	1.7
Japan	8.88	84.9	–5.6	2.8	1.1
Korea	1.94	77.2	–0.7	4.4	0.1
Mexico	1.94	58.3	–7.9	9.4	2.3
New Zealand	0.22	53.6	–8.2	11.0	2.6
Norway	0.68	51.9	–2.5	11.6	0.9
Poland	0.69	53.4	–15.8	11.0	4.7
Portugal	0.41	50.8	–10.2	12.0	3.4
Slovenia	0.08	27.2	–15.5	24.3	9.0
Spain	2.51	62.8	–10.2	8.1	2.7
Sweden	0.81	49.2	–10.0	12.5	3.4
Switzerland	0.80	35.3	–20.0	18.9	8.6
United States	27.26	73.5	–8.3	5.3	1.9
All others	33.62				

Source: Authors' calculations from the OECD STAN (STructural ANalysis) Database, the Economist Intelligence Unit, and a model described in the text.

Notes: The home share is the share a country spends on domestic manufactures out of total country spending on manufactures. The last two columns calculate the implications of the level of the home share, and its changes over time, for countries' gains from trade and how those gains have evolved. We look at the gains from trade only in manufactures.

Source: Eaton and Kortum (2012).

Equation (41) is generally referred to as the ACR formula, because it was first developed in an important paper by Arkolakis, Costinot and Rodriguez-Clare (2012). It is remarkable not only for its simplicity, but also for its generality. Indeed, other models, such as the Armington model (see Problem Set 2) or the Chaney-Melitz model (see Chapter V), admit a very similar formula,[32] which also takes the form

$$\frac{\left({}^{w_n}/_{P_n}\right)_{\text{Trade}}}{\left({}^{w_n}/_{P_n}\right)_{\text{Autarky}}} = (\pi_{nn})^{-\frac{1}{\text{trade elasticity}}}. \tag{42}$$

[32]This is very closely related to the fact that these models also generate a gravity equation. For further details, see Arkolakis, Costinot and Rodriguez-Clare (2012) and Costinot and Rodriguez-Clare (2014).

In each of these models, the trade elasticity corresponds to a different structural parameter (demand elasticity in the Armington model, productivity dispersion in Eaton-Kortum, and shape of the Pareto productivity distribution in Chaney-Melitz). However, for a researcher who just wants to calculate the gains from trade, these underlying differences are irrelevant. No matter what the underlying model is, the trade elasticity can be correctly inferred from estimating the gravity equation in all three cases, and then the gains from trade predicted by all three models are exactly the same.

While this result is remarkable, it should also not be exaggerated: it does not mean that all trade models are the same and that their assumptions do not matter. As we will see shortly, once we introduce differences in demand elasticities between sectors, or intermediate inputs, the simple formula in Equation (42) does not hold any more: then, different models yield different gains from trade, and we need more information on more than just two numbers to calculate them. Still, the ACR result shows that most of the modern standard trade models share a common structure, and that therefore one can get at their quantitative predictions with very similar methods.

6.1.2 Changes in trade costs

Estimates for gains from trade are interesting, but not very useful: there are essentially no autarkic countries, or countries seriously contemplating a return to autarky. Instead, we would like to know the answers to more relevant questions, e.g.: What would be the consequences of a 5% increase in trade costs between Great Britain and the EU? The Eaton-Kortum model can provide an answer to this question as well, using a limited amount of data and assumptions.

To start, notice that the ACR formula can be generalized to assess the impact of any change in model parameters (for instance, a change in trade costs). The change in real wages between the current and the new equilibrium is then given by

$$\frac{\left({}^{w_n}/{}_{P_n}\right)_{\text{After}}}{\left({}^{w_n}/{}_{P_n}\right)_{\text{Before}}} = \left(\frac{\pi_{nn}^{\text{After}}}{\pi_{nn}^{\text{Before}}}\right)^{-\frac{1}{\theta}}. \tag{43}$$

So, knowing the trade elasticity θ, the import share is a sufficient statistic to assess the welfare consequences of a change in trade costs. For instance, in a blog entry from April 2016, Paul Krugman wrote that if Brexit would reduce the share of EU countries in British trade from 50% to 33% (the level before EU accession), the British import share would fall from 30% of GDP to 25%.[33] Then, using the ACR formula with $\theta = 4$ indicates that British real wages would fall by 1.7%.

[33] See http://krugman.blogs.nytimes.com/2016/04/23/boris-is-bad-enough/.

This rough calculation, while useful, is based on an arbitrary assumption about the evolution of the import share. To be more precise, we can use the model to get a prediction for what the import share should be with higher trade costs. Appendix VIII shows how the basic Eaton-Kortum model can be used for such a calculation. However, this model still contains many simplifying assumptions, which one would like to relax for a full quantitative analysis of an event such as Brexit or a trade war between China and the United States. Most importantly, it considers only trade in final goods (ignoring the complex global value chains and input-output networks which characterize our current period of globalization), and ignores tariffs (which are not just deadweight losses such as the iceberg costs d, but create some income for the countries imposing them). Thus, in the next section, we will discuss the role of these elements and develop a full model containing them. We are then ready to use this machinery to analyse a range of currrent questions in trade policy.

6.2 The relative changes methodology

The principle What does the Eaton-Kortum model imply for Brexit? The most straightforward approach to answer this question would just be to solve for the equilibrium of the model in a “before Brexit” and an “after Brexit” scenario, and compare the levels of wages, GDP and trade in both scenarios. To do so, we would need to assume values for all parameters of the model: the average productivity parameter $(T_n)_{n=1}^N$ for every country in the world, the productivity dispersion θ, and the level of trade costs d_{nj} between any pair of countries (n, j), before and after Brexit.

This approach is straightforward, but it is also demanding, as some parameters correspond to concepts which are hard to measure. Therefore, researchers instead rely on a relative changes methodology. Indeed, it turns out that in order to compute the statistics that we are most interested in (e.g., the change in real wages after Brexit), we only need to know a subset of model parameters.

Model equations Let us denote the matrix of pre-Brexit trade costs by (d_{nj}), and the matrix of post-Brexit trade costs by $\left(d'_{nj}\right)$. Using Equation (28), we get

$$\frac{\pi'_{nj}}{\pi_{nj}} = \frac{T_j \left(w'_j d'_{nj}\right)^{-\theta} / T_j \left(w_j d_{nj}\right)^{-\theta}}{\sum_{i=1}^N T_i \left(w'_i d'_{ni}\right)^{-\theta} / \sum_{i=1}^N T_i (w_i d_{ni})^{-\theta}},$$

where all variables with a $'$ refer to the post-Brexit equilibrium. Denoting the proportional change in any variable after the change in trade costs by $\hat{v} \equiv \frac{v'}{v}$, and doing some algebra, we get

$$\hat{\pi}_{nj} = \frac{\left(\hat{w}_j \hat{d}_{nj}\right)^{-\theta}}{\sum_{i=1}^{N} \frac{T_i (w_i d_{ni})^{-\theta} (w_i d_{ni})^{\theta} \left(w'_i d'_{ni}\right)^{-\theta}}{\sum_{s=1}^{N} T_s (w_s d_{ns})^{-\theta}}} = \frac{\left(\hat{w}_j \hat{d}_{nj}\right)^{-\theta}}{\sum_{i=1}^{N} \pi_{ni} \left(\hat{w}_i \hat{d}_{ni}\right)^{-\theta}}. \tag{44}$$

Using the fact that $\hat{w}_j = \hat{Y}_j$ (as population is not affected by the change in trade costs) and that $Y'_j = \sum_{n=1}^{N} \pi'_{nj} Y'_n$ (national income is equal to spending on nationally produced goods), we finally get

$$\hat{Y}_j Y_j = \sum_{n=1}^{N} \hat{\pi}_{nj} \pi_{nj} \hat{Y}_n Y_n = \sum_{n=1}^{N} \frac{\pi_{nj} \left(\hat{Y}_j \hat{d}_{nj}\right)^{-\theta} \hat{Y}_n Y_n.}{\sum_{i=1}^{N} \pi_{ni} \left(\hat{Y}_i \hat{d}_{ni}\right)^{-\theta}}. \tag{45}$$

If we know (a) the pre-Brexit trade shares π_{nj} for every pair of countries n and j, (b) the pre-Brexit GDP levels Y_n for every country n, (c) the trade elasticity θ and (d) the changes in trade costs caused by Brexit, $\hat{d}_{nj}$, for every pair of countries n and j, then Equation (45) defines a system of N equations with N unknowns. This allows us to solve (using a computer) for the changes in GDP triggered by Brexit in each country, $\hat{Y}_n$. Once we have these, we can plug them into Equation (44), and we get the changes in trade shares triggered by Brexit. Then, we can finally plug the change in trade shares into the ACR formula, and we get the changes in real wages triggered by Brexit.

Proceeding this way is attractive, as we need to make only very few assumptions on parameter values. In particular, we do not need to take any stand on the average productivity parameters $(T_n)_{n=1}^{N}$ or on the level of trade costs d_{nj}. Instead, we can rely on easily observable statistics, such as GDPs and trade shares. The only parameters we need are the trade elasticity θ (which we get from estimating a gravity equation) and the changes in trade costs $\left(\hat{d}_{nj}\right)$ triggered by Brexit (which we can get by making assumptions about how Brexit will change tariffs and other trade costs).

We are now fully equipped to use the basic Eaton-Kortum model for a quantitative analysis of trade policy. However, the Eaton-Kortum model abstracts from many issues that are relevant in the real world. Therefore, we consider in the next section three key extensions of the basic model that are typically included in quantitative studies.

6.3 Important extensions of the basic Eaton-Kortum model

In this section, we discuss three important extensions of the Eaton-Kortum model: different sectors, intermediate inputs, and tariffs. These three extensions are incorporated in Caliendo and Parro (2015), a leading example for a quantitative trade model building on the Eaton-Kortum framework. Here, we discuss the three extensions separately and in a simplified form. Appendix VIII contains the equilibrium conditions of the full Caliendo-Parro model.

6.3.1 Sectors and industries

The basic Eaton-Kortum model presented in Section 4 assumes that all goods are of the same “type”: there is a constant elasticity of substitution ε across all goods, and all goods face the same trade costs. This is obviously unrealistic. When analyzing real-world trade policies, we want to have some heterogeneity in trade costs. For instance, manufacturing goods (T-shirts, cars, smartphones,...) are much easier to export than services (haircuts, cleaning services, legal advice,...), and we want our model to reflect that. Furthermore, while goods from the same industry (e.g., T-shirts from different brands) are easy to substitute, goods from different industries are much harder to substitute. Our model also needs to take this into account in order not to give absurd predictions: if trade costs for T-shirts fall, consumers will probably not massively substitute T-shirts for their haircuts.

To introduce these issues, we can use a two-tier CES structure. For instance, assume that consumers consume final goods from I sectors, with Cobb-Douglas preferences. Then, the utility of the representative consumer of country n is

$$U_n = \prod_{i=1}^{I} \left(C_n^i\right)^{\alpha^i}, \tag{46}$$

with $\alpha^i > 0$ and $\sum_i \alpha^i = 1$. In each sector i, there is a unit mass of varieties, indexed on $[0, 1]$. These varieties are assembled under perfect competition with a CES production function

$$C_n^i = \left(\int_0^1 \left(c_n^i(\omega)\right)^{\frac{\varepsilon^i - 1}{\varepsilon^i}} d\omega \right)^{\frac{\varepsilon^i}{\varepsilon^i - 1}}. \tag{47}$$

We assume that $\varepsilon^i > 1$. That is, it is easier to substitute between different varieties of the same sector than between the final goods of different sectors. For each variety, production is given by $y_n^i(\omega) = z_n^i(\omega)\, l_n^i(\omega)$. In each sector, productivities are realizations of a Fréchet distribution, with parameters (T_n^i, θ^i), and trade costs are given by a matrix d_{nj}^i. Thus, trade costs and the productivity distribution now differ across sectors.

The overall structure of this model is very similar to the baseline Eaton-Kortum model, and equilibrium conditions can be derived exactly as in Section 4. However, we now have a much more flexible model, accounting for heterogeneity between goods.

6.3.2 Intermediate goods and global value chains

Relevance A defining feature of the current globalization is the emergence of "global value chains", with the production process for many goods being split up across the world. As a consequence, many countries do not export only final goods, but intermediate inputs to be used in production (and maybe exported again) in another country. The WTO has estimated that in 2011, 49% of world trade in goods and services took place within such global value chains (WTO, 2015). These input-output linkages are very important for issues of trade policy: for example, if the United States imposes a tariff on car imports from Mexico, this does not only affect the Mexican factories building these cars, but also the US firms that sell them intermediate inputs. Moreover, as we will see now, ignoring intermediate inputs would lead us to substantially underestimate gains from trade.

The intermediate inputs multiplier Consider the Eaton-Kortum model presented in Section 4, but assume that the production function for good i in country j is

$$y_j(i) = z_j(i)\left(l_j(i)\right)^{\beta}\left(M_j(i)\right)^{1-\beta}, \quad \text{with } M_j(i) = \exp\left(\int_0^1 \ln m_j(i')\, di'\right). \tag{48}$$

That is, production now does not only use labour, but a Cobb-Douglas combination of labour and an intermediate input index $M_j(i)$. This intermediate input index itself is a Cobb-Douglas aggregate of all goods in the economy. Using our results from Chapter I, we can show that the cost for consumers of country n buying good i from country j is given by

$$h_{nj}(i) = d_{nj}\frac{\left(\frac{w_j}{\beta}\right)^{\beta}\left(\frac{P_j}{1-\beta}\right)^{1-\beta}}{z_j(i)}. \tag{49}$$

That is, production costs now not only depend on wages, but also on the aggregate price index of the producing economy.[34] This is because the way intermediate inputs are aggregated in production is exactly the same than the way they are aggregated in consumer utility. Note that this is a simplistic production structure: it means that all goods use intermediate inputs in exactly the same proportions, and in the same proportions as in aggregate consumption. This is obviously not true in reality (and the full Caliendo-Parro model in the Appendix allows for more realistic production structures).

[34] To derive Equation (49), proceed in two steps. First, derive the optimal demand for each intermediate for a given level of aggregate intermediate purchases $M_j(i)$. This yields an Equation similar to (18) (but with $\varepsilon = 1$, as we are considering the Cobb-Douglas case), and shows that the price of one unit of the intermediate index $M_j(i)$ is P_j. Then, we can solve for the optimal choice of labour and aggregate intermediate input, using the same methods.

With this simple modification, the equilibrium conditions of the basic Eaton-Kortum model are almost unchanged: we just need to replace the wage w_j by the cost index $\left(\frac{w_j}{\beta}\right)^{\beta}\left(\frac{P_j}{1-\beta}\right)^{1-\beta}$. Therefore, we get that the share of spending of country n on its own goods is

$$\pi_{nn} = \frac{T_n\left(w_n^{\beta}P_n^{1-\beta}\right)^{-\theta}}{\Phi_n} = \frac{T_n P_n^{-\theta}\left(\frac{w_n}{P_n}\right)^{-\beta\theta}}{\Phi_n}, \tag{50}$$

Furthermore, as in the basic Eaton-Kortum model, the ideal price holds $P_n = \frac{1}{\gamma}\Phi_n^{-\frac{1}{\theta}}$. Replacing this into Equation (50) yields

$$\frac{w_n}{P_n} = \gamma^{\frac{1}{\beta}}\left(\frac{T_n}{\pi_{nn}}\right)^{\frac{1}{\beta\theta}}, \tag{51}$$

and gains from trade are given by

$$\frac{\left(w_n/P_n\right)_{\text{Trade}}}{\left(w_n/P_n\right)_{\text{Autarky}}} = \left(\pi_{nn}\right)^{-\frac{1}{\beta\theta}}. \tag{52}$$

This shows that gains from trade are decreasing in β, and therefore increasing in the intermediate input share $1-\beta$. A direct consequence of this is that gains from trade are larger in a model with intermediate inputs than in a model without intermediate inputs.

The reason for this is that intermediate inputs introduce multiplier effects. Indeed, assume that opening up to trade allows Italy to import cheaper intermediate inputs. This generates a direct gain, as it allows Italy to lower the price of final goods that consumers consume. However, the cheaper final goods can also used as intermediate inputs, and so there is a further reduction in intermediate goods prices, leading to a further decrease in final goods prices, and so on (for a discussion of the macroeconomic role of this intermediate input multiplier, see Jones, 2011).

Finally, a more realistic production structure (as in the full Caliendo-Parro model), opens up further channels. For instance, it allows us to account for the fact that even non-traded sectors use traded intermediate inputs (e.g., hairdressers use imported hair dryers), and therefore gain from trade. We would have missed these gains in a model that only focuses on final goods.

6.3.3 Tariffs

The basic Eaton-Kortum model presented in Section 4 summarized all trade barriers between two countries with the iceberg cost d_{nj}. These iceberg costs can be thought of as representing tariffs as well as other obstacles to trade. However, modelling things this way abstracts from one of the most important features of tariffs: they create revenue for the government that collects them. When we want to have a full picture of the effects of tariff policy, we need to take these revenues into account. To do so, let us assume - as we did for gravity estimation - that trade costs are given by $d_{nj} = (1 + t_{nj})\,\widetilde{d}_{nj}$, where t_{nj} is the tariff on exports from j to n and $\widetilde{d}_{nj}$ captures all other trade frictions.[35] We also assume that the government transfers all tariff revenues to the representative consumer. These modifications do not affect the main equilibrium conditions of the model. The only equation that changes is the expression for national income of country n, which becomes

$$Y_n = w_n L_n + \sum_{j=1}^{N} t_{nj} \frac{Y_{nj}}{1 + t_{nj}}. \tag{53}$$

Thus, national income is now the sum of labour income and tariff revenue collected by the government. Tariff revenues for imports from country j are a fraction t_{nj} of the pre-tariff value of imports, equal to $\frac{Y_{nj}}{1+t_{nj}}$, where Y_{nj} is total spending of consumers from country n on goods from country j.

6.4 Applications and insights

6.4.1 Changes in trade policy

Caliendo and Parro (2015) use their model to assess the consequences of the 1994 North American Free Trade Agreement (NAFTA) between the United States, Canada and Mexico. Their main results are shown in Table 2. They indicate that NAFTA triggered very small gains for the United States, larger gains for Mexico, and a slight loss for Canada. Furthermore, they show that intermediate inputs are important: without them, the welfare effects from trade would be only half as large.

Table 2: Caliendo and Parro's estimates for the impact of NAFTA

	With Intermediates		Without Intermediates
	Welfare	Real Wage	Welfare
Mexico	+ 1.31%	+ 1.72%	+ 0.66%
Canada	- 0.06%	+ 0.32%	- 0.04%
United States	+ 0.08%	+ 0.11%	+ 0.04%

[35] These tariffs are "ad valorem" tariffs, as they apply to the value of imports: a tariff of 20% means that an importing firm has to pay a fraction 20% of the value of its imports to the government. There also exist "specific" tariffs which apply to quantities (e.g., 2€ for every imported banana). Both of these tariffs have exactly the same effects.

Considering these results, you may wonder why they are so small, and even negative for Canada. One explanation for the small size of the effect is that tariffs before NAFTA were not extremely high, especially not between the United States and Canada. So, Canada had little to gain: it was already well integrated with the US, and Mexico was a small and relatively distant economy (which, following the gravity equation, suggests that trade between the two countries should be limited). The fact that welfare changes are even negative is due to the loss of tariff revenues, and indicates that despite its overall free-trade message, the optimal level of tariffs in a Ricardian model is not zero! We will come back to this point shortly. Caliendo and Parro also speculate that if they had taken into account also changes in non-tariff trade costs, welfare effects would have been larger.

Table 3: The welfare consequences of higher tariffs

Table 4.2 Welfare Effect of Tariffs Under Perfect Competition

	Unilateral U.S. 40% Tariff				Uniform Worldwide 40% Tariff		
	One Sector	Multiple Sectors			One Sector	Multiple Sectors	
		Without Intermediates	Without Intermediates, With Dispersion	With Intermediates		Without Intermediates	With Intermediates
Country	1	2	3	4	5	6	7
AUS	−0.10%	−0.13%	−0.11%	−0.28%	−1.26%	−1.38%	−2.85%
AUT	−0.09%	−0.06%	−0.05%	−0.13%	−2.97%	−2.04%	−4.31%
BEL	−0.16%	−0.12%	−0.09%	−0.26%	−3.96%	−2.63%	−6.34%
BRA	−0.10%	−0.08%	−0.07%	−0.16%	−0.81%	−0.43%	−0.86%
CAN	−1.20%	−1.16%	−0.96%	−2.28%	−2.06%	−2.14%	−4.16%
CHN	−0.22%	−0.14%	−0.12%	−0.46%	−1.56%	−0.43%	−2.28%
CZE	−0.05%	−0.03%	−0.02%	−0.08%	−3.16%	−1.34%	−4.55%
DEU	−0.16%	−0.10%	−0.08%	−0.20%	−2.48%	−0.74%	−1.83%
DNK	−0.20%	−0.09%	−0.07%	−0.26%	−3.04%	−1.32%	−3.63%
ESP	−0.06%	−0.04%	−0.03%	−0.07%	−1.46%	−0.71%	−1.88%
FIN	−0.09%	−0.04%	−0.03%	−0.10%	−2.36%	−0.94%	−2.82%
FRA	−0.09%	−0.06%	−0.05%	−0.13%	−1.51%	−0.60%	−1.43%
GBR	−0.16%	−0.15%	−0.12%	−0.31%	−1.66%	−1.50%	−3.17%
GRC	−0.08%	−0.02%	−0.01%	−0.06%	−1.84%	−1.65%	−3.03%
HUN	−0.13%	−0.06%	−0.05%	−0.17%	−4.19%	−2.54%	−7.03%
IDN	−0.09%	−0.06%	−0.05%	−0.14%	−1.56%	−0.82%	−2.33%
IND	−0.16%	−0.13%	−0.11%	−0.25%	−1.17%	−0.71%	−1.65%
IRL	−0.91%	−0.56%	−0.47%	−1.58%	−4.41%	−2.17%	−6.65%
ITA	−0.07%	−0.03%	−0.03%	−0.07%	−1.47%	−0.46%	−1.31%
JPN	−0.10%	−0.05%	−0.05%	−0.11%	−0.92%	0.24%	0.04%
KOR	−0.21%	−0.14%	−0.12%	−0.34%	−2.31%	0.22%	−1.06%
MEX	−1.08%	−0.87%	−0.73%	−1.67%	−1.74%	−1.11%	−2.35%
NLD	−0.22%	−0.16%	−0.13%	−0.34%	−3.33%	−1.70%	−4.04%
POL	−0.04%	−0.03%	−0.03%	−0.08%	−2.21%	−1.28%	−3.18%
PRT	−0.06%	−0.05%	−0.04%	−0.09%	−2.12%	−1.85%	−3.67%
ROM	−0.03%	0.00%	0.01%	0.01%	−2.08%	−2.15%	−4.56%
RUS	−0.03%	−0.05%	−0.04%	−0.12%	−1.30%	−2.84%	−4.94%
SVK	−0.05%	−0.01%	−0.01%	−0.04%	−3.97%	−2.51%	−6.22%
SVN	−0.05%	−0.04%	−0.03%	−0.10%	−3.50%	−2.44%	−5.68%
SWE	−0.15%	−0.08%	−0.07%	−0.19%	−2.71%	−1.23%	−3.18%
TUR	−0.03%	−0.01%	−0.01%	−0.02%	−1.34%	−0.45%	−1.24%
TWN	−0.45%	−0.34%	−0.29%	−0.76%	−3.40%	−1.85%	−5.13%
USA	0.21%	0.41%	0.26%	0.63%	−0.80%	−0.44%	−1.00%
RoW	−0.49%	−0.43%	−0.37%	−0.97%	−2.69%	−2.45%	−5.74%
Average	−0.20%	−0.14%	−0.12%	−0.33%	−2.28%	−1.36%	−3.35%

Note: counterfactual results are computed using the exact hat algebra. All data is from WIOD and trade elasticities are from Caliendo and Parro (2010). See the online Appendix for details.

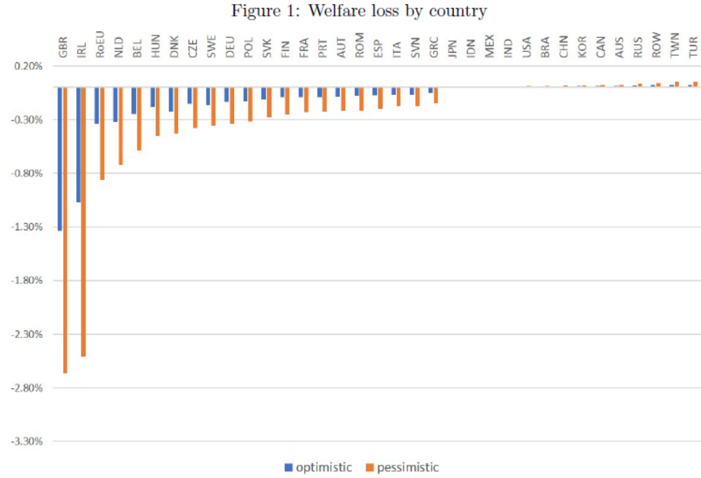
Source: Costinot and Rodriguez-Clare (2014).

Costinot and Rodriguez-Clare (2014) construct a model that is very similar to the one of Caliendo and Parro, and use it to study the consequences of a unilateral 40% increase in US tariffs. Their results are shown in Table 3. This illustrates that all countries of the world lose from this change, except the United States. This illustrates the double effect of tariffs. On the one hand, they are an obstacle to trade, and therefore keep countries from reaping the gains from trade. On the other hand, for the country which imposes the tariffs, they are an improvement in the terms of trade. Tariffs increase the post-tariff price of imports, which reduces demand and therefore decreases the pre-tariff price of imports. This is good news for the importing country, as it effectively pays only the pre-tariff price (the difference between the pre and the post-tariff price being the tariff itself, which is collected by the government and can be rebated to consumers).

So, Ricardian theory does not imply that free trade is the optimal economic policy. Instead, there is an optimal tariff which maximizes a country's welfare (for further details, see Costinot, Donaldson, Vogel and Werning, 2015). However, if all countries of the world were to retaliate and also increase their tariffs by 40%, there are no changes in the terms of trade, and everyone would lose, as illustrated in columns 5 to 7.

Figure 7: The impact of Brexit on different countries

Figure 1: Welfare loss by country



Notes: The figure plots the welfare loss by country for the optimistic and pessimistic scenario. Assumptions are the same as the notes to Table 3. We assume that the other EU countries have to fill the budget hole left by the UK proportionally to their GDP. This brings them a net fiscal loss of 0.015% in the optimistic case and 0.051% in the pessimistic case. The list of countries can be found in Table A.1.

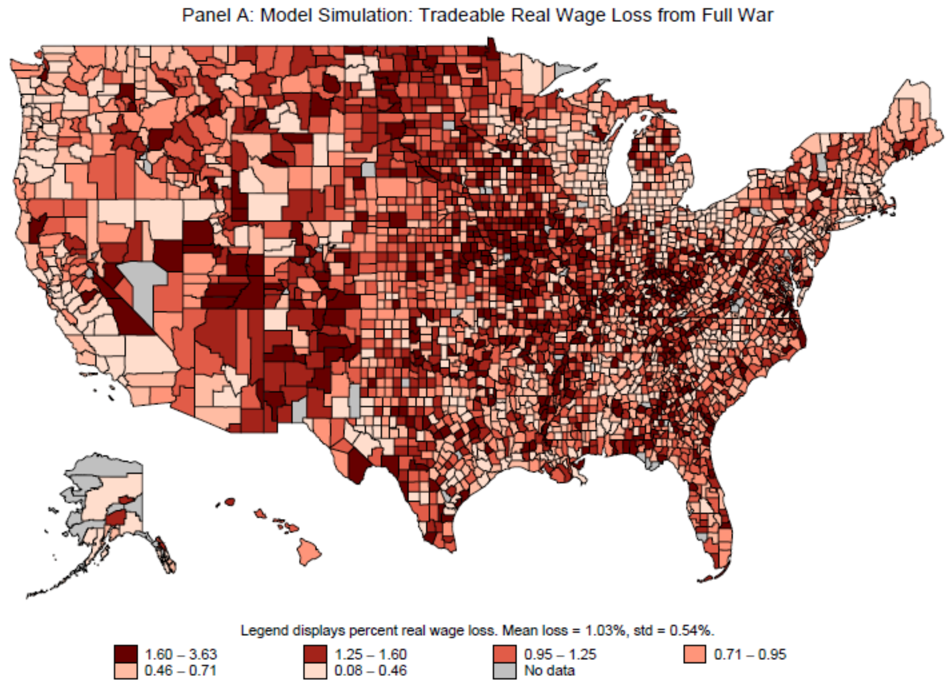
Source: Dhingra *et al.* (2017)

Dhingra *et al.* (2017) use a variation of this model to assess the implications of Brexit. They assess different scenarios for trade costs, which correspond to different scenarios for the outcome of Brexit negotiations: a strong increase in trade costs for a "hard Brexit" (trade under WTO rules) or a moderate increase for a "soft Brexit" (retained membership of the Single Market). The model predicts a 1.3% welfare loss for Great Britain in case of a soft Brexit, and a 2.7% welfare loss in case of a hard Brexit (see Figure 7). Other European countries lose as well, but their losses are relatively small, except for Ireland. This underlines again the importance of geography for trade, as underlined by the gravity equation: because Ireland is particularly close to the United Kingdom, it has the most to lose from higher trade barriers with it.

Finally, in a recent paper, Fajgelbaum, Goldberg, Kennedy and Khandelwal (2019) study the effects of the 2018 trade war on US welfare. While they do not use an Eaton-Kortum style model, the methods that they use are still very similar to the ones we have studied so far. The authors summarize the 2018 events as follows. "*After more than a half-century of leading efforts to lower international trade barriers, in 2018 the U.S. enacted several waves of tariff increases on specific products and countries. Import tariffs increased from 2.6% to 16.6% on 12,043 products covering $303 billion (12.7%) of annual U.S. imports. In response, trade partners imposed retaliatory tariffs on U.S. exports. These counter-measures increased tariffs from 7.3% to 20.4% on 8,073 export products covering $127 billion (8.2%) of annual U.S. exports.*" They then proceed to analyze the effects of these tariffs. Using the Romalis (2007) methodology presented above, they find that tariffs greatly lowered trade volumes for affected products: imports of varieties targeted by import tariffs fell by more than 30%, and exports of products targeted by retaliatory tariffs fell by 10%. However, as the United States is still a relatively closed economy, and tariffs applied only to a small fraction of imports and exports, the overall losses are modest, and US GDP was only lowered by 0.27%. These losses were however unequally distributed over space, with Midwest farming regions being hit hardest (see Figure 8). Indeed, for political reasons, retaliatory tariffs by China and other countries mainly targeted these (largely Republican) regions.

To conclude, the results from these quantitative studies suggest that realistic increases in trade barriers have moderate effects on welfare and real wages: they are neither enormous nor negligible. As we will see in Chapter VII, this has important political economy consequences, because the losses from globalization tend to be concentrated on certain regions or people, while its gains are more likely to be diffuse. Another important insight is that international linkages through supply chains matter, as the welfare impact of trade is much larger in models which take them into account. Current research efforts strive to make models even more realistic on this side. For instance, de Gortari (2018) argues that the input-output structure is even more specialized than standard models would suggest. While Caliendo and Parro assume, for instance, that Mexican cars exported to Germany or to the US are built with the same intermediate inputs, De Gortari shows that for 74% of inputs into Mexican cars

Figure 8: Regional consequences of the 2018 US trade war



Source: Fajgelbaum, Goldberg, Kennedy and Khandelwal (2019).

exported to the US come from the US, while for Germany, the corresponding number is 18%. This paper, as well as Antràs and de Gortari (2020), make some progress in building a more sophisticated model of international supply chains.

6.4.2 The Covid-19 pandemic

The ongoing Covid-19 pandemic has disrupted production around the globe. International trade was hit particularly hard: in April 2020, the WTO was forecasting a decline in global trade between 13 and 32%, substantially larger than the decline in global GDP.

There is a large and growing amount of research on the trade consequences of the pandemic, using the methodologies we have studied in this chapter.[36] For instance, Sforza and Steininger (2020) use the Caliendo-Parro model to assess the interaction between globalization and the pandemic. They find that the impact of the pandemic would have been lower in a less open world economy, but not substantially so (as even their less open economy still has a substantial amount of trade).

The pandemic has also raised a number of new aspects that are not considered in standard trade

[36] For an overview by a leading expert, see https://www.youtube.com/watch?v=erq8pqBpFhI

Table 4: Trade restrictions for essential medical goods

Table 4: Trade of essential medical goods during COVID-19

	Number of countries	Number of countries		Share of countries (%)	
		Import liberalization	Export curbs	Import liberalization	Export curbs
Surplus	22	4	19	18.18	86.36
Deficit	87	25	40	28.74	45.98

Note: Surplus (deficit) refers to countries with positive (negative) net exports in essential medical goods in 2018.

Source: Leibovici and Santacreu (2020).

models. One important issue is analyzed by Leibovici and Santacreu (2020): some traded goods are "essential" (such as masks and medicines, but also food), and may be very hard to substitute in case of a sudden disruption. This has become painfully clear during the pandemic, as exporting countries of essential goods restricted their exports (see Table 4). Using a model with occasional pandemic shocks, they then show that net importers of essential goods suffer more during pandemics. In general, the Covid-19 pandemic has shown the necessity to take into account dynamic features: importing all essential goods such as food and drugs from other countries may be efficient and welfare-enhancing in normal times, but can be very costly during a pandemic.

6.4.3 Other shocks

In Problem Set 2, we saw that in the Dornbusch-Fischer-Samuelson model, an increase in productivity for a country's comparative disadvantage good increases the welfare in that country, but decreases the welfare of its trading partner. In a famous article, Samuelson (2004) argued that this result should be worrisome: if China's technological catch-up was concentrated in its comparative disadvantage sectors, it could lead to lower real wages in the United States. However, di Giovanni, Levchenko and Zhang (2014) have shown that this intuition is not a world with many countries. Analysing a variation of the Eaton-Kortum model, they show that the United States (and Europe) would actually gain if China's productivity growth were biased towards its comparative disadvantage sectors.

This surprising result is due to the fact that China's current productivity profile is actually similar to the one of most (developing) countries in the world: it has high productivity in some basic manufacturing industries (such as textiles and apparel), and low productivity in more high-tech manufacturing sectors (such as machinery). By catching up in these comparative disadvantage industries, China becomes more similar to Europe or to the United States, but less similar with respect to the rest of the world. Therefore, the overall gains from trade increase, because the overall differences in opportunity costs across countries increase.

Table 5 summarizes the main results of the paper. First, Panel A shows the welfare gains from China participating in world trade. These gains are largest in China itself, the neighbouring countries of South, East and Central Asia (which have low costs for trading with China), and developed OECD countries which are quite dissimilar to China. On the other hand, they are small in countries which are far away from China, and quite similar in terms of productivity (they are actually negative for several countries, such as for example Honduras, El Salvador and some North African countries). These countries do not trade that much with China, and are negatively affected by the fact that China pushes them out of the US or EU market.

Table 5: The consequences of Chinese productivity growth

TABLE 3—WELFARE CHANGES

	Mean	Median	Min	Max	Countries
Panel A. Welfare gains from trade with China					
China	3.72				
OECD	0.13	0.12	−0.03	0.30	22
East and South Asia	0.23	0.20	−0.20	0.80	12
East. Europe and Cent. Asia	0.14	0.09	−0.08	0.78	11
Latin America and Caribbean	0.09	0.09	−0.27	0.39	15
Middle East and North Africa	0.12	0.13	0.04	0.22	6
Sub-Saharan Africa	0.08	0.06	−0.04	0.21	8
Panel B. Welfare gains from balanced growth in China					
China	11.43				
OECD	0.01	0.02	−0.01	0.04	22
East and South Asia	0.03	0.04	−0.05	0.09	12
East. Europe and Cent. Asia	0.01	0.01	−0.02	0.06	11
Latin America and Caribbean	−0.01	0.00	−0.06	0.04	15
Middle East and North Africa	−0.01	−0.01	−0.07	0.02	6
Sub-Saharan Africa	0.00	0.01	−0.02	0.02	8
Panel C. Welfare gains from unbalanced growth in China					
China	10.57				
OECD	0.17	0.12	−0.07	0.77	22
East and South Asia	0.84	0.74	0.22	1.70	12
East. Europe and Cent. Asia	0.42	0.34	0.07	1.52	11
Latin America and Caribbean	0.50	0.49	0.09	1.68	15
Middle East and North Africa	0.48	0.52	0.19	0.77	6
Sub-Saharan Africa	0.23	0.21	−0.03	0.57	8

Notes: Units are in percentage points. This table reports the changes in welfare from three counterfactual scenarios. Panel A presents the welfare gains in the benchmark for the 2000s, relative to the scenario in which China is in autarky. Panel B presents the changes in welfare under the counterfactual scenario that growth is balanced in China across sectors, relative to the benchmark. Panel C presents the changes in welfare under the counterfactual scenario of unbalanced growth in China, relative to the benchmark. The technological changes assumed under the counterfactual scenarios are described in detail in the text.

Source: di Giovanni, Levchenko and Zhang (2014).

Second, Panels B and C show that the gains from productivity growth biased towards China's comparative disadvantage sectors are much larger than those from uniform productivity growth. Again, the gains are largest in China's closest neighbours, but they are also large in virtually all other regions of the world.

Finally, perhaps the most original application of modern Ricardian trade theory deals with data which is more than 3500 years old. Around the 19th BCE, the Assyrian empire covered most of modern-day

Turkey. Several of the cities that existed at the time are now "lost": historians know that they existed, but they do not know where they were (in general, there are several archaeological sites which could correspond to a city, but it is difficult to say which is the right one). Barjamovic, Chaney, Coşar and Hortaçsu (2019) use information on trade between these cities (coming from ancient texts) to locate these lost cities. To do so, they built an Eaton-Kortum style model, and then use the resulting system of equations to estimate the distance between cities which is needed to make the model fit the trade data. Combining the distance estimates with the location of some known cities, they can formulate an estimate of where the lost ones should be. The methodology seems to work well, as their results are generally very close to the conjectures of historians. Therefore, the methodology could be used to find the cities for which historians have not made conjectures, or disagree among each other.

As mentioned above, the fact that the Eaton-Kortum model and its extensions are consistent with gravity gives researchers some confidence that "trade theory with numbers" is empirically meaningful. However, this is not a proper empirical test of the model, in the sense that it does not show us that trade is really explained by relative productivity differences. The next section briefly surveys the attempts that have been made to provide some more stringent empirical tests. It then turns to investigating the deep determinants of productivity differences across countries.

7 Further empirical evidence and extensions

7.1 Empirical evidence beyond gravity

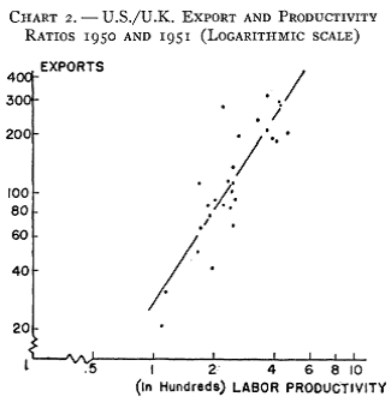
In the 1950s and 1960s, several studies tried to test whether the logic of Ricardo's basic two goods-two countries example held in the real world. MacDougall (1951) and Balassa (1963) argued that Ricardo's theory implied that goods with a higher ratio of the productivity of country A to the productivity of country B should have a higher ratio of Exports from A to B to Exports from B to A. Figure 9 shows that this was indeed the case for the United States and the United Kingdom in the beginning of the 1950s, which was seen as vindication for the Ricardian theory.[37]

However, this empirical test is actually inconsistent with Ricardo's theory. Indeed, in Ricardo's example (as in the Dornbusch-Fischer-Samuelson and Eaton-Kortum models) cross-exports of the same good just do not happen: either country A exports the good to country B, or country B exports to country A, but never both. Thus, while MacDougall and Balassa's tests indicated that relative productivities

[37]In these studies, productivity was simply measured as labour productivity (output per worker).

seemed to be important for trade, they were not rigorous tests of Ricardian theory. Before the advent of the Eaton-Kortum model, empirical tests were blocked by these difficulties: the two-goods, two-country model's predictions were too strong to be successfully tested, but there were no Ricardian models with many countries whose predictions could have been tested.

Figure 9: A simple empirical assessment of Ricardo's ideas



Source: Balassa (1963).

With the Eaton-Kortum model, this changed, as researchers now had clear predictions for trade flows in a many-country world. As mentioned before, these were first of all used for to check that the model was consistent with gravity equations. Costinot, Donaldson and Komunjer (2012) then went a step further, proposing to "*offer the first theoretically consistent Ricardian test*". They consider an extension of the Eaton-Kortum model which is quite similar to the Caliendo-Parro model we have studied above, with I different industries that each have a Fréchet distribution of productivities with average productivity T_j^i and common productivity dispersion θ. As we had seen above, such a model yields predictions for trade shares at the industry-level, and therefore also an industry-level gravity equation. In particular, the authors show that the sales from country j to country n in any sector i are given by

$$\ln Y_{nj}^i - \ln Y_{jj}^i = \delta_{nj} + \delta_n^i + \theta \ln T_j^i + \varepsilon_{jn}^i. \tag{54}$$

Thus, the exports of country j to country n in sector i depend on factors which are specific to the country pair (δ_{nj}, which captures for instance the trade cost parameter d_{nj}), factors specific to the importing country n and the sector i (δ_n^i, which captures among other things country n's average productivity in sector k, T_n^i), and finally on country j's productivity in sector i, $\ln T_j^i$. If the Ricardian principle of comparative advantage is true, then the coefficient θ should be positive: an increase in a country's productivity for a given sector (relative to its trade partner's productivity, captured by δ_n^i, and relative to its overall productivity, captured by δ_{nj}) should increase its exports in that sector. This intuition is not very different from MacDougall's or Balassa's. However, by deriving the estimating equation (54) from a complete many-country model, we now have a more rigorous test: in particular, we now control for a country's average productivity and the one of its trade partner. Costinot et al. estimate that $\theta \approx 6.5$, and significantly different from 0,[38] which they interpret as strongly supportive of Ricardo's ideas: relative productivity differences do matter for real-world trade flows.

The Eaton-Kortum model has also stimulated researchers to develop other multi-country Ricardian models. The most important such model is due to Costinot (2009). Costinot's model is a bit more complicated and less widely used than Eaton-Kortum, but it has the advantage that the productivity distribution is not limited to a Fréchet distribution. Therefore, the model can in principle be used for a different empirical test of Ricardo's ideas: if one has data on the true productivity distributions and transport costs for every country in the world, one can just plug those into the model's equations, and see whether the predicted trade flows of the model correspond to those in the real world.

While this approach is intuitive, it presents a huge practical problem: as some countries do not produce certain goods, one cannot observe their productivity for these goods.[39] For instance, in Ricardo's original example, if there is complete specialization, England produces only cloth and Portugal produces only wine. But then, there is no way that we can compute England's productivity for wine production or Portugal's productivity for cloth production, and therefore, we cannot test whether this specialization occurred because England has a higher relative productivity for cloth than for wine.[40]

However, Costinot and Donaldson (2012) found a clever way around this problem for the special case of agricultural goods. Indeed, knowing the climate, wind, rainfall etc, of a given region, it is possible to calculate how good it would be at producing a given crop, even if it is not actually grown there. Plugging this information on productivity into the Costinot model, they find that the correlation with real-world agricultural output per country and the predictions of the model is quite high. So, "*Ricardo's theory of comparative advantage is not just mathematically correct and non-trivial; it also*

[38]There are endogeneity issues: maybe productivity is not exogenous, but affected by trade itself (for instance, if trade creates technology transfers). Therefore, the authors estimate (54) also with an Instrumental Variable (IV) approach, instrumenting sectoral productivity by R&D expenditure. This makes their results even stronger. However, maybe trade affects firms' R&D incentives, which would question the exogeneity of Costinot et al.'s instrument.

[39]One should add to this the complicated problem of defining and measuring productivity even when there is data, but that is a secondary issue.

[40]The paper by Costinot, Donaldson and Komunjer (2012) gets around this issue by assuming that there is some productivity dispersion within every industry: so, even if Italy imports more T-shirts from China than it exports T-shirts to China, there is always a little bit of such intra-industry trade.

[for agricultural goods] has significant explanatory power in the data".

7.2 Where do productivity differences come from?

For many agricultural goods, the answer to this question is obvious. It takes no sophisticated (economic) theory to explain why Colombia is better at growing coffee than Iceland: this is due to climatic conditions which are largely beyond human control. However, agricultural goods only represent a small fraction of world trade nowadays, and for manufacturing goods, it is much more difficult to find natural sources of productivity differences. Indeed, as it is possible to move machines and ideas across countries[41], why would not every country in the world be able to produce shoes, cars, planes etc. with the same productivity?

In the last years, a line of "institutional trade theories", reviewed in Nunn and Trefler (2014), have tried to address this question. Their claim is that institutions - countries' legal and judiciary systems, and even culture - matter to determine whether a country is performant in producing in certain industries. Nunn and Trefler pay particular attention to differences in contract enforcement across countries, and give the following example: *"Consider a complex product such as a commercial airliner. Its production requires high levels of innovative effort by all parties involved and this effort is so difficult to verify in a legal setting that only the most incomplete of contracts can be written between these parties. In contrast, more standardized products such as blue jeans do not require any relationship-specific, non-contractable inputs. Thus, a country with good contracting institutions will have relatively low costs of producing airliners and relatively high costs of producing blue jeans. That is, contracting institutions will be a source of comparative advantage*." Intuitively, if inputs are not very specific (for instance, jeans cloth), the firm can just switch to another supplier if its current one is not doing a good job, and having an extensive contract with suppliers is not very important. However, if an certain input into production is very specific (for example, a steering wheel that only fits an Airbus A320), then it is important that the airplane company can be sure that its supplier does a good job, and has a comprehensive contract with the supplier specifying all steps to be taken, and compensation in case something goes wrong.

For international trade, this implies that countries with a good contracting environment should have a comparative advantage in contract-intensive industries. This was shown theoretically by Levchenko (2007), and verified empirically by Nunn (2007). Nunn proposed to measure the "contract intensity" of

[41] Comin and Mestieri (2013) provide evidence for technology diffusing relatively quickly across countries.

an industry by the thickness of the market for its inputs: an industry with many buyers and sellers on input markets is not very contract-intensive, while an industry with few buyers and sellers is. With this, Nunn estimated an equation similar to the one of Costinot, Donaldson and Komunjer (2012) seen above, but using, instead of productivity of country j in sector k, $\text{Contract intensity}_k * \text{Contract enforcement}_j$, where the degree of contract enforcement in a country is captured by a World Bank measure. He finds a positive and significant coefficient on this term, suggesting that contracting institutions matter for comparative advantage. In related work, Tabellini (2008) argues that trust may act as a substitute for contract enforcement.

Contracting institutions are however not the only potentially relevant ones. For instance, Rajan and Zingales (1998) have famously shown that some economic sectors are more dependent on external finance (credit, stocks, bonds) than others. This can also shape trade patterns: Manova (2008) has shown that when a country engages in financial liberalization, its exports increase most in the most finance-dependent sectors, which is consistent with the country getting a comparative advantage in these industries.

In sum, the literature on the institutional origins of comparative advantage is still quite recent, but it provides an interesting perspective on the potential sources of the abstract productivities z that we have encountered in this chapter.

Part III

Heckscher-Ohlin Trade Theory

1 Introduction

Heckscher-Ohlin trade theory is based on the same fundamental principle than Ricardian trade theory: countries gain from trade if they have different opportunity costs for the same goods, and specialize according to comparative advantage. However, differences in costs do not come from differences in productivity any more, but from differences in factor endowments.

The first verbal formulation of the theory was developed by the Swedish economist Eli Heckscher (Heckscher, 1919). His student Bertil Ohlin put these ideas into a mathematical model (Ohlin, 1924). The modern formulation of the Heckscher-Ohlin theory, however, is largely due to Paul Samuelson (and some people therefore refer to it as the Heckscher-Ohlin-Samuelson theory).

Heckscher and Ohlin formulated their theory under the impression of the 19th century's "Great Specialization". European countries exported manufacturing goods, while the global periphery (including the colonies, the United States or Argentina) exported agricultural goods and raw materials. Heckscher and Ohlin did not believe that this pattern of trade was only due to productivity differences. Instead, they noted that these two sets of countries were very different in terms of their production factor endowments: while the periphery had seemingly endless agricultural land at its disposal, European countries were much more densely populated. Their theory fit these two facts together, by claiming that Europe exported labour-intensive manufacturing goods precisely because it was relatively well endowed with labour (and had therefore low production costs in manufacturing), and periphery countries exported land-intensive agricultural goods precisely because they were relatively well endowed with land (and had therefore low production costs in agriculture).

The new theory was also well placed to explain the effects of trade on the income distribution within countries. Actually, this was the main motive for Heckscher to develop it (his 1919 book was entitled "*The Effect of Foreign Trade on the Distribution of Income*"). Up to the current day, many economists continue to assess the consequences of trade and globalization on different categories of the population by using the Heckscher-Ohlin framework.

The exposition of the Heckscher-Ohlin model in these notes partly follows Ventura (2005). Section

2 introduces the concept of an "integrated economy", popularized by Paul Samuelson. Sections 3 and 4 then analyse international trade both when it can replicate this integrated economy, and when it cannot. In particular, they derive the most famous results of the model: the Heckscher-Ohlin theorem, the Stolper-Samuelson theorem and the Rybczynski theorem. Finally, Section 5 covers the long history of empirical tests of the Heckscher-Ohlin model.

2 The integrated economy

In 1949, Paul Samuelson, revisiting Heckscher-Ohlin theory, made a surprising claim: under certain conditions, international trade in goods can be a substitute for international factor mobility (Samuelson, 1949). That is, it may be that a world in which production factors cannot move across countries, but goods can, is just as well off as a world in which both goods and production factors can move.

To understand this claim, and the conditions under which it holds, we introduce in this section the concept of an integrated economy. In the integrated economy, there are different countries, but borders are de facto irrelevant, because production factors and goods are perfectly mobile. We will determine the equilibrium of this economy, and then ask ourselves Samuelson's question: would it be possible to reach this equilibrium even if production factors were not mobile?

2.1 A world without borders

2.1.1 Assumptions and optimal behaviour

Consider a world with I goods and two production factors, capital and labour. The world is endowed with K units of capital and L units of labour, which are supplied inelastically. Note that capital and labour are not necessarily equally distributed across the population. For instance, the distribution of capital is highly unequal in all countries of the world (see Piketty, 2014): while most people's income is mainly made up by wages, some people (including the richest segments of the population) get a lot of income from capital. Thus, changes in the prices of labour and capital will change the income distribution between these two categories.

We will, however, continue to make the assumption that all people have the same CES preferences, so that, as argued in Chapter I, the distribution of income does not matter for any good's demand function. Instead, demand only depends on aggregate income, which is now given by $\mathcal{W} = wL + rK$, where r (the rental rate of capital) is the price of one unit of capital. Furthermore, we assume that preferences are Cobb-Douglas with equal spending shares, so that the demand for good i is

$$p_i c_i = \frac{1}{I}\left(wL + rK\right). \tag{55}$$

Firms operate under perfect competition. They can produce good i with a Cobb-Douglas production function

$$y_i = z_i k_i^{\alpha_i} l_i^{1-\alpha_i}. \tag{56}$$

Perfect competition implies, as always, that firms charge a price equal to their marginal cost. As we have seen in Chapter I, the marginal cost of production for this production function is given by

$$\lambda = \frac{1}{z_i}\left(\frac{r}{\alpha_i}\right)^{\alpha_i}\left(\frac{w}{1-\alpha_i}\right)^{1-\alpha_i}, \tag{57}$$

and input demand functions are given by

$$r = \lambda \alpha_i \frac{y_i}{k_i} \text{ and } w = \lambda\left(1-\alpha_i\right)\frac{y_i}{l_i}. \tag{58}$$

This gives the optimal ratio of capital to labour chosen by the firm:

$$\frac{k_i}{l_i} = \frac{\alpha_i}{1-\alpha_i}\frac{w}{r}. \tag{59}$$

Equation (59) shows that industries with a high value of α_i use a lot of capital and little labour (that is, they are capital-intensive). On the other hand, industries with a low value of α_i are labour-intensive.

2.1.2 Equilibrium

As usual, the equilibrium in this world is defined as a set of prices (for goods, labour and capital) such that consumers and firms maximize their utility/profits taken these prices as given, and all markets clear. There are two sets of markets here:

1. Goods markets, where for every good i, we must have $c_i = y_i$.
2. Factor markets, where market clearing implies $\sum_{i=1}^{I} l_i = L$ and $\sum_{i=1}^{I} k_i = K$.

Solving for production in equilibrium is easy. Using the fact that prices equal marginal cost, we get

$$rk_i = \alpha_i p_i y_i \tag{60}$$

$$wl_i = (1 - \alpha_i)\, p_i y_i \tag{61}$$

Thus, note that each production factor earns a constant share of the firms' total revenues, irrespective of its price. This share is equal to the coefficients α_i and $1 - \alpha_i$, which are therefore called the factor shares. This result is very useful: replacing Equation (61) into Equation (55) gives

$$wl_i = (1 - \alpha_i) \frac{1}{I} (wL + rK) . \tag{62}$$

Summing up Equation (62) across goods, we get

$$wL = (1 - \alpha) (wL + rK) , \tag{63}$$

where α is defined as $\alpha \equiv \frac{1}{I} \sum_{i=1}^{I} \alpha_i$. Replacing this again into (62) gives the allocation of labour across industries:

$$l_i = \frac{1 - \alpha_i}{1 - \alpha} \frac{L}{I} . \tag{64}$$

Equation (64) shows that an industry's share of the labour force is increasing in its labour share $1 - \alpha_i$. The exact analogue holds for capital:

$$k_i = \frac{\alpha_i}{\alpha} \frac{K}{I} . \tag{65}$$

Knowing how much capital and labour are used in every industry, we can directly deduce how much the world produces and consumes of every good. Therefore, we can calculate world real income, which is equal to

$$\frac{\mathcal{W}}{P} = U = Z K^{\alpha} L^{1-\alpha} , \tag{66}$$

where $Z = \prod_{i=1}^{I} \left(\frac{z_i}{I} \left(\frac{\alpha_i}{\alpha} \right)^{\alpha_i} \left(\frac{1-\alpha_i}{1-\alpha} \right)^{1-\alpha_i} \right)^{\frac{1}{I}}$. Thus, aggregate real income is again a Cobb-Douglas function of aggregate factor endowments. What are equilibrium wages and rental rates? From the preceding, we can deduce

$$\frac{w}{r} = \frac{1 - \alpha}{\alpha} \frac{K}{L} . \tag{67}$$

Combining this with Equation (66) gives

$$\frac{w}{P} = (1 - \alpha) Z \left(\frac{K}{L} \right)^{\alpha} \text{ and } \frac{r}{P} = \alpha Z \left(\frac{K}{L} \right)^{\alpha - 1} . \tag{68}$$

2.2 Samuelson's angel

In the previous section, we have determined the equilibrium of a world without countries. However, introducing countries (or nationalities) is irrelevant as long as production factors are perfectly mobile. Calling all workers "Italian" and all capitalists "Swiss", for instance, does not change anything for the equilibrium, as long as Italian workers and Swiss capitalists can freely work together. Thus, the integrated economy represents a world in which countries do not matter: "*all factors [are] perfectly mobile, and nationalism ha[s] not yet reared its ugly head*", in Samuelson's words.

Now, however, "*suppose that an angel came down from heaven*" and created borders across which production factors could not move. Would this destroy the happy world of the integrated economy? Not necessarily, Samuelson argued, as long as the angel followed a simple rule: if in every industry, she assigned a common fraction of the labour and capital used in production to every country, nothing would change. The world as a whole would produce as much as before, and wages and rental rates would be the same as in the integrated economy (and thus equalized across countries).

To understand Samuelson's argument, consider two countries (Italy and Switzerland) in a world with two goods (food and cars). Samuelson's allocation rule then says that the angel can give, for instance, 10% of all labour and capital used in food production to Switzerland, and the remaining 90% to Italy, while in car production, she gives 20% of factors to Switzerland and 80% to Italy. Why does splitting up factors like this does not reduce world production? There are two reasons for this:

1. **Constant returns to scale in production.** If Switzerland is given 10% of the capital and 10% of the labour used in the world food industry, it will produce 10% of the latter's output. Italy, with the remaining 90%, will produce 90% of output. Thus, splitting up production across two countries has not reduced production overall! This is because the production function for food has constant returns to scale: if it had increasing returns, then getting the same world production by splitting it up in two parts were impossible.

   Furthermore, note that while the food production function has constant returns to overall, there are decreasing returns to every single factor. This is the reason for which the angel cannot be allowed to allocate to Switzerland a different percentage of the capital and of the labour of the food industry: if it gave Switzerland, say, 10% of the labour and 20% of the capital of food production, and Italy the remaining 90 and 80%, production would be lowered.[42]

[42] This can be most clearly seen in the case the angel would give all labour to one country and all capital to the other, in which case overall production would be zero with the Cobb-Douglas production function. Note, however, that this does not imply countries must have the same overall capital-labour ratios. For instance, assume food uses lots of labour

2. **Equal technologies (and therefore productivities) in every country**. The allocation rule also makes the implicit assumption that the new countries of Italy and Switzerland can continue to produce food with the same production function, and in particular, with the same productivity z_{Food}. Suppose that this were not the case, and that Italy has a higher productivity for food ($z_{\text{Food,IT}} > z_{\text{Food,CH}}$). In a world without borders, that would imply that all production factors used for food come to Italy, to work with the highest productivity. However, once the angel has thrown some production factors into Switzerland, they can no longer combine with Italian productivity, and world food output must necessarily be lower.

   Therefore, notice that replicating the integrated economy is almost impossible in a Ricardian model. In Ricardo's famous example, trade makes England and Portugal better off, but they can never be as well off as when English workers were allowed to move to Portugal and to work with the superior Portuguese technologies.

These two restrictions are also the most important assumptions of Heckscher-Ohlin trade theory. So, the possibility that trade in goods can substitute for factor mobility is a particularity of Heckscher-Ohlin theory: in Ricardian models with productivity differences (Chapter II) or in models with increasing returns (Chapter IV) this is (almost) impossible.

The parable of the angel already hints at the role that trade in goods must play. For instance, assume that Italians and Swiss consumed as much food as cars in the integrated economy. Then, if the angel gives Italy almost all of the food industry and Switzerland most of the car industry, Italians and Swiss must exchange (Italian) food against Swiss (cars) to maintain their integrated economy consumption. In the next section, we will derive a formal mathematical formulation of Samuelson's angel's allocation rule, and by doing so, derive some first important results of the Heckscher-Ohlin theory.

3 Heckscher-Ohlin theory with factor price equalization

3.1 Formalizing Samuelson's rule

3.1.1 Assumptions

We continue to make the assumptions of Section 2.1.1. However, we now assume that the world is made up by J different countries, endowed each with K_j units of capital and L_j units of labour.[43]

and little capital, and cars use little labour but lots of capital. If the angel gives 90% of the food industry to Italy and 90% of the car industry to Switzerland, Italy will have a lower ratio of capital to labour than Switzerland.

[43] Of course, countries' endowments add up to the world's endowment, so that $\sum_{j=1}^{J} K_j = K$ and $\sum_{j=1}^{J} L_j = L$.

These production factors cannot move across countries.

We make the assumption that there are no cross-country productivity differences, and for simplicity set all productivities to 1 (that is, $z_{ij} = z_i = 1$ for all goods i and countries j). Furthermore, we assume that the factor shares (the α_i's) are the same across countries, but different across industries.. These differences in factor shares across industries, combined with differences in factor endowments across countries, will determine the patterns of trade in the Heckscher-Ohlin model.[44]

Finally, for simplicity, we assume that there are no trade costs.

An equilibrium in this model is a set of prices for all goods and production factors such that firms and consumers maximize and markets clear. There are two sets of markets:

1. **Global** goods markets, where for every good i, we must have $\sum_{j=1}^{J} c_{ij} = \sum_{j=1}^{J} y_{ij}$.

2. **Local** factor markets, for capital and labour, where in every country, market clearing implies $\sum_{i=1}^{I} l_{ij} = L_j$ and $\sum_{i=1}^{I} k_{ij} = K_j$.

Note that as factor markets are local, wages and rental rates can a priori be different across countries. In the next section, we prove that this equilibrium (which we call the trade equilibrium) coincides with the integrated economy one if and only if countries' endowments of capital and labour respect the allocation rule of Samuelson's angel.

3.1.2 Replicating the integrated economy

Factor price equalization (FPE) *The trade equilibrium coincides with the integrated economy equilibrium if and only if the trade equilibrium features factor price equalization (that is, if wages w_j and rental rates r_j are the same in all countries of the world).*

Proving this statement is easy. Consider a trade equilibrium price vector which holds FPE. Then, this vector is also an equilibrium price vector for the integrated economy: it holds the requirement that there is a unique price of labour and capital in the world, and it ensures market clearing (as the trade equilibrium market clearing condition imply the integrated economy ones).[45] When is FPE achieved?

[44] This is why the model is sometimes also referred to as the "factor proportions" model. Note that the fact that there are now two factors of production alone is not sufficient to generate new reasons for trade: if all industries had the same capital share α, there would be no reasons for trade here, as there would be no possibilities to specialize according to factor endowments.

[45] Note that in Ricardian models, we almost never have FPE: in the Dornbusch-Fischer-Samuelson or Eaton-Kortum models, real wages in general differ across countries. This is just another illustration of the fact that replicating the integrated economy is almost impossible in these models.

FPE and Samuelson's angel *FPE occurs if and only if in all countries, all industries use the same capital-labour ratio as the integrated economy.*

Proof. ($\Rightarrow$) If the trade equilibrium features FPE, the wage-rental ratio in every country is equal to its integrated economy value. However, as we have showed previously (see Equation (59)), the capital-labour ratio in any industry i only depends on the wage-rental ratio and on the factor shares α_i. Thus, as the wage-rental ratio is the same as in the integrated economy, and the factor shares are also the same, every industry in every country must produce with the same capital-labour ratio as the corresponding industry in the integrated economy.
($\Leftarrow$) If all industries use the same capital-labour ratio as the integrated economy, then all countries must have the same wage-rental ratio (which is the same as in the integrated economy). Moreover, as shown in Equation (57), the marginal cost of production of good i produced in country j is $\frac{w_j}{1-\alpha_i}\left(\frac{1-\alpha_i}{\alpha_i}\frac{r}{w}\right)^{\alpha_i}$ (where we have already used the fact that the ratio r/w is equalized across countries). This implies that every country must have the same wage in equilibrium: if this were not the case, only the country with the lowest wages would produce (and the labour market would not clear in higher-wage countries). $\square$

Next, we turn to a simple and useful graphical interpretation of Samuelson's rule.

3.1.3 A graphical illustration for two countries and two goods

Assume now that $I = J = 2$. Furthermore, lets give names to the goods: we have a L-good and a K-good, with $\alpha_L < \alpha_K$ (that is, the K-good is more capital-intense than the L-good).
Saying that the L-good in country 1 is produced with the same capital-labour ratio than in the integrated economy, as required by Samuelson's rule, implies that there exists a non-negative real number a_1 such that

$$\begin{pmatrix} l_{L1} \\ k_{L1} \end{pmatrix} = a_1 \begin{pmatrix} l_L \\ k_L \end{pmatrix}$$

An analogous condition must hold for the K-good: there must be same real number b_1 such that $\begin{pmatrix} l_{K1} \\ k_{K1} \end{pmatrix} = b_1 \begin{pmatrix} l_K \\ k_K \end{pmatrix}$. Factor market clearing implies

$$\begin{pmatrix} L_1 \\ K_1 \end{pmatrix} = a_1 \begin{pmatrix} l_L \\ k_L \end{pmatrix} + b_1 \begin{pmatrix} l_K \\ k_K \end{pmatrix} \tag{69}$$

The same thing is true for the other country: there must be real numbers a_2 and b_2 such that

$$\begin{pmatrix} L_2 \\ K_2 \end{pmatrix} = a_2 \begin{pmatrix} l_L \\ k_L \end{pmatrix} + b_2 \begin{pmatrix} l_K \\ k_K \end{pmatrix} \tag{70}$$

So, we can restate the conditions for FPE in this simple case as follows.

Lemma. *In the two country, two goods case, a necessary and sufficient condition for FPE is that there exist $a_1, b_1, a_2, b_2 \in \mathbb{R}^+$ such that Equations (69) and (70) hold.*

Note that it is easy to show that $a_1 + a_2 = b_1 + b_2 = 1$ (just sum Equations (69) and (70) and solve the linear system), which shows that when we solved the system for one country, we directly have the solution for the other country.

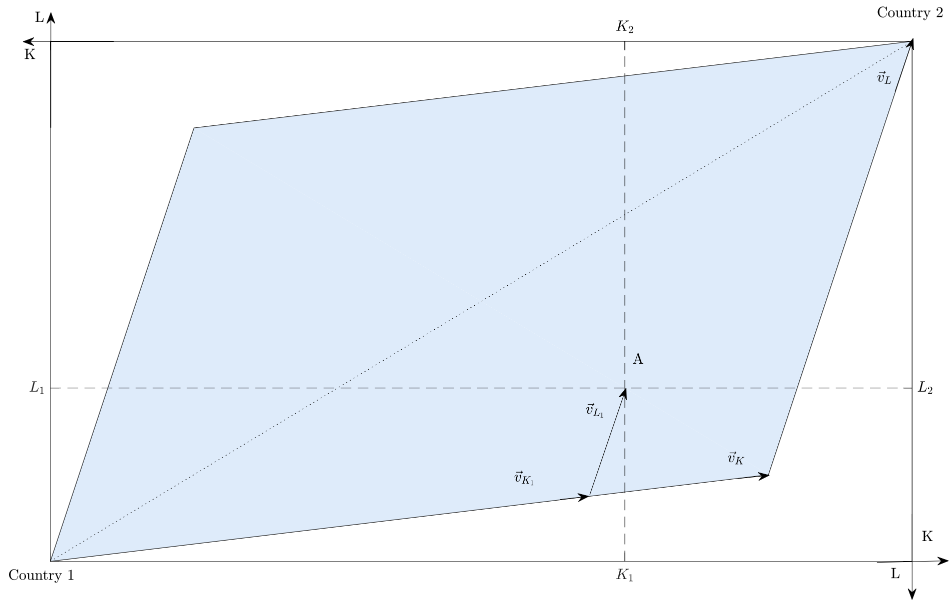
This result shows us that an easy way to check, for any given factor endowment, whether there will be FPE or not. But it shows even more than that: once we have solved for the constants a_1, b_1, a_2 and b_2, we know how much of each industry Samuelson's angel has allocated to every country, and therefore, we know each countries' production. Demand functions show us how much every country consumes of every good, and from the differences between production and consumption, we can calculate exports and imports. Problem Set 3 provides an example for these calculations.

Finally, this result lends itself to a nice graphical representation. Consider an Edgeworth box, where the axes represent the world's endowment of capital and labour. In this box, we can mark the endowments of both countries (point A).

Now, lets put on this diagram the factor choices of the integrated economy in every industry, that is, the two vectors $\overrightarrow{v_L} = \begin{pmatrix} l_L \\ k_L \end{pmatrix}$ and $\overrightarrow{v_K} = \begin{pmatrix} l_K \\ k_K \end{pmatrix}$. As $\alpha_L < \alpha_K$, we have $\frac{k_L}{l_L} < \frac{k_K}{l_K}$, so that in the graph, the slope of $\overrightarrow{v_K}$ is lower than that of $\overrightarrow{v_L}$. Furthermore, as $l_L + l_K = L$ and $k_L + k_K = K$, once we have drawn one vector, we can deduce the other directly.

To replicate the integrated economy, country 1's endowment must be a linear combination, with positive coefficients, of the vectors $\overrightarrow{v_K}$ and $\overrightarrow{v_L}$. All endowment points that verify this lie in the cone spanned by these two vectors. In the same way, the endowment of country 2 must lie inside the cone spanned by the two vectors seen from the origin of country 2. The intersection of these two cones is shaded in blue in Figure 10, and called the FPE set. In it, both Equation (69) and (70) hold, and therefore all endowment points achieve an equilibrium with FPE, replicating the integrated economy. In this diagram, we can even graphically solve the exact allocation of industries (that is, for a_1 and b_1) for the endowment point A, as shown in Figure 10.

Figure 10: The FPE Parallelogram



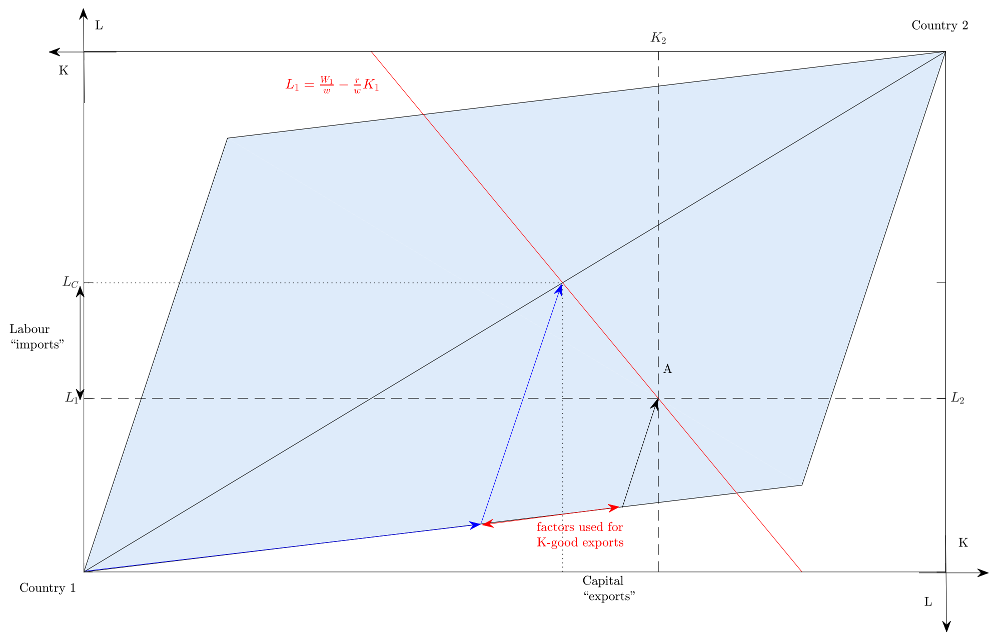
The figure shows that FPE occurs when countries' factor endowments are not too different. For instance, FPE becomes impossible if all labour or all capital are concentrated in one country: this forces countries to produce with extreme factor proportions, which do not coincide with the integrated economy ones. Furthermore, note that the size of the FPE set is determined by the difference in factor shares α_L and α_K. If factor shares are very different, the integrated economy produces with quite extreme factor proportions, and a larger set of endowment points lead to FPE. If they are very close, however, both industries produce nearly with the same capital-labour ratio, and the FPE set shrinks. In the extreme case where $\alpha_L = \alpha_K$, it is reduced to the diagonal of the Edgeworth box.

3.2 The Heckscher-Ohlin theorem

3.2.1 Trade and factor contents

What role does trade in goods play for replicating a world with perfect factor mobility? The following theorem characterizes the patterns of trade in an economy with FPE.

Figure 11: A graphical illustration of the Heckscher-Ohlin theorem



The Heckscher-Ohlin theorem *In the two country, two goods case, when* $\frac{K_1}{L_1} > \frac{K_2}{L_2}$ *(i.e., when country* 1 *is relatively more capital-intensive than country* 2*) and when there is FPE, country* 1 *exports the capital-intensive* K*-good and imports the labour-intensive* L*-good .*

You will do an analytical proof of this statement in Problem Set 3. However, we can already provide a nice graphical proof, introducing some useful new concepts.

In Figure 10, we cannot graph directly the amount of the two goods which are produced or consumed. Instead, we will focus on the amount of factors embedded in consumption and in production. How do we do that? First, note that a country's income is given by

$$\mathcal{W}_1 = wL_1 + rK_1 \iff L_1 = \frac{\mathcal{W}_1}{w} - \frac{r}{w}K_1,$$

where w and r are the integrated economy factor prices. Thus, for any income $\mathcal{W}_1$, we can draw an iso-income line, regrouping all combinations of capital and labour yielding the same national income. We are interested in one iso-income line in particular, the one passing through the endowment point A (marked in red in Figure 11). This iso-income line corresponds to the actual income of country 1 in the trade equilibrium. Now, consider the point where our iso-income line crosses the diagonal. This endowment point describes a down-scaled version of the integrated economy (with the same factor proportions as the latter) which has the same income as country 1. This small integrated economy would consume exactly what it produces. However, having the same income, the same preferences and facing the same prices as country 1[46], its consumption will be exactly equal to that of country 1.

This implies that the total amount of labour embedded in the consumption of country 1 is L_C, and the total amount of capital K_C. K_C is lower than the country 1's capital endowment K_1: thus, it is as if country 1 (which is more capital-intensive than country 2, because A lies below the diagonal) were "exporting" capital. Precisely, its production uses more capital than its consumption, which implies that it must export the more capital-intensive good. Capital does not move directly between countries, but indirectly, being embedded in different goods.

Section 5 shows that most empirical studies on the Heckscher-Ohlin model focus on this factor content of trade, rather than on actual quantities traded. Here, we can however be more precise, by drawing the production that would take place with the integrated economy factor proportions (blue vectors). The differences between the blue and the black vectors (which indicate how much factors are used in the production of country 1) indicate the amount of factors used for exports and imports.

Note that (unlike Ricardian models) Heckscher-Ohlin theory generates incomplete specialization: even a country with a high capital-labour ratio typically continues to produce the labour-intensive good.

3.2.2 Gains from trade

As there is FPE, any country j's real income in the trade equilibrium is equal to its real income in the integrated economy. Therefore, we can show that[47]

$$\left(\frac{\mathcal{W}_j}{P}\right)^{\text{FPE}} = \frac{wL_j + rK_j}{P} = \left((1-\alpha)\frac{L_j}{L} + \alpha\frac{K_j}{K}\right) ZK^{\alpha}L^{1-\alpha}. \tag{71}$$

So, in the trade equilibrium with FPE, every country gets a share of world income which is proportional to its capital and labour endowment. How does this compare to autarky? Solving for the autarky equilibrium in every country just we solved for the integrated economy equilibrium, we get[48]

$$\left(\frac{\mathcal{W}_j}{P_j}\right)^{\text{Autarky}} = ZK_j^{\alpha}L_j^{1-\alpha}. \tag{72}$$

[46]It has the same factor proportions as the integrated economy, and therefore the same prices, which are also the prices of country 1 by FPE.

[47]To see this, note $wL_j = wL\frac{L_j}{L}$, and then use Equation (63).

[48]Recall that we have set all productivities z_{ij} to 1. Therefore, $Z = \prod_{i=1}^{I}\left(\frac{1}{I}\left(\frac{\alpha_i}{\alpha}\right)^{\alpha_i}\left(\frac{1-\alpha_i}{1-\alpha}\right)^{1-\alpha_i}\right)^{\frac{1}{I}}$, and this term is identical in Equations (71) and (72).

and therefore,

$$\frac{\left({\mathcal{W}_j}/{P}\right)^{\text{FPE}}}{\left({\mathcal{W}_j}/{P_j}\right)^{\text{Autarky}}} = \frac{(1-\alpha)\,\frac{L_j}{L} + \alpha\frac{K_j}{K}}{\left(\frac{L_j}{L}\right)^{1-\alpha}\left(\frac{K_j}{K}\right)^{\alpha}}. \tag{73}$$

This ratio is always larger than 1, showing that there are gains from trade: national income in the trade equilibrium (with FPE) is higher than in autarky.[49]

What are the deep reasons for gains from trade? We can see these when reconsidering two concepts which we introduced in the Dornbusch-Fischer-Samuelson model. The integrated economy produces with the optimal capital/labour ratios in every industry. However, in autarky, the fact that a country must use all its production factors means that it cannot produce with these capital/labour ratios: instead, it is forced to use a lot of its abundant factor in all of its industries (i.e., it is forced to do factor deepening). With trade, it can change its production structure and produce more in the industries that use its abundant factor most (and export the part it does not consume). Structural transformation has replaced factor deepening.

However, while overall national income increases through trade, the gains from trade are unevenly distributed between workers and capitalists. We turn to this in the next section.

3.2.3 Winners and losers

In the integrated economy, real wages and rental rates were given by Equation (68):

$$\frac{w}{P} = (1-\alpha)\,Z\left(\frac{K}{L}\right)^{\alpha} \text{ and } \frac{r}{P} = \alpha Z\left(\frac{K}{L}\right)^{\alpha-1}.$$

In autarky, instead, real wages and rental rates in country j are

$$\frac{w_j}{P_j} = (1-\alpha)\,Z\left(\frac{K_j}{L_j}\right)^{\alpha} \text{ and } \frac{r_j}{P_j} = \alpha Z\left(\frac{K_j}{L_j}\right)^{\alpha-1}.$$

Consider a country which is more capital-intensive than the rest of the world (${K_j}/{L_j} > {K}/{L}$). The above formulas show that in this country, trade will increase the real rental rate, but lower the real wage. More generally, in every country, trade leads to gains for the relatively abundant factor, but to losses for the relatively rare factor.

Why does this happen? In this capital-intensive economy, in autarky, capital is relatively abundant, and therefore has a low marginal product and earns a low return. Labour, on the other hand, is

[49] Formally, this ratio is larger than 1 if $(1-\alpha)\ln\left(\frac{L_i}{L}\right) + \alpha\ln\left(\frac{K_i}{K}\right) \leq \ln\left((1-\alpha)\,\frac{L_i}{L} + \alpha\frac{K_i}{K}\right)$. This latter statement holds for any $\alpha \in (0,1)$, because the natural logarithm is a concave function.

relatively rare, which means that it earns a high marginal product and high wages. When opening up to the rest of the world, however, the economy specializes in capital-intensive goods: the demand for capital increases, the demand for labour falls, and therefore real rental rates increase and real wages fall.

Thus, Heckscher-Ohlin trade theory shows that trade brings about systematic changes in the income distribution, with some categories of the population losing and others gaining. As we have shown in the previous section, the winners must always win more than the losers lose (as total national income goes up). Therefore, in theory, it would be possible that the winners compensate the losers through an appropriate transfer scheme. As we will see in Chapter VII, recent studies for the United States show however that this compensation hardly happens in practice.

Importantly, all results shown in this section (the Heckscher-Ohlin theorem, gains from trade and the existence of winners and losers) are also true if there is no FPE. Intuitions are the same, just proofs are a little harder.

Things get more difficult, however, when we allow for more than two goods, factors or countries. Indeed, as in the Ricardian case, it is not straightforward to generalize intuitions: when adding a third good with $\alpha_L < \alpha_M < \alpha_K$, how do trade patterns change? How do you define a capital-intensive country when there are three production factors? We will study first how introducing more factors and goods affects the logic of FPE, and then study the general predictions of the Heckscher-Ohlin model, with many goods and countries (but in general limited to two factors).

3.3 FPE with more than two goods, factors or countries

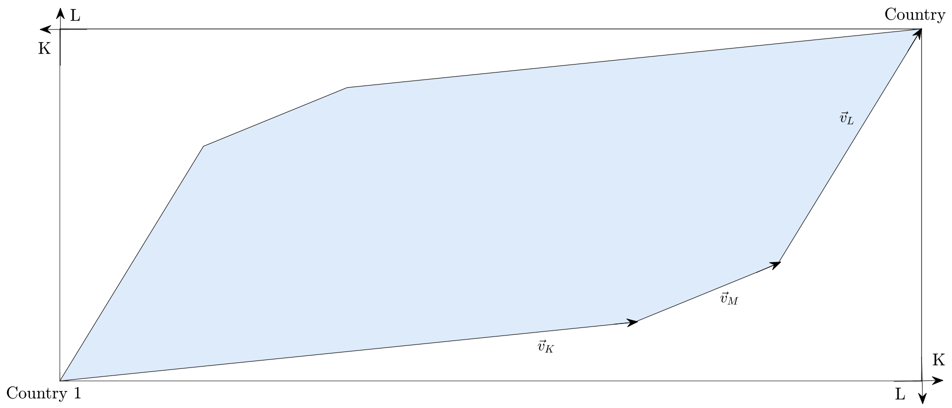
Adding more than two goods, factors or countries to our model does not change the logic of factor price equalization: as long as Samuelson's angel respects its allocation rule, FPE will be guaranteed.

Three goods For instance, suppose there is a third good with capital share α_M, holding $\alpha_L < \alpha_M < \alpha_K$. Then, Equation (69) is replaced by

$$\begin{pmatrix} L_1 \\ K_1 \end{pmatrix} = a_1 \begin{pmatrix} l_L \\ k_L \end{pmatrix} + b_1 \begin{pmatrix} l_K \\ k_K \end{pmatrix} + c_1 \begin{pmatrix} l_M \\ k_M \end{pmatrix}. \tag{74}$$

This is a linear system with two equations and three unknowns. Therefore, it admits either no solution (meaning that there cannot be FPE) or an infinity of solutions. In the latter case, FPE is possible,

Figure 12: The FPE set with three goods



but production patterns are indeterminate. Figure 12 illustrates the FPE set.

However, even though production is indeterminate, we can do the same analysis as before with respect to the factor content of production and consumption. The predictions of the Heckscher-Ohlin model on the factor content of trade are much more general than its predictions on the actual goods traded, which is why empirical studies have focused on the former.

Three factors Now, suppose that there is a third production factor, land, denoted by N, and that the production functions of each good are given by $y_{ij} = k_{ij}^{\alpha_i} n_{ij}^{\beta_i} l_{ij}^{1-\alpha_i-\beta_i}$. Still, the same reasoning as before is valid, and Equation (69) is replaced by

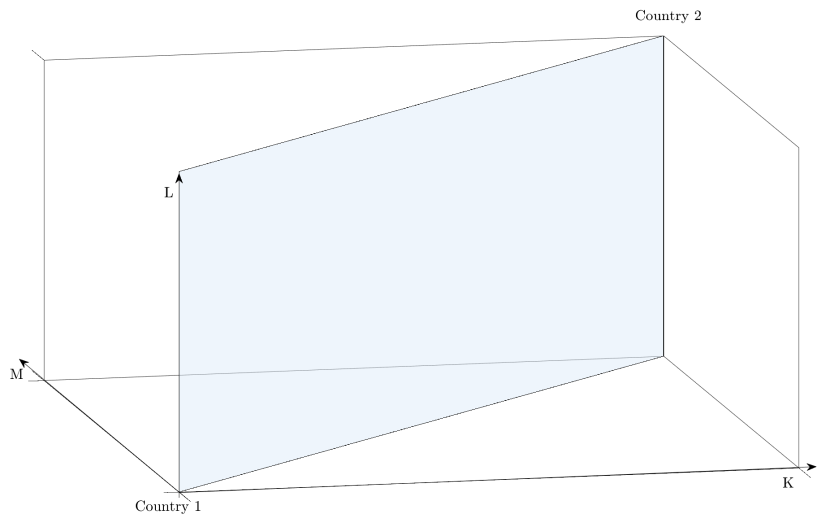
$$\begin{pmatrix} L_1 \\ N_1 \\ K_1 \end{pmatrix} = a_1 \begin{pmatrix} l_L \\ n_L \\ k_L \end{pmatrix} + b_1 \begin{pmatrix} l_K \\ n_K \\ k_K \end{pmatrix}. \tag{75}$$

This is a linear system of three equations with two unknowns. Therefore, it has a solution only if these equations are not independent (i.e., if one of them can be written as a linear combination of the others). FPE is still possible, but more difficult: here, in the three-dimensional factor space, the FPE set will be a two-dimensional object (a plane), as shown in Figure 13.

Intuitively, we can conclude that increasing the number of goods with respect to the number of factors increases the possibilities for FPE, because it gives countries more possibilities to specialize. Increasing the number of factors, however, makes FPE more difficult.[50]

[50] As previously mentioned, introducing productivity differences makes FPE next to impossible. For instance, if we

Figure 13: The FPE set with three factors



4 Heckscher-Ohlin theory without factor price equalization

We now turn to the general case of the Heckscher-Ohlin model, with many countries and goods (but still only two production factors), and without imposing FPE. In this case, solving the model is not as easy as in the FPE set, but we can still show some important results, including some of the most famous theorems in international trade.

4.1 Equilibrium conditions

We still retain all assumptions from Section 3.1.1, but now characterize the model's equilibrium without taking a stand on whether there is FPE or not.

First, note that a country produces a given industry iff it has the lowest marginal costs worldwide. Then, these lowest marginal costs equal the world market price. Therefore, for every country j and for every industry i,

$$\left(\frac{w_j}{1-\alpha_i}\right)^{1-\alpha_i}\left(\frac{r_j}{\alpha_i}\right)^{\alpha_i} \geq p_i. \tag{76}$$

The equation holds with equality if country j produces good i, and as a strict inequality otherwise. Furthermore, we know that in every industry in which country j produces, $w_j l_{ij} = (1-\alpha_i)\, p_i y_{ij}$.

would have $z_{K1} > z_{K2}$ and $z_{L1} < z_{L2}$, the FPE set in Figure 10 would be reduced to one point: the only way to replicate the integrated economy would be if country 1 had exactly as much labour and capital as the K-industry used in the integrated economy, and country 2 exactly as much as the L-industry.

Using this, we can rewrite national factor market clearing conditions as

$$\sum_{i=1}^{I} (1-\alpha_i) \frac{p_i y_{ij}}{w_j} = L_j. \tag{77}$$

$$\sum_{i=1}^{I} \alpha_i \frac{p_i y_{ij}}{r_j} = K_j, \tag{78}$$

where $y_{ij} = 0$ if country j does not produce good i in equilibrium.

The system of equations formed by (76), (77) and (78) pins down the solutions of the model for w_j, r_j and the y_{ij} as a function of the set of world market prices $(p_i)_{i\in\{1,..,I\}}$. Calculating this solution is, however, very cumbersome, because we need to make a lot of case distinctions: for every inequality, we have to check whether it holds strictly or not. And this is only the first step: afterwards, we need to calculate equilibrium prices, bringing demands back into the picture.

Thus, the Heckscher-Ohlin model is difficult to solve outside the FPE set.[51] However, fortunately, it turns out that even in this complicated setup, predictions on the factor content of trade are easy to derive, as we will see in Section 5. Furthermore, a range of important results can be shown with a partial equilibrium analysis, taking world goods prices as given. The next section illustrates how this works, paving the way for the Rybczynski and Stolper-Samuelson theorems.

4.2 Diversification, structural transformation, and factor deepening

To start with, assume there are only two goods, a K-good and an L-good, with $\alpha_K > \alpha_L$. Then, a country can be in one of three possible situations: it can produce only the L-good, produce only the K-good or produce both goods.

It can be shown that the position of the country only depends on its capital-labour ratio. A country with a high capital-labour ratio produces only the K-good, a country with a low capital-labour ratio produces only the L-good, and a country with an intermediate capital-labour ratio produces both.[52]

[51] The model is difficult because we have to distinguish a multitude of cases, depending on whether a country produces a certain good or not. This is exactly the same problem as in the Ricardian model, where we saw that the Dornbusch-Fischer-Samuelson model was able to overcome it by introducing a continuum of goods. Could this be a solution for the Heckscher-Ohlin model, too? It turns out that it is, as shown by (of course) Dornbusch, Fischer and Samuelson (1980). The model is not quite as easy as the Ricardian one, but it does get rid of the cumbersome case distinctions.

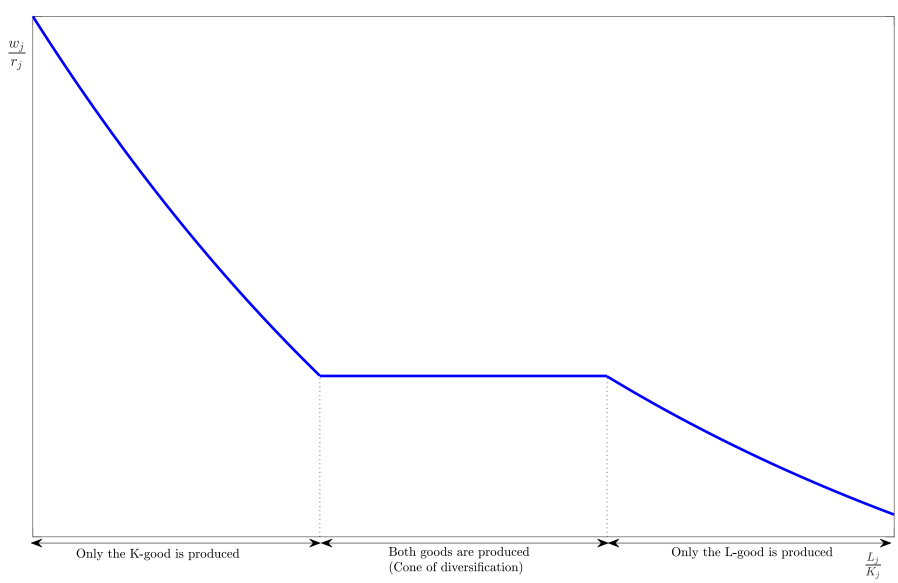
[52] How can one derive the exact cut-offs for the capital-labour ratio? Consider a country which produces only the K-good. For this country, Equation (76) must hold with equality for the K-good. Furthermore, $l_{Kj} = L_j$ and $k_{Kj} = K_j$: all factors are used for producing the K-good. Knowing this, we can use Equations (77) and (78) to determine w_j and r_j. Then, for our initial hypothesis that the country produces only the K-good to be valid, it must be that these wages and prices imply a marginal cost for the L-good which is strictly higher than p_L. This condition provides the cut-off for the capital-labour ratio, showing that the country effectively produces only the K-good if $\frac{K_j}{L_j} > \left(\frac{p_K}{p_L}\right)^{\frac{1}{\alpha_L-\alpha_K}} \left(\frac{1-\alpha_L}{1-\alpha_K}\right)^{\frac{1-\alpha_L}{\alpha_K-\alpha_L}} \left(\frac{\alpha_L}{\alpha_K}\right)^{\frac{\alpha_L}{\alpha_K-\alpha_L}}$. Likewise, it only produces the L-good if

Furthermore, we can determine the value of the wage-rental ratio in each of these different cases. In countries which specialize completely, Equation (59) shows that this ratio is given by the countries' endowment, and we have $\frac{w_j}{r_j} = \frac{1-\alpha_i}{\alpha_i}\left(\frac{K_j}{L_j}\right)$, where i stands for the good in which the country specializes. In countries which produce both goods, the wage-rental ratio can be obtained by solving the system formed by Equation (76), which holds with equality for both goods. This gives

$$\frac{w_j}{r_j} = \begin{cases} \frac{1-\alpha_K}{\alpha_K}\frac{K_j}{L_j} & \text{if the country produces only the } K\text{-good.} \\ \left(\frac{\alpha_L^{\alpha_L}}{\alpha_K^{\alpha_K}}\frac{(1-\alpha_L)^{1-\alpha_L}}{(1-\alpha_K)^{1-\alpha_K}}\frac{p_L}{p_K}\right)^{\frac{1}{\alpha_K-\alpha_L}} & \text{if the country produces both goods.} \\ \frac{1-\alpha_L}{\alpha_L}\frac{K_j}{L_j} & \text{if the country produces only the } L\text{-good.} \end{cases} \tag{79}$$

Figure 14 plots the wage-rental ratio as a function of a country's labour-capital ratio. It shows that if the country is completely specialized (that is, produces only one good), the wage-rental ratio decreases in the labour-capital ratio. However, when the country is not specialized, factor prices do not depend on the capital-labour ratio. The intermediate zone where this happens is called a cone of diversification.

Figure 14: Wage-rental and Labour-capital ratios



Outside the cone of diversification, if a country gets more labour-intensive, it uses more labour for the production of the good in which it is specialized. As there are decreasing returns to labour, this lowers

$\frac{K_j}{L_j} < \left(\frac{1-\alpha_L}{1-\alpha_K}\right)^{\frac{1-\alpha_K}{\alpha_K-\alpha_L}} \left(\frac{\alpha_L}{\alpha_K}\right)^{\frac{\alpha_K}{\alpha_K-\alpha_L}} \left(\frac{p_K}{p_L}\right)^{\frac{1}{\alpha_L-\alpha_K}}$.

the marginal product of labour, and the wage-rental ratio declines. Inside the cone of diversification, however, trade allows the country to escape decreasing returns through structural transformation. Now, if a country gets more labour-intensive, it does not use more labour in all industries, but instead changes its production mix, producing more of the labour-intense good and less of the capital-intense one. This is the essence of the famous Rybczynski theorem, explained in the next section.

Note that all countries inside the cone of diversification have the same factor prices. Thus, if all countries of the world are inside the same cone of diversification, there is FPE.[53] Otherwise, some countries with extreme factor proportions must specialize completely. The predictions of the Heckscher-Ohlin theorem still hold true for them: they produce the good which uses their relatively abundant factor most intensely, and export it in order to finance imports of the other good. Trade increases their national income and the return to their abundant factor, while it lowers the return to their scarce factor. However, factor prices do not move enough to be equalized to those of countries in the cone.

4.3 Two classic theorems

4.3.1 The Rybczynski theorem

Theorem. *(Rybczynski, 1955) For a small country belonging to a cone of diversification, a small increase in the capital endowment does not affect factor prices. Instead, it leads to a more than proportional increase of the production of K-goods, and a decrease in the production of L-goods.*

Of course, the exact analogue statement holds for an increase in the labour endowment. Note that the theorem is formulated for a small country: this implies that changes in the country's endowments do not affect world prices, so that we can take these as given.

Proof. We have already shown in Section 4.2 that for a country in the cone of diversification, the wage-rental ratio does not depend on factor endowments. Therefore, as long as the increase in capital is small enough not to move the country outside of the cone, factor prices do not change.[54]

Furthermore, we can combine Equations (77) and (78) in order to eliminate $p_L y_{Lj}$ and get

$$y_{Kj} = \frac{1}{p_K}\left(\frac{\alpha_L w_j}{\alpha_K - \alpha_L}\right)\left(\frac{1-\alpha_L}{w_j}\frac{r_j}{\alpha_L}K_j - L_j\right)$$

[53] If we consider more than two goods, the number of cones of diversification may change.

[54] The wage-rental ratio does not change, and as goods prices do not change, neither the wage nor the rental rate changes, as can be seen from Equation (76).

This equation shows that an increase in K_j increases the production of the K-good, while an increase in L_j lowers the production of the K-good. This is exactly what the Rybczynski theorem predicts: the only thing left to show is that the increase in y_{Kj} triggered by an increase in K_j is more than proportional, that is, that the elasticity of y_{Kj} with respect to K_j is larger than 1. After some algebra, it comes that

$$\frac{\partial y_{Kj}}{\partial K_j}\frac{K_j}{y_{Kj}} = \frac{1}{1 - \frac{\alpha_L}{1-\alpha_L}\left(\frac{w_j L_j}{r_j K_j}\right)} = \frac{1}{1 - \frac{\alpha_L(1-\alpha)}{\alpha(1-\alpha_L)}},$$

where the last equality is due to the fact that our country produces both the K-good and the L-good, and we therefore have, as in the integrated economy, $\frac{w_j}{r_j} = \frac{1-\alpha}{\alpha}\frac{K_j}{L_j}$, with $\alpha = \frac{\alpha_K+\alpha_L}{2}$. Now, as $0 < \frac{\alpha_L}{\alpha} < 1$ and $0 < \frac{1-\alpha}{1-\alpha_L} < 1$, we have $0 < \frac{\alpha_L(1-\alpha)}{\alpha(1-\alpha_L)} < 1$, which completes the proof. □

The Rybczynski theorem shows that an open economy can react to endowment changes by structural transformation, and need not change its factor prices. This makes it very different from a closed economy, where an increase in the endowment of one factor must lead to a fall in its relative price.

This result has played an important role in the discussion on the consequences of immigration in rich countries, generally interpreted as an increase in the supply of low-skilled workers. While many commentators are worried that low-skilled immigration depresses the wages of domestic low-skilled workers, the Rybczynski theorem suggests that this need not occur in an open economy: instead, there may just be an expansion of industries using low-skilled labour intensively. Several studies have found support for this mechanism. Card (1990) studied the 1980 “Mariel boat lift”, an episode during which the Cuban government allowed 125’000 people to emigrate to the United States (and essentially, to Miami). Card shows that although the Miami labour force increased by 7%, wages of low-skilled workers (and unemployment) evolved just as in other, comparable US cities. Hanson and Slaughter (2002), in a broader study on the effects of immigration in the United States, also found some evidence for Rybczynski effects. However, recently, Borjas (2017) has criticized Card’s study (and especially, Card’s choice of cities to which Miami should be compared). Re-examining the effect of the Mariel boatlift, he finds that actually, there was a substantial wage effect: “*the wage of high school dropouts in Miami dropped dramatically, by 10 to 30 percent*”. The debate on the economic effects of immigration remains open.

Another interesting application of the Rybczynski theorem is Ventura (1997), who studies the case of the East Asian “miracle economies” (South Korea, Hong Kong, Taiwan and Singapore) in the 1990s. These countries had an extremely high investment rate, and thus accumulated capital very fast. Still, the returns to capital did not seem to fall, which was puzzling to many observers. Ventura notes, however, that this was just a consequence of the Rybczynski theorem's structural transformation: the miracle economies were able to avoid decreasing returns to capital by trading with the rest of the world and specializing ever more in capital-intensive industries.

Finally, the reallocation of resources implied by the Rybczynski effect may also have negative consequences, related to the famous "Dutch Disease" debate. This term was coined by *The Economist* in 1977, who noted that the discovery of large natural gas reserves in the Netherlands in the 1950s had lead to an expansion of the natural gas industry, but drawn labour and capital away from the manufacturing sector. In the presence of externalities such as learning-by-doing (likely to be stronger in manufacturing than in natural gas), this could have negative long-run consequences (Corden and Neary, 1982; Krugman, 1987). However, Allcott and Keniston (2017) argue that this does not appear to be the case in the data: "*tradable manufacturing subsectors do contract during resource booms, but their productivity is unaffected, so there is no evidence of foregone local learning-by-doing effects*".

4.3.2 The Stolper-Samuelson theorem

Theorem. *(Stolper and Samuelson, 1941) For a small country belonging to a cone of diversification, an increase in the world market price of the K-good increases the price of the production factor used intensively for K-good production more than proportionally, and decreases the price of the other factor.*

Proof. Consider Equation (76), which must hold with equality for both goods for a country belonging to the cone of diversification. Taking natural logarithms, we get

$$(1-\alpha_L)(\ln(w_j) - \ln(1-\alpha_L)) + \alpha_L(\ln(r_j) - \ln(\alpha_L)) = \ln(p_L)$$

$$(1-\alpha_K)(\ln(w_j) - \ln(1-\alpha_K)) + \alpha_K(\ln(r_j) - \ln(\alpha_K)) = \ln(p_K)$$

Solving this system gives, after some algebra,

$$\ln(r_j) = Constant + \frac{1-\alpha_K}{\alpha_L - \alpha_K}\ln(p_L) - \frac{1-\alpha_L}{\alpha_L-\alpha_K}\ln(p_K)$$

Recall that you can calculate the elasticity of a variable x with respect to a variable y as $\frac{\partial \ln(x)}{\partial \ln(y)}$.

Therefore, we get

$$\frac{\partial \ln(r_j)}{\partial \ln(p_L)} = \frac{1-\alpha_K}{\alpha_L - \alpha_K} < 0 \text{ and } \frac{\partial \ln(r_j)}{\partial \ln(p_K)} = \frac{1-\alpha_L}{\alpha_K - \alpha_L} > 1$$

□

The Stolper-Samuelson theorem reiterates the effects of trade on the income distribution, which we have encountered already in Section 3.2.3, generalizing them to a case without FPE. It has had a long-lasting impact since its discovery in 1941, "*because it proved something seemingly obvious to non-economists: free trade with low-wage nations could hurt workers in a high-wage country.*"[55]
Indeed, the theorem (and Heckscher-Ohlin trade theory more generally) have often been invoked to explain increasing wage inequality in developed countries over the last 30 years, and in particular, the very low wage growth for low-skilled workers. Indeed, as developing countries are relatively well-endowed with low-skilled labour, their entry into world trade since the 1980s should lead, according to the theory, to a fall in the relative price of low-skilled goods, a decrease in wages for low-skilled workers and an increase in wages for high-skilled workers in rich countries.
We will have a deeper look at the empirical evidence on these issues in Chapter VII. For now, just note that while the Stolper-Samuelson theorem (and Heckscher-Ohlin theory) may a priori be able to explain the evolution of wage inequality in rich countries, there are also problems. Most importantly, the theory also predicts that in poor countries, trade should lead to an increase in low-skilled wages and a fall in high-skilled wages, and therefore a reduction in wage inequality. This has not happened, as has been pointed out by critical studies (Davis and Mishra, 2007).

Now that we know the Heckscher-Ohlin model's main predictions, we turn to its empirical assessments. As the next section shows, they have had a long and changing history.

5 Empirical tests of the theory

5.1 The Leontief paradox

The simplest possible Heckscher-Ohlin model, with two countries, two goods and two factors, predicts that relatively capital-intensive countries export relatively capital-intensive goods. This prediction

[55] This quotation is from an article by *The Economist*, which recently included the Stolper-Samuelson theorem in a series on five great economic ideas. See http://www.economist.com/news/economics-brief/21703350-third-our-series-looks-stolper-samuelson-theorem-inconvenient-iota.

in mind, Leontief (1953) conducted the first empirical test of the theory. Leontief calculated how much capital and labour was used to produce US exports and US imports in 1947, using a newly developed input-output table.[56] In 1947, with the European countries still strongly affected by wartime destructions, the United States were clearly the most capital-intensive country in the world. However, Leontief's calculations yielded, to a general surprise, that the capital-labour ratio of US exports was lower than the capital-labour ratio of US imports!

Leontief proposed a tentative explanation for his paradox, arguing that maybe labour should not be measured just by the number of workers. Probably, American workers were more productive than workers of other countries: thus, when measuring labour in "efficiency units", it could turn out that the United States were actually relatively labour-intensive rather than relatively capital-intensive.

While Leontief's result was highly influential, later researchers criticised his methodology. Leamer (1980) showed that with trade deficits, or with more than two factors, the Heckscher-Ohlin model may actually be compatible with Leontief's findings. Most importantly, he argued that his test was misguided: the Heckscher-Ohlin model does not make strong predictions on the relative factor content of exports and imports. However, it does make strong predictions on the relative factor content of consumption and production, as we have seen in Section 3. These predictions would become crucial for testing the theory, and we return to them in the next section, generalizing our previous results.

5.2 The Heckscher-Ohlin-Vanek equations

Consider the basic Heckscher-Ohlin model with arbitrarily many countries, goods and factors, and assume that the trade equilibrium achieves FPE.

Denote by $\mathbf{a}_L$ a vector giving for any good the units of labour which are demanded to produce one unit of the good. As there is FPE, and there are no technology differences, this vector is the same in all the countries. Write $\mathbf{y}_j$ for the vector of productions of every good in country j, and $\mathbf{c}_j$ for the vector of consumptions. Then, define

$$\mathbf{t}_j = \mathbf{y}_j - \mathbf{c}_j.$$

$\mathbf{t}_j$ is called the net export vector of country j. It contains for every good the difference between

[56] Leontief himself was instrumental for creating the first input-output tables, and was awarded the Nobel Prize for them in 1973. These tables list how much each industry uses the goods of other industries. We need them here because most goods do not only use capital and labour in production, but also intermediate inputs. To determine the total amount of capital and labour which went into production, one needs to add back the labour and capital used to produce intermediates (and therefore needs to know how many and which intermediates were used). Leontief only had an input-output table for the United States: to calculate the capital and labour used in the production of US imports, he therefore needed to assume that all other countries of the world have the same input-output table.

production and consumption (positive if the country exports a good, negative if it imports it). From this, we can determine the labour content of trade as

$$t_{Lj} = \langle \mathbf{a}_L, \mathbf{t}_j \rangle$$

When t_{Lj} is positive, country j is a net exporter of labour: more labour is embedded in its exports than in its consumption. When t_{Lj} is negative, then the country is a net importer of labour. The Heckscher-Ohlin model allows us to relate these factor contents to a country's factor endowments. To do so, note that labour market clearing implies $\langle \mathbf{a}_L, \mathbf{y}_j \rangle = L_j$. Furthermore, the labour content of country j's consumption is $\langle \mathbf{a}_L, \mathbf{c}_j \rangle$. As all countries have the same Cobb-Douglas preferences and face the same prices, the share of a country in the world's consumption of a given good is equal to that country's share in world income.[57] Therefore, $\mathbf{c}_j = s_j \mathbf{c} = s_j \mathbf{y}$, where s_j is country j's share in world income and the last equality comes from the fact that world consumption equals world production. Therefore, we get

$$t_{Lj} = L_j - s_j \langle \mathbf{a}_L, \mathbf{y} \rangle$$

which simplifies to[58]

$$t_{Lj} = L_j - s_j L. \tag{80}$$

The same analysis can obviously be done for every production factor. Writing $\mathbf{t}_j$ for the vector of net factor exports and $\mathbf{v_j}$ ($\mathbf{v}$) for country j's (the world's) endowment vector, we get the Heckscher-Ohlin-Vanek (HOV) equations, named after Vanek (1968):

$$\mathbf{t}_j = \mathbf{v}_j - s_j \mathbf{v}. \tag{81}$$

The HOV equations state that a country should "export" a factor in which is relatively well endowed (in the sense that its share in the world supply of that factor is higher than its share in world income). In this 1980 paper, Leamer suggested that this is the equation that Leontief should have tested.

[57] This is a direct consequence of the fact that with CES (and therefore also Cobb-Douglas) preferences, the share of income spend on a certain good does not depend on the level of income, and is therefore the same across all countries if they all face the same prices. So, we have, for any good i, $p_i c_{ij} = x\mathcal{W}_j$ and $p_i c_i = x\mathcal{W}$ (where x stands for the common spending share), so that $c_{ij}/c_i = \mathcal{W}_j/\mathcal{W}$.

[58] Here, we need FPE again: otherwise, $\mathbf{a}_L$ varies across countries, and it is not true that $\left\langle \mathbf{a}_L^j, \mathbf{y} \right\rangle = L$.

5.3 The HOV equations and empirical tests

A negative result The HOV equations were first tested by Bowen, Leamer and Sveikauskas (1987), who collected data on exports, imports and 12 production factors in 27 countries. With this, they first calculated the factor content of countries' net exports (i.e., the left hand side of the HOV equations).[59] Then, they separately calculated the right hand side, using their measurements of factor endowments. Given the measurement errors involved, they argue that one cannot reasonably expect these two objects to be exactly equal. Thus, they proposed (among other things) a "sign test", asking how often t_{vj} and $V_j - s_j V$ have the same sign. For the 324 comparisons ($12 * 27$) they made, this was only true in 61% of the cases, which is not great: even tossing a coin would have gotten 50%... Other tests were equally disappointing, and so, it seemed that the Heckscher-Ohlin model did not very well at explaining real-world trade.

Putting productivities back into the picture However, this was not the end of it. Indeed, rejecting the HOV equations does not mean that the Heckscher-Ohlin model is flawed: the equations are only valid in a special case of the model, when endowments are such that there is FPE. As real-world factor prices are far from being equalized across countries, it is unsurprising that a test based on FPE failed miserably. Moreover, the basic Heckscher-Ohlin model had made the assumption that there were no productivity differences across countries. Heckscher and Ohlin indeed believed that technology diffusion would quickly make the same productivities available to all countries, but they may have misguided in their optimism.

In 1993, Trefler reintroduced productivity differences into the Heckscher-Ohlin model. He did so in a special way, following Leontief's intuition that production factors in different countries may have different productivities (Trefler's paper is entitled "*International Factor Price Differences: Leontief Was Right!*"). Thus, Trefler defined the "effective" labour and capital endowments of a country, $\hat{L}_j$ and $\hat{K}_j$, as

$$\hat{L}_j = a_{Lj} L_j \text{ and } \hat{K}_j = a_{Kj} K_j$$

where a_{Lj} and a_{Kj} are country-specific productivities of labour and capital, and the production function for any good i is $y_{ij} = \hat{k}_{ij}^{\alpha_i} \hat{l}_{ij}^{1-\alpha_i}$. This model is equivalent to the basic Heckscher-Ohlin model: once we replace capital and labour with effective capital and effective labour, all the previous results

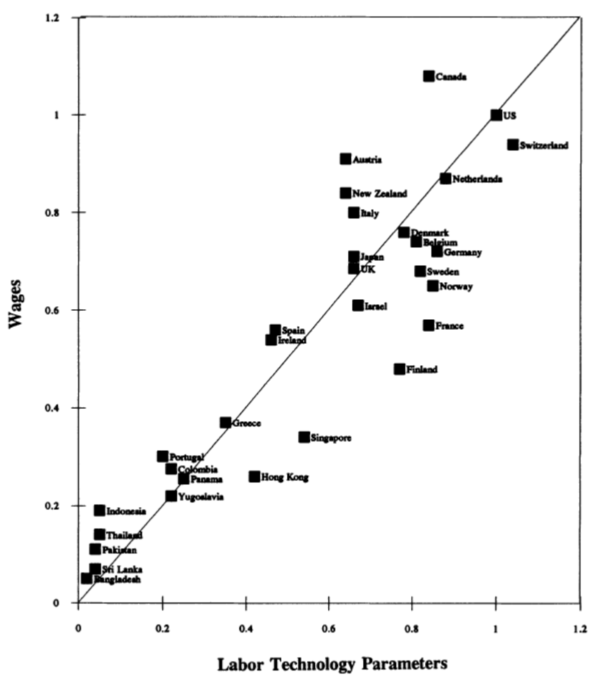
[59] Just as Leontief, they needed to use an input-output table, and just as Leontief, they used the US input-output table for all countries.

derived in this chapter go through. In particular, FPE will now be "conditional": prices for effective factors are equalized, which means that actual factor prices are pinned down by productivity ratios. Taking the example of labour, where $\hat{w}_j = \frac{w_j}{a_{Lj}}$, if $\hat{w}_j$ is equalized across countries, $\frac{w_j}{w_{j'}} = \frac{a_{Lj}}{a_{Lj'}}$.

This augmented model can fit the HOV equations perfectly, but success is now too easy, because we have too many degrees of freedom. Indeed, we have as many productivity parameters a_{ij} as there are HOV equations, meaning that with the appropriate choices for the a_{ij}'s, all of the HOV equations hold. How can we then test if the theory is true?

Trefler proposed two methods. First, one can check whether the a_{ij}'s that make the HOV equations hold are positive: this must of course be the case for the theory to make sense, but there is nothing that guarantees them to be positive if the theory is inaccurate. Second, one can check if the a_{ij}'s make sense: if the theory is true, and we are in FPE, then for instance wages across countries should hold $\frac{w_j}{w_{j'}} = \frac{a_{Lj}}{a_{Lj'}}$. Figure 15 shows that this works quite well: the labour productivity ratios obtained in Trefler's calculations are closely aligned with relative wages in reality.

Figure 15: Relative wages and labour productivities in Trefler's model



Source: Trefler (1993).

Trefler (1995) proposed a different approach: instead of allowing for different productivities for every factor, one could also just assume productivity is a country-specific constant ($a_{ij} = a_j$). In that case, it is no longer guaranteed that the HOV equations hold with equality. However, Trefler finds that this model already makes them hold reasonably well, and concludes that simple, uniform productivity differences can already go a long way in reconciling the Heckscher-Ohlin model with the data. Finally, Davis and Weinstein (2001) have extended the basic Heckscher-Ohlin model to account not only for productivity differences, but also for non-traded goods, and show that this further improves its performance.

In sum, this empirical literature shows that the basic Heckscher-Ohlin model, without any productivity differences, does poorly in explaining world trade. This does not mean that the model is useless, but just that it is crucial to also take into account other factors determining world trade, most importantly, the cross-country productivity differences emphasized by Ricardian models.

Part IV

Increasing Returns

1 Introduction

Both the Ricardian and the Heckscher-Ohlin trade theory explain why countries exchange different goods, specializing according to comparative advantage. They also suggest that countries that trade should be dissimilar: the larger their differences in relative productivities and/or relative factor endowments, the larger are gains from trade.

Until World War 2, these intuitions seemed compatible with the patterns of world trade, marked by the "Great Specialization", in which an industrialized core exchanged manufacturing goods against primary goods from the periphery. However, trade patterns began to change definitively after 1945, when there was a great expansion of trade between developed nations, driven by intra-industry trade. Table 6 illustrates these two trends for Great Britain. In 1913, most of Britain's trade was with non-European countries, while in 1992, European countries represented over 60% of both exports and imports. Furthermore, in 1913, Britain exported manufacturing goods and imported non-manufacturing goods. In 1992, Britain still exported manufacturing goods, but not also imported mainly manufacturing goods. Furthermore, going into the details would show that this trade was taking place within more narrowly defined manufacturing industries: it is not that Britain would export cars and import TV sets, but instead, Britain both imports and exports cars, and imports and exports TV sets. This type of North-North, intra-industry trade is difficult to explain with theories of comparative advantage.

Table 6: The rise of North-North and intra-industry trade: the example of Great Britain

Table 3. Commodity and Geographical Composition of U.K. Trade
Percent[a]

Year	*Exports of manufactures*	*Imports of manufactures*	*Exports to Europe*	*Imports from Europe*
1913	75.5	20.2	39.5	44.6
1992	81.9	78.4	63.8	63.7

Source: Mitchell (1988), Barraclough (1978), and Great Britain, Central Statistical Office (1994).
a. Numbers are percentages of the relevant piece of trade. Export columns indicate percentage of total merchandise exports; import columns indicate percentage of total merchandise imports.

Source: Krugman (1995).

As early as 1933, Bertil Ohlin had come up with an alternative reason for trade: if production had increasing returns, there could be a motive to specialize and trade even for completely identical countries (Ohlin, 1933).[60] Imagine, for instance, a world with two countries and two goods, A and B. When the production of these goods has increasing returns, it is more efficient to have one large firm in the first country producing good A (and exporting it to the second country), and one large firm in the second country producing good B, rather than two small firms in every market. While this argument could explain trade between similar nations, in 1966, Bela Balassa added another point which could account for the intra-industry nature of trade: goods traded within industries were not exactly the same, but differentiated, such as different types or brands of cars or clothes (Balassa, 1966).

Thus, the elements to explain North-North intra-industry trade were in place quite early. However, as Krugman explains in his 2008 Nobel Prize Lecture, "*this straightforward-seeming explanation [...] was not at all part of the standard corpus of international trade theory circa 1975. [...] Why? The answer was that unexhausted economies of scale at the firm level necessarily imply imperfect competition, and there were no readily usable models of imperfect competition to hand"* (Krugman, 2008).

This deadlock was overcome through a number of advances in the theory of Industrial Organization, which combined product differentiation and increasing returns to yield models of monopolistic competition. In these models, increasing returns make every firm a monopolist for some variety of a differentiated good, but the firm still competes with the producers of the other varieties. The model of Dixit and Stiglitz (1977), in particular, proved to be the "readily usable model of imperfect competition" trade theory had been waiting for. It allowed Krugman and others to develop "New Trade Theory", which could finally explain North-North, intra-industry trade in a rigorous way, and delivered a number of new predictions. Furthermore, through its focus on firms, it paved the way for the next major innovation, the incorporation of firm heterogeneity in trade models (see Chapter V).

2 Increasing returns, product differentiation and trade

2.1 Monopolistic competition

The typical monopolistic competition model assumes that an industry produces not just one homogeneous good, but a large number of differentiated varieties. For example, the car industry produces sports cars, family vans, small city cars, etc., and offers a large number of brands with different characteristics.

[60] For a review of Ohlin on increasing returns, see Paul Krugman, "*Was it all in Ohlin?*" (Krugman, 1999).

Consumers have some elasticity of substitution between these varieties.[61]

The production of every variety has increasing returns, and is therefore produced by a monopolist in equilibrium. The monopolist sets its price in order to maximize profits, but needs to take into account that increasing its price will lower its sales, as consumers substitute away to other varieties.

A crucial simplification here is the assumption that there is a large number of varieties, but each individual monopolist produces only a small number of them (in most models, only one). This implies that the effect of the firm's decisions on aggregate variables (such as the industry's overall price level, or its total sales) is negligible, and we can disregard it. Therefore, there is no strategic interaction between firms in the model, and no need to bring in tools from game theory to solve it.

2.2 Increasing returns and gains from trade: the Krugman model

Using the basic monopolistic competition framework outlined above, Krugman (1979) proposed a model showing how identical countries can gain from trade.

2.2.1 Assumptions

There are two identical countries, populated by identical consumers which inelastically supply L units of labour and consume a mass M of differentiated varieties.[62] Their preferences are given by

$$U_j = \int_0^M v\left(c_j\left(i\right)\right) di, \tag{82}$$

where v is an increasing and concave function. Furthermore, the function $c \mapsto \frac{v'(c)}{v''(c)c}$ is assumed to be (weakly) increasing. The Krugman model is the only one in these notes which does not feature CES preferences. Indeed, as we will see later, CES preferences shut down one of the model's main effects. Nevertheless, the preferences chosen by Krugman share the "love of variety" features of CES preferences with an elasticity ε larger than 1: as the marginal utility of every variety is decreasing, the consumer would ideally like to split up her consumption across as many varieties as possible.

Importantly, the mass of varieties M is not fixed, but will be determined in equilibrium. Precisely, it will be pinned down by a free entry condition: Krugman assumes that it is costless to invent a new

[61] We will analyse a CES-type model (similar to the one in Chapter I), where consumers end up consuming a little bit of every available variety. It has been shown that this "love of variety"-model is essentially equivalent to the (presumably more realistic) "ideal variety"-model developed by Lancaster (1979), in which consumers are different, and every consumer wants to consume just one variety that is "ideal" for her. However, calculations are easier with the love of variety-model.

[62] Krugman assumes there is only one industry, but it is easy to extend the model to several ones.

variety and start producing it. Therefore, as long as incumbent monopolists make a profit, new firms enter the market and create new varieties. That process only stops when the profits of all firms have been driven to zero.

The producer of a given variety i has the production function

$$y_j(i) = \max(0, z(l_j(i) - f)).$$

Production has increasing returns because there is a fixed cost of f units of labour which the firm has to pay to start producing (this could capture headquarter or administration costs, for example). After paying this cost, technology is linear, as in the Ricardian models we have seen: for every additional unit of labour employed, the firm produces z units of its variety.

Increasing returns can be seen most clearly when looking at the firm's costs. The total cost of the firm for producing $y_j(i)$ units is $\left(f + \frac{y_j(i)}{z}\right) w_j$. Therefore, the average cost per unit is $\left(\frac{f}{y_j(i)} + \frac{1}{z}\right) w_j$. The average cost decreases in the scale of production $y_j(i)$, as the firm spreads out the fixed cost over more and more units. The marginal cost, however, is constant at $\frac{w_j}{z}$.

2.2.2 The autarky equilibrium

To simplify notation, I omit everywhere the country index j. As the two countries considered are identical, their autarky equilibria are obviously identical, too.

Consumer choices Every consumer solves

$$\begin{cases} \max_{\{c(i)\}_{i\in[0,M]}} & U = \int_0^M v(c(i))\, di \\ s.t. & \int_0^M p(i)\, c(i)\, di = w \end{cases}$$

As in Chapter I, this problem can be solved by the "trick" of deriving under the integral for a particular variety n.[63] This yields the first-order condition

$$v'(c(n)) = \lambda p(n), \tag{83}$$

where λ is the Lagrange multiplier on the consumer's budget constraint. The Lagrange multiplier

[63]Unlike in Chapter I, however, we cannot a priori assume that there is a representative consumer. Indeed, without making further assumptions on the function v, it is not guaranteed that demands will not depend on the income level. However, given that all varieties are exactly symmetrical, the consumer will actually end up spending the same percentage of its income (independently of its income level) on all of them.

is implicitly defined by replacing Equation (83) into the consumer's budget constraint, which gives $\int_0^M p(i) v'^{-1}(\lambda p(i)) di = w$. However, we will not need it in what follows.

Firms The monopolist selling variety i solves

$$\max_{p(i)} p(i) Lc(i) - \left(f + \frac{Lc(i)}{z}\right) w.$$

This expression uses the fact the monopolist knows its demand function, and therefore that the quantity $y(i)$ that he sells must equal $Lc(i)$, the demand for his product by all consumers. Furthermore, the monopolist knows how this demand varies with the price it charges. The first-order condition for its problem is

$$c(i) + p(i) \frac{\partial c(i)}{\partial p(i)} - \frac{w}{z} \frac{\partial c(i)}{\partial p(i)} = 0,$$

Multiplying this expression by $\frac{p(i)}{c(i)}$ makes the price-elasticity of demand, $\frac{\partial c(i)}{\partial p(i)} \frac{p(i)}{c(i)}$, appear. Denoting the absolute value of the price-elasticity of demand by $\varepsilon_i(c(i))$ (so that $\varepsilon_i(c(i)) = -\frac{\partial c(i)}{\partial p(i)} \frac{p(i)}{c(i)}$), we get

$$p(i) = \left(\frac{\varepsilon_i(c(i))}{\varepsilon_i(c(i)) - 1}\right) \frac{w}{z}. \tag{84}$$

Equation (84) indicates a very general result: a monopolist sets its price as a multiple of its marginal cost. This multiple is called the mark-up. Several things should be noted:

- Equation (84) is only well defined when the price-elasticity ε_i is larger than 1. Indeed, otherwise, quantities would respond so little to prices that the monopolist could make an infinite profit by charging an infinitely high price.

- The mark-up is always higher than 1, so the monopolist charges a price which is higher than its marginal cost.

- The mark-up is decreasing in the price-elasticity. When the price-elasticity is close to 1, quantities do not respond much to prices, and the monopolist can charge a very high price. When the price-elasticity gets very large, on the other hand, the mark-up falls, as consumers reduce demand a lot after price increases. In the extreme where the price-elasticity tends to infinity, the mark-up tends to 1, and we recover perfect competition, with prices equal to marginal costs.

- With CES preferences, the price-elasticity would be constant and equal to the parameter ε (see

Chapter I). Thus, the mark-up would be constant, too. Therefore, CES preferences simplify calculations, but as we will see later, they also eliminate one of the model's results.

Finally, using consumers' first-order condition, we can derive an expression of the price-elasticity of demand just as a function of quantities and of the utility function v. Deriving Equation (83) with respect to $p(i)$ gives

$$v''(c(i))\frac{\partial c(i)}{\partial p(i)} = \lambda$$

and therefore,

$$\varepsilon_i(c(i)) = -\frac{\partial c(i)}{\partial p(i)}\frac{p(i)}{c(i)} = -\frac{v'(c(i))}{v''(c(i))\,c(i)}. \tag{85}$$

Thus, our initial assumption on the shape of v implies that the (absolute value of the) price-elasticity of demand is (weakly) decreasing in the quantity consumed. Intuitively, the higher the quantity a consumer consumes, the less he responds to prices, which seems to be a reasonable assumption.

Solving the model As the utility function is the same for every variety, and marginal costs are also identical, all firms choose the same prices and produce the same varieties in equilibrium: $p(i) = p$ and $y(i) = y$. Therefore, Equation (84) can be rewritten

$$\frac{pz}{w} = \frac{\varepsilon(c)}{\varepsilon(c) - 1}. \tag{86}$$

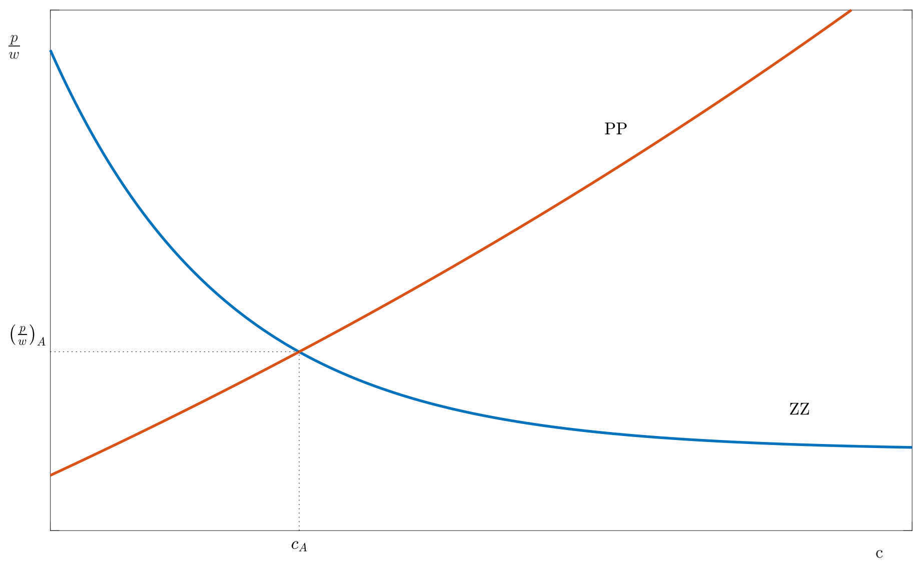
This equation defines a relationship between the ratio of prices to wages, $\frac{p}{w}$, and the quantity of each variety consumed by an individual consumer, c. It is plotted as the PP curve in Figure 16. As the price-elasticity is decreasing in quantity, the mark-up is increasing in quantity, and the PP curve is upward-sloping. Importantly, note that w/p is not equal to the real wage (which would be the wage divided by some ideal price index we have not yet determined). Indeed, it would be misleading to interpret w/p as a sufficient statistic for welfare: it captures only how much of each variety the consumer can buy with his wage, but for welfare, we also need to know the number of varieties.

Next, we can remark that in equilibrium, free entry drives the profits of every monopolist to 0. Profits are given by $py - w\left(f + \frac{y}{z}\right)$, and as $y = Lc$, the requirement that they are 0 writes

$$\frac{p}{w} = \frac{f}{Lc} + \frac{1}{z}. \tag{87}$$

This gives a second relationship between $\frac{p}{w}$ and c, marked by the ZZ curve in Figure 16. The ZZ curve is downward-sloping, and its intersection with the PP curve gives us the values of $\frac{p}{w}$ and c in the autarky equilibrium.

Figure 16: Autarky equilibrium in the Krugman model



The last thing left to do is to determine the mass of varieties, M. For this, we use the labour market clearing condition. The amount of labour used for the production of one variety is $l = f + \frac{Lc}{z}$. Therefore, labour market clearing implies

$$M\left(f + \frac{Lc}{z}\right) = L \;\Rightarrow\; M = \frac{L}{f + \frac{Lc}{z}}. \tag{88}$$

Note, in particular, that the mass of varieties is increasing in the country's total population L. In Problem Set 4, you will study in greater detail the effects of population growth in this model.

2.2.3 Trade and gains from trade

Now, assume both countries can trade varieties without trade costs.

Two approaches to variety At this point, we need to choose between two modelling approaches. A first approach would be to assume that the varieties which the two countries produced in autarky were identical. For example, in each country, there was one producer of the variety "electric sports cars". Then, opening up to trade creates some head-to-head competition between the two electronic sports cars producers, and as there are increasing returns, one of them must necessarily exit the market, and only the other one remains, serving now both markets.[64]

An alternative approach is to make an Armington assumption[65] and assume that varieties produced in different countries are different. In our previous example, it may be that Italy produces Ferrari electric sports cars, and Germany Mercedes electric sports cars, and the differentiation by brand name makes consumers perceive these as different goods. In this case, there is no head-to-head competition after opening up. However, as we will see, some producers will still be forced to exit, through the more indirect effects of increased competition from other, foreign varieties.

These two approaches are completely equivalent in the context of the Krugman model. For its somewhat greater simplicity, we take the Armington approach.

The trade equilibrium As both countries are exactly identical, all their aggregate variables coincide in equilibrium. I therefore continue not to use country subscripts to ease notation, and focus on one country, called Home.

Trade allows consumers to consume foreign varieties on top of domestic ones. There is no need to solve for any specialization in this model: each variety produced is also exported, as its producer has a monopoly on both markets. We continue to write M for the mass of varieties consumed by a Home consumer. While in the autarky equilibrium, this was also equal to the mass of varieties produced in Home, this is not true any more in the trade equilibrium. Instead, only half of these varieties are produced at Home, the other half are imported from the other country.

The possibility to buy foreign varieties however does not affect Home consumer's first-order condition, which is still given by (83). Furthermore, the pricing rule for monopolists is also unchanged: monopolists' sales are now given by $2Lc$ rather than by Lc, but you can easily check that this does not affect Equation (84).[66] Thus, trade does not affect the PP curve, as shown in Figure 17.

However, the monopolist's profits are now $p\,(2Lc) - w\left(f + \frac{2Lc}{z}\right)$, so the free entry condition becomes

$$\frac{p}{w} = \frac{f}{2Lc} + \frac{1}{z}. \tag{89}$$

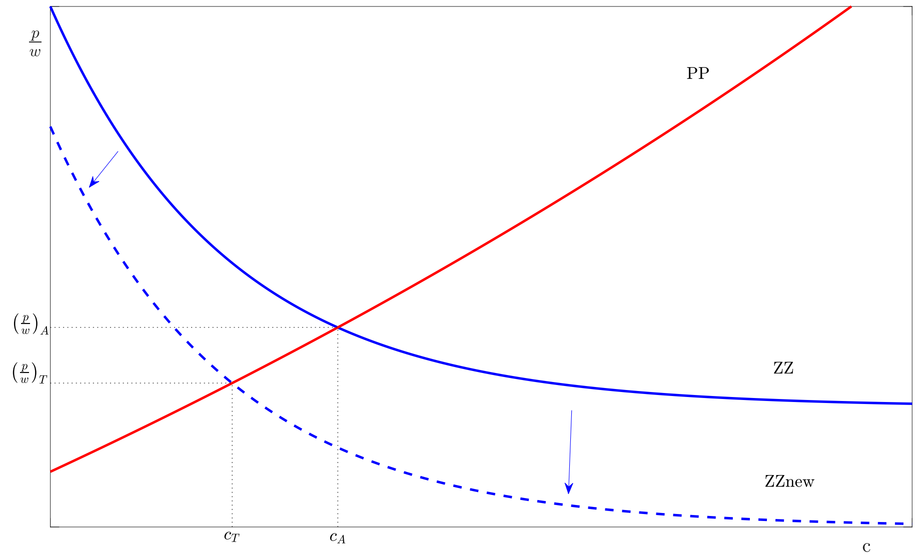
Therefore, trade makes the ZZ line shift downwards, as shown in Figure 17.

[64] As both producers are exactly identical, there would be no way of predicting which of them exits and which one remains. This would be determined by luck, or some kind of (unmodelled) first mover advantage.

[65] See Problem Sets 1 and 2 for two examples of Armington models.

[66] The reason for why this does not affect the pricing rule is that just multiplying the demand function by 2 does not change the price elasticity of demand, which is the only relevant thing for the monopolist's optimal price choice.

Figure 17: The effect of free trade



As a consequence, in both countries, trade raises the ratio w/p (which implies that firms' mark-ups decrease), and lowers an individual consumer's consumption of any given variety. This may seem strange at first sight: the consumer can buy more of any variety i with her wage, yet her consumption of variety i goes down. How can this be? The explanation is that the total number of varieties available has changed, too.

The mass of varieties can be determined as before, by using (Home) labour market clearing:

$$\frac{M}{2}\left(f + \frac{2Lc}{z}\right) = L \;\Rightarrow\; M = \frac{2L}{f + \frac{2Lc}{z}}. \tag{90}$$

This shows that the mass of varieties available to Home consumers is higher in the trade equilibrium than in autarky. Summarizing the effects from trade in the Krugman model,[67]

$$\frac{1}{2}M^{\text{Trade}} < M^{\text{Autarky}} < M^{\text{Trade}} \quad \text{and} \quad (w/p)^{\text{Autarky}} < (w/p)^{\text{Trade}} \quad \text{and} \quad c^{\text{Autarky}} > c^{\text{Trade}}.$$

Gains from trade Usually, we have proven the existence of gains from trade by calculating a country's real income (using an appropriate ideal price index). Doing this here would be a bit involved, as we have not specified the function v, but we actually can show the existence of gains from trade in

[67] The first inequality can be shown by proving $2c^{\text{Trade}} \geq c^{\text{Autarky}}$.

a simpler way. Indeed, a worker of the home country is affected by trade in two ways.

1. For any given variety, the number of units the worker can buy with her wage has increased (because $\frac{w}{p}$ has increased).

2. The number of varieties to choose from has increased.

Both of these developments expand the workers budget set, i.e., increase the set of feasible consumption choices. Therefore, the utility of the worker cannot go down, and if she chooses a different consumption allocation, she must be better off. Here, she chooses a different allocation (lowering her amount of every variety consumed in order to consume a greater mass of varieties) because her preferences (just like the CES preferences of Chapter I) feature love for variety.

How do increasing returns generate gains from trade? In the trade equilibrium, every firm produces for two markets instead of one, which (because of increasing returns) lowers its average production cost, and therefore increases its profits, all else equal. However, there is also more intense competition, because each firm has a greater number of competitors, and this forces firms to lower prices. Harsher competition puts some firms out of business, so the equilibrium mass of varieties produced by each country in the trade equilibrium is lower than in the autarky equilibrium. These firms, however, charge lower mark-ups. Moreover, the overall number of firms is higher than in autarky, because of increasing returns: as every firm is more efficient, the market can support a larger number of them.[68]

Summarizing, in Krugman's model, gains from trade between identical countries come from increasing returns, and manifest themselves in a greater variety of goods to choose from, and lower mark-ups for each of these goods. In the next section, we will study some additional predictions derived from this basic framework.

2.3 Extensions and applications

2.3.1 The Krugman model with CES preferences

In many subsequent applications of the model, starting with Krugman (1980), researchers have not used the general utility function given in Equation (82), but the more specific CES preferences that we studied in Chapter I. What does this change?

[68] This last point can also be seen when considering the situation from a resource point of view. Intuitively, when the countries open up to trade, increasing returns would allow them to produce the same mass of varieties than in autarky by using less labour. That, however, would leave some labour unused, which cannot be: the labour which has been freed up instead flows into the production of new varieties.

As we have seen in Chapter I, with CES preferences, the price elasticity of demand is constant, and equal to ε. Therefore, the model is well-defined if ε is larger than 1, and in that case, the equation for the PP curve is

$$\frac{p}{w} = \frac{\varepsilon}{\varepsilon - 1} \frac{1}{z}, \tag{91}$$

that is, the PP curve just becomes a horizontal line. The ZZ curve, on the other hand, is unchanged. With this new PP curve, we can now analytically solve for all relevant variables of the model. In the autarky equilibrium, this gives

$$c^{\text{Autarky}} = \frac{fz\,(\varepsilon - 1)}{L} \text{ and } M^{\text{Autarky}} = \frac{L}{f\varepsilon}. \tag{92}$$

In the trade equilibrium, instead

$$c^{\text{Trade}} = \frac{fz\,(\varepsilon - 1)}{2L} \text{ and } M^{\text{Trade}} = \frac{2L}{f\varepsilon}. \tag{93}$$

So, with CES preferences, the mass of varieties consumed in every country exactly doubles as a consequence of trade. Moreover, there are now no changes in mark-ups. Thus, one of the channels of the original Krugman model disappears: all gains from trade now come through increasing variety, and none through lower prices per variety.

Finally, with CES preferences, we can also provide an analytical proof for gains from trade. Using the expression for the ideal price index derived in Chapter I, the real wage in autarky is

$$\left(\frac{w}{P}\right)^{\text{Autarky}} = \frac{w}{\left(\frac{L}{f\varepsilon}\right)^{\frac{1}{1-\varepsilon}} \frac{\varepsilon}{\varepsilon-1}\frac{w}{z}} = \frac{\varepsilon - 1}{\varepsilon} z \left(\frac{L}{f\varepsilon}\right)^{\frac{1}{\varepsilon - 1}},$$

while the real wage in the trade equilibrium is

$$\left(\frac{w}{P}\right)^{\text{Trade}} = 2^{\frac{1}{\varepsilon-1}} \left(\frac{\varepsilon - 1}{\varepsilon} z \left(\frac{L}{f\varepsilon}\right)^{\frac{1}{\varepsilon-1}}\right).$$

This shows that there are gains from trade (as we have assumed $\varepsilon > 1$). These gains are decreasing in the elasticity ε. Indeed, when ε goes to infinity, all varieties become perfect substitutes, and an increase in the mass of varieties does not make workers better off.

2.3.2 Asymmetric countries and the home market effect

Krugman (1980) used this CES model to study an equilibrium with asymmetric countries and transport costs. In this work, he showed two important results.

First, the larger country produces and exports a greater number of varieties, as the size of its domestic labour force allows it to benefit more from increasing returns to scale, and therefore to acquire cost advantages in a greater number of industries. For the same reasons, the larger country also has higher wages. This effect on wages crucially depends on transport costs. Indeed, without transport costs, the small country can fully benefit from the increasing returns generated by the large country. With transport costs, however, the consumers in the larger country, who can get a larger number of varieties from firms located close by, are better off (see Exercise 1 of Problem Set 4 for details). This point is a strong difference between models of increasing returns and Ricardian ones. In a Ricardian model, a large country must produce a lot, which deteriorates its terms of trade. In the Krugman model, larger production helps exploiting increasing returns and lowering costs.

Second, Krugman shows that his model can deliver a "home market effect", as first described by Linder (1961). Linder's idea was that if consumers in a country have a strong preference for a particular product, the country is likely to become an exporter of this product. In a Ricardian or Heckscher-Ohlin framework, this idea is counterfactual: high domestic demand for a given product makes it likely that a country is an importer of that product, not an exporter. Indeed, productivity and domestic demand are uncorrelated in these models, and there is no mechanism through which high domestic demand for, say, tomatoes can increase the productivity of tomato-producing firms and give the country a comparative advantage for tomatoes. Krugman shows that in a model with increasing returns, things are different. He considers a model with two industries (where every industry has a mass of differentiated varieties and increasing returns), and assumes that consumers in the Home country like industry A more than industry B. Then, under certain parameter conditions, the Home country specializes in the production of industry A. Again, this result crucially depends on trade costs: Home specializes precisely because it wants to provide its home consumers with the products they like free of trade costs.

Costinot, Donaldson, Kyle and Williams (2019) provide empirical evidence for such home market effects in the pharmaceutical industry. A concrete example from their paper neatly shows the point. "*Consider the drug famotidine, [..] used to treat peptic ulcers and gastro-esophageal reflux, and discovered in Japan - a country known for particularly high incidence rates of peptic ulcers. In our data, individuals in Japan are nearly twice as likely to die from digestive disorders than are individuals in the rest* *of the world (0.266 deaths per 1,000 population annually in Japan, relative to 0.170 on average in other countries). Looking at data on Japan's exports and imports, sales of Japanese drugs targeting pepticulcers and gastro-esophageal reflux diseases outside Japan account for 10.35% of world sales, compared to an average of 4.54% for all other disease categories. More strikingly, Japan is a net importer in the pharmaceutical sector as a whole, but is a net exporter of drugs targeting peptic ulcers, reflux, and related digestive diseases.*"

The home market effect has important implications, showing that comparative advantage is not exogenous, but could be manufactured. For instance, a country could protect its industries from import competition, increase domestic demand for them, and once this demand has made industries sufficiently competitive, open up again. These considerations have given rise to a large literature on "strategic trade policy" in the 1980s (see Brander (1995) for an overview).

2.3.3 Empirical applications

Scale, selection, and gains from variety Even before New Trade Theory, policy makers often argued that trade liberalization would allow firms to operate on a larger scale and make them more efficient through increasing returns. Krugman's 1979 model indeed predicts that firms should become larger after trade liberalization. However, this effect has not been borne out empirically. For instance, Head and Ries (1999) for Canada and Tybout and Westbrook (1995) for Mexico show that the average scale of firms did not increase much after trade liberalizations. This does not necessarily mean that increasing returns are not important: in the CES version of the Krugman model, firms actually do not increase their scale after a trade liberalization, as greater competition from foreign varieties exactly cancels the effect of larger market size.

The results of these empirical investigations lead trade economists to focus on two other aspects. First, some tried to quantify the gains from trade due to expanding variety. The most famous study here is due to Broda and Weinstein (2006), who estimated that the increasing number of imported varieties in the United States increased consumer welfare by 1.2% between 1972 and 2001.[69] More recently, Feenstra and Weinstein (2017) examined the period 1992-2005, and found that increasing variety and decreasing markups (another effect emphasized by the Krugman model) increased welfare by around 1%, with decreasing markups accounting for about half of the total effect.

[69] Such computations obviously rely on strong assumptions. First, the authors need to define what a "variety" is in the real world, and then, they need to estimate the elasticity of substitution between varieties. Both of these tasks are far from trivial.

Second, other studies focused on the prediction that some firms are forced out of the market after trade liberalizations. However, these firms did not seem to be randomly selected, but rather were the previously least performant firms on their national markets. This simple observation started a vast literature on firm heterogeneity, which we study in the next chapter.

Gravity As shown by Helpman (1987), the Krugman model also delivers a gravity equation, similar to the ones we have encountered already (see Chapter II and Problem Set 2). Furthermore, the model also yields an ACR-type formula for the gains from trade. We will study these issues in greater detail in the next chapter, where we extend the Krugman model to allow for firm heterogeneity.

2.3.4 Taking stock: different reasons for trade

New Trade Theory was designed to explain North-North, intra-industry trade. But of course, it does not imply that all trade must be of this type. Therefore, it seems quite natural to combine the new theory with the classical (Ricardian and Heckscher-Ohlin) ones in a general model, having both intra-industry trade due to increasing returns, and inter-industry trade due to differences in productivity and factor endowments. The first such "synthesis" model is due to Helpman and Krugman (1985), and there have been many subsequent ones, including Romalis (2004) and Ventura (2005).

From the three reasons for trade we have studied, which one is the most important? There is no general answer to this question, as their relative importance depends on the countries and the historical period considered. With the greater openness of emerging countries to trade, it seems that during the last two decades, the "classical" motives for trade have regained greater importance. As Krugman puts it in his Nobel Prize Lecture, *"there's no reason the world has to keep moving in a direction that makes new theories more relevant. [...] The rise of the new trade theory was motivated to a large extent by the rising relative importance of similar-similar trade: two-way exchanges of goods among advanced economies. For the last two decades, however, the trend has been in the other direction, with rapidly rising trade between advanced economies and much poorer, lower-wage economies, especially China. [...] Nobody doubts that trade between the United States and Mexico, where wages are only 13 percent of the U.S. level, or China, where they are only about 4 percent, reflects comparative advantage rather than arbitrary, scale-based specialization. The old trade theory has regained relevance."*

However, New Trade Theory has not only managed to explain successfully North-North intra-industry trade, but also paved the way for the major development, the introduction of firm heterogeneity. This is the object of the next chapter.

Part V

Heterogeneous Firms

1 Introduction

In the Ricardian and Heckscher-Ohlin models, there is no big role for firms. Of course, they produce all goods, but as there is perfect competition, their number and their size is indeterminate and irrelevant: one could as well assume that there is just one representative firm producing all the output of an industry. In the Krugman model, there are firms, but they are all identical.

However, when researchers started working in the 1980s and 1990s on large firm-level datasets, they found that in the real world, even firms belonging to the same industry are very different. First, there are huge differences in firm size (as measured by sales or employment). Second, there are huge differences in economic performance: some firms need considerably less resources than others to achieve the same amount of sales, which has been interpreted as indicating that they are more productive, or produce more appealing products. Finally, only a small minority of all firms in a country export, and these firms are in general larger and more productive than the average firm. This suggests that there are some hurdles to exporting, creating a selection process which only the "best" firms are able to pass.

In 2003, Marc Melitz managed to combine all of these insights in a highly successful model (Melitz, 2003). He started from the standard Krugman (1979) framework, and extended it by introducing differences in productivity between firms and fixed exporting costs. His model then emphasizes that trade liberalizations change the allocation of resources between domestic firms: the most productive ones manage to expand to the foreign market, while the least productive ones contract or exit. Thus, as resources are reallocated to the most productive firms, the overall productivity of every trading industry increases.

Melitz's model does not introduce a fundamental new motive for trade: trade still occurs because of increasing returns to scale, just as in the Krugman model. However, it highlights a new mechanism with which an economy reacts to trade, showing that if we would assume (as in the Krugman model) that all firms are identical, we would substantially underestimate the gains from trade. Furthermore, it allows us to better understand all the complex changes triggered by trade.

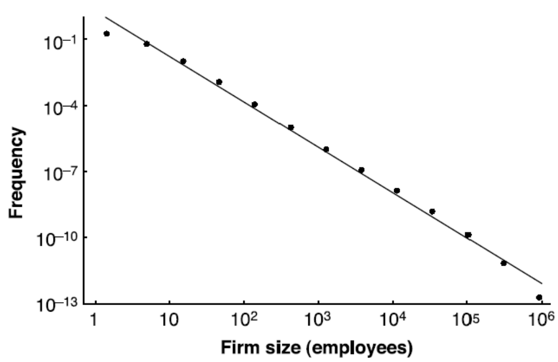
Section 2 of this chapter describes some empirical stylized facts on firm-level differences in size, productivity and exports. Section 3 shows how Melitz combined these insights in his model, and Section 4 shows the implications for gains from trade. Section 5 reviews the empirical literature on the effects of trade liberalization on productivity, and finally, Section 6 gives a short overview of the many applications of the Melitz model over the last decade.

2 Firm heterogeneity in the data

2.1 Productivity and the size distribution of firms

In most countries and most industries, firm size (which is typically measured either by the number of employees or by total sales) follows Zipf's law (Axtell, 2001; Gabaix, 2009). That is, the size of a firm is inversely proportional to its rank in the firm size distribution: the largest firm has approximately twice as many employees than the second largest, three times as much as the third largest, etc.[70] This distribution is highly skewed: the number of large firms is very small, but their size is very large. This implies that large firms have an outsized importance for aggregate economic variables.[71] Furthermore, it begs a question: why are there so large differences in firm sizes?

Figure 18: The size distribution of firms in the United States, 1997



Source: Axtell, 2001

First, it has been shown that size differences are not just due to industry differences, that is, to the

[70]Zipf's law is a fascinating statistical regularity. It describes the distribution of firm sizes, but also city sizes or the frequency of words in the English language (see Gabaix, 2009 and the Wikipedia article "Zipf's law"). Zipf's law is a power law, of the same family as the Fréchet distribution we encountered in the Eaton-Kortum model.

[71]Gabaix (2011) argues that some firms are so large that their idiosyncratic shocks (the Volkswagen pollution scandal, a strike at General Motors...) have a sizeable impact on a whole economy. For instance, the decline of Nokia has been seen as a reason behind the economic problems of Finland in recent years.

fact that a supermarket chain typically has more employees than a software firm: the size distribution is skewed even within the supermarket or the software industry. Instead, productivity differences could be an explanation. For instance, Syverson (2011) reports that, for a study of 218 different US manufacturing industries, "*the average difference in logged total factor productivity (TFP) between an industry's 90th and 10th percentile plants is 0.651. This corresponds to a TFP ratio of* $e^{0,651} = 1,92$. *To emphasize just what this number implies, it says that the plant at the 90th percentile of the productivity distribution makes almost twice as much output with the same measured inputs as the 10th percentile plant.*" In developing countries such as India and China, these magnitudes are typically even bigger (see Hsieh and Klenow, 2009).

Productivity differences provide a relatively easy explanation for size differences, as more productive firms can hire and sell more. However, they also pose several issues of measurement and interpretation. Strictly speaking, a firm's productivity is the ratio between the physical output it produces and the resources (capital and labour) that it uses for production. However, real world data, almost never has information on physical quantities, but instead only information about sales (prices multiplied by quantity). Therefore, standard productivity measures really capture how much sales a firm can produce with its resources. These differences in "revenue productivity" may come from differences in physical productivity (a capacity to produce more quantities from the same inputs), differences in product appeal (a capacity to produce a better product with the same inputs) or differences in mark-ups (which may come from differences in market power, or differences in the quality of the goods produced). While physical productivity and product appeal are largely isomorphic in most economic models, mark-ups have very different implications, and it is therefore important to identify their impact on commonly used productivity measures. In Section 5.2, we will return to this issue.

2.2 The special status of exporters

The majority of firms do not participate in international trade. For instance, Bernard, Jensen, Redding and Schott (2018) show that in 2007, only 37% of US manufacturing firms were exporters. In service sectors, the fraction of exporting firms is certainly even lower. Furthermore, as Table 7 shows, there is a lot of heterogeneity between industries: in some subsectors of manufacturing, many firms export, in others, very few.

Exporting firms are not only few, they are also special. Table 8 illustrates these differences by running the simple regression

$$x_i = \alpha + \beta \mathrm{E}_i + \varepsilon,$$

where x_i is some outcome variable for firm i (employment, sales, etc.) and E_i a dummy variable equal to 1 if firm i exports, and to 0 otherwise. As the table shows, the estimate for β is positive for all outcome variables considered, showing exporters have on average higher employment, higher sales, higher TFP, higher wages, etc. Furthermore, this is robust to controlling for industry fixed effects.

Table 7: Export frequency in the United States, 2007

TABLE 1
FIRM EXPORTING

		Percent of firms (1)	Fraction of firms that export (2)	Mean exports as a share of total shipments (3)
311	Food manufacturing	6.8	0.23	0.21
312	Beverage and tobacco product	0.9	0.30	0.30
313	Textile mills	0.8	0.57	0.39
314	Textile product mills	2.7	0.19	0.12
315	Apparel manufacturing	3.6	0.22	0.16
316	Leather and allied product	0.3	0.56	0.19
321	Wood product manufacturing	4.8	0.21	0.09
322	Paper manufacturing	1.5	0.48	0.06
323	Printing and related support	11.1	0.15	0.10
324	Petroleum and coal products	0.5	0.34	0.13
325	Chemical manufacturing	3.3	0.65	0.23
326	Plastics and rubber products	3.9	0.59	0.11
327	Nonmetallic mineral product	4.3	0.19	0.09
331	Primary metal manufacturing	1.5	0.58	0.31
332	Fabricated metal product	20.6	0.30	0.09
333	Machinery manufacturing	8.7	0.61	0.15
334	Computer and electronic product	3.9	0.75	0.28
335	Electrical equipment, appliance	1.7	0.70	0.47
336	Transportation equipment	3.4	0.57	0.16
337	Furniture and related product	6.5	0.16	0.14
339	Miscellaneous manufacturing	9.3	0.32	0.16
Aggregate manufacturing		100.0	0.35	0.17

Notes: Data are for 2007 and are for firms that appear in both the US Census of Manufactures and the LFTTD. Column 1 summarizes the distribution of manufacturing firms across three-digit NAICS manufacturing industries. Column 2 reports the share of firms in each industry that export. Firm exports are measured using customs information from LFTTD. Column 3 reports mean exports as a percent of total shipments across all firms that export in the noted industry. Percentages in column 1 need not sum exactly to 100 due to rounding.

Source: Bernard, Jensen, Redding and Schott (2018).

These "exporter premia" raise an important question: are they due to selection or causation? That is, do we observe exporter premia because only the largest and most productive firms export, or do we observe them because exporting makes firms larger and more productive?

In reality, exporter premia are probably a mix of both things. However, the general consensus in international trade has long been that selection effects are more important. This view was based on the work of Clerides, Lach and Tybout (1998) and Bernard and Jensen (1999), who showed that productivity and wages are higher for exporting firms even **before** they actually start to export.

Selection is also a fundamental feature of the Melitz model, in which firms' productivity is fixed and not affected by exporting. In recent years, however, researchers have again paid greater attention to ways in which exporting can improve firm performance, through learning or by stimulating more R&D.

Table 8: Exporters are different

TABLE 2
EXPORTER PREMIA

	Exporter premia		
	(1)	(2)	(3)
log employment	1.28	1.11	—
log shipments	1.72	1.35	0.24
log value added per worker	0.33	0.19	0.21
log TFP	0.03	0.04	0.04
log wage	0.21	0.09	0.10
log capital per worker	0.28	0.16	0.20
log skill per worker	0.06	0.01	0.11
Additional covariates	None	Industry fixed effects	Industry fixed effects, log employment

Notes: Data are for 2007 and are for firms that appear in both the US Census of Manufactures and the LFTTD. All results are from bivariate ordinary least squares regressions of firm characteristic in first column on a dummy variable indicating firm's export status. Firm exports measured using customs information from LFTTD. Columns two and three include industry fixed effects and industry fixed effects plus log firm employment, respectively, as additional controls. Total factor productivity (TFP) is computed as in Caves et al. (1982). Capital and skill per worker are capital stock and non-production workers per total employment, respectively. All results are significant at the 1 percent level except the log Skill per Worker results in column 2 which are not significant at the 10 percent level.

Source: Bernard, Jensen, Redding and Schott (2018).

3 The Melitz model

3.1 Assumptions

The Melitz model builds on the standard Krugman (1980) model with CES preferences. We consider two identical countries with labour supply L. However, Melitz introduces three major novelties: heterogeneous productivities, fixed and variable export costs, and entry costs.

As in the Krugman model, there is monopolistic competition, and each variety is produced by exactly one firm. Production uses labour, and the firm producing variety i in country j has the production function

$$y_j(i) = \max(0, z(i)(l(i) - f)). \tag{94}$$

The only difference between this production function and the Krugman one is that productivity $z(i)$ is now firm-specific: firms with high values of $z(i)$ have high productivity (and therefore low marginal costs), while firms with low values of $z(i)$ have low productivity (and high marginal costs).

Firms, and therefore varieties, can be created by entrants who pay a fixed cost of f_E units of labour (while in the Krugman model, $f_E = 0$). The productivity of newly created firms is determined by a lottery: every firm draws its productivity from a cumulative distribution function G, with $G(0) = 0$. Firms are assumed to be risk-neutral, and therefore, they pay the entry cost if their expected profit is positive. After learning the outcome of their draw, they must decide whether they want to produce with this productivity, or exit right away. Indeed, for firms with a very low draw for $z(i)$, production may not be worthwhile, as their low revenues may not cover the fixed cost f.

Finally, when a firm decides to produce, it needs to decide whether it wants to export or not. To export, the firm needs to pay a fixed cost of f_X units of labour. In the words of Melitz and Redding (2014), this fixed cost "*captures market access costs (e.g. advertising, distribution, and conforming to foreign regulations) that do not vary with exporter scale*". Unproductive firms may not want to pay this fixed cost, as their foreign sales are insufficient to cover it. There are also iceberg trade costs, as in the Eaton-Kortum model: to deliver one unit of its variety to the other country, a domestic firm needs to produce and ship $\tau \geq 1$ units, as part of the variety "melts" in transit.

Thus, summing up, a firm's life cycle in the Melitz model consists of five steps:

1. The firm decides whether to enter the industry (and pay the entry cost).
2. The firm learns its productivity $z(i)$.
3. The firm decides whether to produce or to exit.
4. If it decides to produce, it decides between selling domestically, only exporting, or selling domestically and exporting.
5. Finally, the firm decides on the price of its variety, at home and (potentially) abroad.

Now, we can solve for the model's autarky equilibrium. As usual, we drop all country subscripts.

3.2 The autarky equilibrium

Consumer demand is still given by Equation (16) of Chapter I. Thus, we can focus all our attention on firms.

Prices and profits To study firms' behaviour, we proceed by backward induction. That is, we start from the end of the firm's life cycle and work our way to the beginning. So, consider a firm which has learned its productivity, decided to produce (for the domestic market, as exporting is not possible in autarky), and needs to choose its optimal price. The problem of this firm is exactly identical to the one we encountered in Chapter IV, and therefore, firms use the same pricing rule: a firm with productivity $z(i)$ chooses a price

$$p(i) = \frac{\varepsilon}{\varepsilon - 1} \frac{w}{z(i)}. \tag{95}$$

Note that more productive firms charge lower prices, and must therefore sell greater quantities. Thus, there is a direct, one-to-one mapping between firm productivity and firm size as measured by (physical) quantity sold. As we will see later, firm sales and employment are also increasing in productivity.

From this price choice, we can deduce the firm's profit. Profits are equal to sales ($p(i)\,c(i)$) minus costs ($w\left(f + \frac{c(i)}{z(i)}\right)$). Using Equations (16) and (95), we therefore get

$$\pi(i) = z(i)^{\varepsilon-1} B - wf, \tag{96}$$

where $\pi(i)$ stands for the profit of the producer of variety i, and $B \equiv \frac{1}{\varepsilon-1}\left(\frac{\varepsilon}{\varepsilon-1}\right)^{-\varepsilon} P^{\varepsilon} w^{1-\varepsilon} U$. B is an auxiliary variable which depends on wages, the aggregate price level, and aggregate utility. Every individual firm takes B as given, as it is too small to influence aggregate values.

Equation (96) shows that a firm's profit is strictly increasing in its productivity. This will be crucial for all that follows.

Production or Exit? Now that we know the profit of a producing firm, we can take one step back in the firm's life cycle. Consider a firm which has learned its productivity $z(i)$, and needs to decide whether to produce or not. If the firm does not produce, it does not earn any profit. If it produces, it will earn a profit given by Equation (96). Therefore, a firm will produce if and only

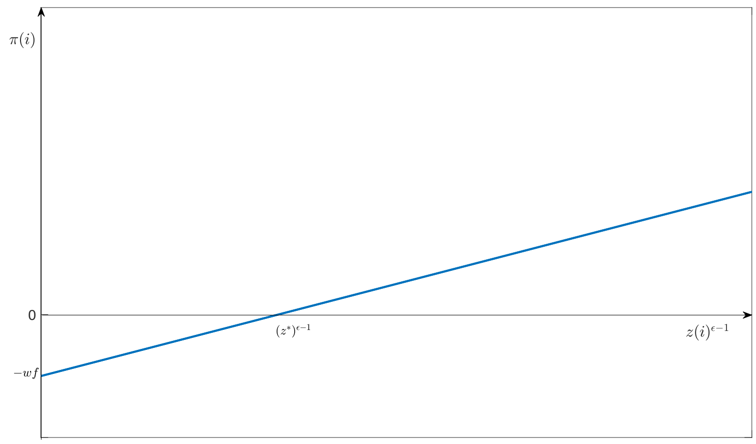
$$z(i)^{\varepsilon-1} B \geq wf. \tag{97}$$

Thus, there is some selection. Firms which are very unproductive (get a $z(i)$ draw close to 0) do not produce, while firms with high productivity do. More precisely, as the left-hand side of Equation (97) is strictly increasing, there is a unique cut-off productivity level z^* such that firms decide to produce if and only if $z(i) \geq z^*$. This cut-off productivity is defined by

$$z^{*\varepsilon-1} B = wf. \tag{98}$$

Figure 19 illustrates this situation. We can note several important things here. First, the Melitz model

Figure 19: Autarky production cut-off in the Melitz model



predicts that only the most productive firms will produce, while the most unproductive exit the market quickly. This seems to be the case in the real world, where several studies have shown that exiting firms have lower productivity than average (Bartelsman and Doms, 2000).

Second, we can deduce from the cut-off z^* the productivity distribution of producing firms. What is the probability that a producing firm has a productivity lower or equal to some real number x? Clearly, it is 0 if $x < z^*$ (there are no firms with productivity below z^* in the market). If $x \geq z^*$, then this is equal to $\mathbb{P}(z \leq x \mid z \geq z^*) = \frac{\mathbb{P}(z^* \leq z \leq x)}{\mathbb{P}(z \geq z^*)}$. Therefore, the cumulative distribution function for the productivity of existing firms, $\widetilde{G}$, is defined as

$$\forall x > 0, \quad \widetilde{G}(x) = \begin{cases} 0 & \text{if } x < z^* \\ \frac{G(x) - G(z^*)}{1 - G(z^*)} & \text{if } x \geq z^* \end{cases}. \tag{99}$$

The density of this cumulative distribution function is the function $x \mapsto \frac{g(x)}{1-G(z^*)}$. Intuitively, the productivity distribution of producing firms is just the initial productivity distribution, truncated at the cut-off level z^*. This productivity distribution is, as we will see, one-to-one related to the size distribution of firms in the Melitz model.

Entry Finally, we are ready to go to the first step of the firms' life cycle, the entry decision. As firms are risk-neutral, they pay the entry cost wf_E if and only if the expected profits from entering are higher or equal than this cost. What are the expected profits from entering? With probability $G(z^*)$, the firm draws a productivity smaller than the cut-off level, immediately exits and earns 0 profits. With probability $1 - G(z^*)$, the firm draws a productivity higher than the cut-off level, distributed according to the distribution defined just above, and earns a positive profit defined by Equation (96). Therefore, the expected profit from entering is

$$G(z^*) \cdot 0 + (1 - G(z^*)) \int_{z^*}^{+\infty} \left(z^{\varepsilon-1}B - wf\right) \frac{g(z)}{1 - G(z^*)} dz.$$

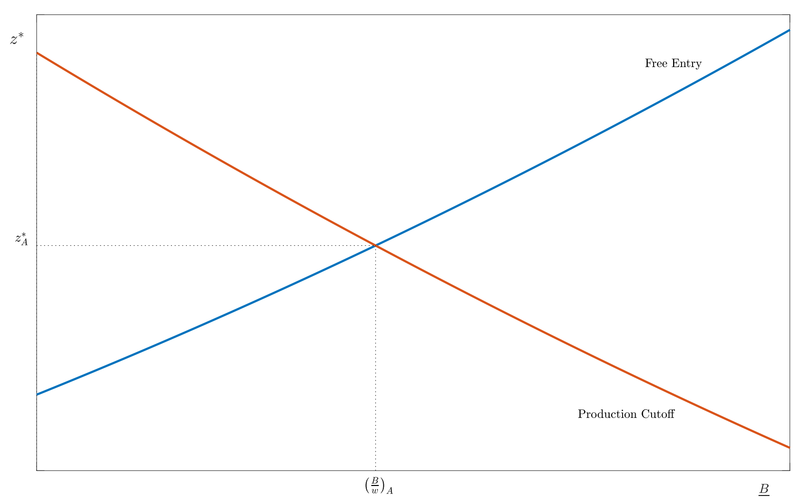
In equilibrium, this expected profit equals the entry cost (if it were higher, there would be more entry, if it were lower, then some firms would make negative profits). Therefore, we have

$$\int_{z^*}^{+\infty} \left(z^{\varepsilon-1}B - wf\right) g(z)\, dz = wf_E. \tag{100}$$

Having characterized firm's optimal behaviour, we can now solve for the autarky equilibrium. Equations (98) and (100) can be rewritten

$$\begin{cases} z^{*\varepsilon-1}\frac{B}{w} = f. \\ \int_{z^*}^{+\infty} \left(z^{\varepsilon-1}\frac{B}{w} - f\right) g(z)\, dz = f_E. \end{cases} \tag{101}$$

Figure 20: Autarky equilibrium for z^* and $\frac{B}{w}$



Thus, we now have a system of two equations and two unknowns, $\frac{B}{w}$ and z^*. The graphical solution of this system is illustrated in Figure 20.

$\frac{B}{w}$ is a summary statistic for aggregate demand and labour costs: high values of $\frac{B}{w}$ (high demand and/or low labour costs) push up profits for all firms. The first equation gives a decreasing relationship between $\frac{B}{w}$ and z^*: when the aggregate component of profits is high, the cut-off productivity level needed for a firm to start producing is low. The second equation gives an increasing relationship between $\frac{B}{w}$ and z^*: when the aggregate component of profits is is high, the entry cut-off must rise to lower entrants' expected profits and bring them back in line with the entry cost f_E.

Aggregation Knowing the equilibrium values of $\frac{B}{w}$ and z^*, we can now proceed to solve for the model's other variables. First, denote by M_E the mass of firms which pay the entry cost. By the law of large numbers, a fraction $(1 - G(z^*))$ of them end up producing, so that

$$M = (1 - G(z^*))\, M_E. \tag{102}$$

As in the Krugman model, we can solve for the equilibrium mass of varieties by using the labour market clearing condition. While labour supply is equal to L, labour demand is the sum of labour used for entry costs ($M_E f_E$ in total) and for the production of final goods. In particular, a firm producing with productivity $z(i)$ demands $f + \frac{c(i)}{z(i)}$ units of labour. Using Equations (16) and (95), we get after some algebra

$$l(i) = f + z(i)^{\varepsilon-1}(\varepsilon - 1)\frac{B}{w} = (\varepsilon - 1)\left(z(i)^{\varepsilon-1}\frac{B}{w} - f\right) + \varepsilon f$$

Note that this equation proves that more productive firms employ more workers, as we had claimed before. To calculate total labour demand for production, we can first calculate the expected labour demand for a randomly chosen producing firm. This is just the expectation of $l(i)$, which is

$$\int_{z^*}^{+\infty}\left[(\varepsilon - 1)\left(z^{\varepsilon-1}\frac{B}{w} - f\right) + \varepsilon f\right]\frac{g(z)}{1 - G(z^*)}dz = \frac{(\varepsilon - 1)\, f_E}{(1 - G(z^*))} + \varepsilon f,$$

where we have made use of Equation (100). By the law of large numbers, expected labour demand equals average labour demand, and therefore, aggregate labour demand is obtained by just multiplying that number by M, the total mass of producing firms. Thus, using Equation (102), the labour market clearing condition can finally be written as

$$L = M_E f_E + M \left(\frac{(\varepsilon - 1) f_E}{(1 - G(z^*))} + \varepsilon f \right).$$

This then gives

$$M_E = \frac{L}{\varepsilon (f_E + (1 - G(z^*)) f)} \text{ and } M = \frac{(1 - G(z^*)) L}{\varepsilon (f_E + (1 - G(z^*)) f)}. \tag{103}$$

Note that when $f = 0$ (and therefore, $z^* = 0$), this reduces to the same expression as in the Krugman model. This already hints at an important insight: heterogeneity is important in the Melitz model only because it interacts with fixed costs to create some selection. If there were only heterogeneity, but no selection, nothing fundamental would be changed.

Finally, we can solve for the real wage, our usual measure of per-capita welfare. One way to do so is to use the definition of the ideal price index P which we had introduced in Chapter I. Using firms' price-setting rule, this writes

$$P = \left(\int_0^M \left(\frac{\varepsilon}{\varepsilon - 1} \frac{w}{z(i)} \right)^{1-\varepsilon} di \right)^{\frac{1}{1-\varepsilon}}$$

Using the fact that we know the productivity distribution, and rewriting this expression, we finally get

$$\frac{w}{P} = \frac{\varepsilon - 1}{\varepsilon} M^{\frac{1}{\varepsilon - 1}} \left(\int_{z^*}^{+\infty} z^{\varepsilon - 1} \frac{g(z)}{1 - G(z^*)} dz \right)^{\frac{1}{\varepsilon - 1}}. \tag{104}$$

Thus, the real wage is increasing in the mass of goods in the market, and in (a transformation of) the average productivity of the firms in the market. Both of these variables are determined endogenously, and we know their equilibrium values because we know the equilibrium value of the cut-off z^*. We are now ready to study the effects of trade in this economy.

3.3 The trade equilibrium

Assume now that the home country is allowed to trade with a second, completely identical country. As in Chapter IV, we make the Armington assumption that every country produces its own varieties, which will simplify calculations. Note that if we would not make this assumption, the Melitz model would feature Ricardian gains from trade: the two countries could have different productivities for the same variety, and specialize by comparative advantage. Shutting down this channel makes the model more directly comparable to Krugman's (with all trade being due to increasing returns), and allows us to focus on the novel mechanisms put forward by Melitz.

As in the Krugman model, the trade equilibrium is symmetric, and therefore I do not use country subscripts.

3.3.1 Consumer and firm behaviour

Consumer demand for any variety is still given by Equation (16), no matter whether that variety is produced at home or abroad. Thus, all the action will again take place on the firm side of the model, where we can proceed by backward induction, just as in the autarky equilibrium.

Prices and profits For sales to the domestic market, the problem of the firm has not changed,[72] and therefore its price-setting rule is still given by Equation (95). However, now some firms may also decide to export their products.

On the foreign market as well, demand for a firm's variety has price elasticity ε, and the firm therefore charges a constant mark-up $\frac{\varepsilon}{\varepsilon-1}$ over its marginal cost. The marginal cost is higher than on the domestic market, because of the iceberg trade costs: to sell one unit of the variety in the foreign country, the firm needs to produce and ship τ units. Thus, the price charged by an exporting firm on the foreign market is

$$p^*(i) = \frac{\varepsilon}{\varepsilon-1}\frac{\tau w}{z(i)}. \tag{105}$$

These price choices directly yield firms' profits. A firm which produces only domestically makes a profit

$$\pi_D(i) = z(i)^{\varepsilon-1} B - wf, \tag{106}$$

where, as in autarky, $B \equiv \frac{1}{\varepsilon-1}\left(\frac{\varepsilon}{\varepsilon-1}\right)^{-\varepsilon} P^{\varepsilon} w^{1-\varepsilon} U$. A firm which decides to both sell domestically and export, on the other hand, makes a profit

$$\pi_X(i) = z(i)^{\varepsilon-1}\left(1+\tau^{1-\varepsilon}\right) B - w\left(f+f_X\right), \tag{107}$$

where I have used the fact that in the symmetric equilibrium, P and U take the same value in both countries. Note, finally, that no firm would want to only export, and not sell domestically. Indeed, for

[72] This is because even if the firm now also sells abroad, there is no interaction between what it does abroad and what it does at home: its own operations abroad neither affect the firm's marginal cost, nor the demand curve that it faces.

such a firm, starting to sell domestically would not entail any additional fixed costs and necessarily yield a profit, as it can charge a price above its variable cost.

Export, domestic production or exit? Next, consider a firm which has learned its productivity $z(i)$, and needs to decide whether to export, produce for the domestic market only, or exit altogether. The profits from the three options are plotted in Figure 21, which uses two important facts:

- For a firm with productivity 0, $\pi_X < \pi_D < 0$, as the firm can make no profits and would only need to pay the fixed costs.
- π_D and π_X are linear functions of $z(i)^{\varepsilon-1}$, with a positive slope. The slope of π_X is higher than the one of π_D.

Thus, eventually, the π_X-line will both cross the x-axis and the π_D-line. Furthermore, we have drawn the figure such that at the point where π_D crosses the x-axis, π_X is still negative. This will occur whenever f_X is sufficiently large, and we will make that assumption from now on.[73]

Figure 21: Productivity cut-offs with trade

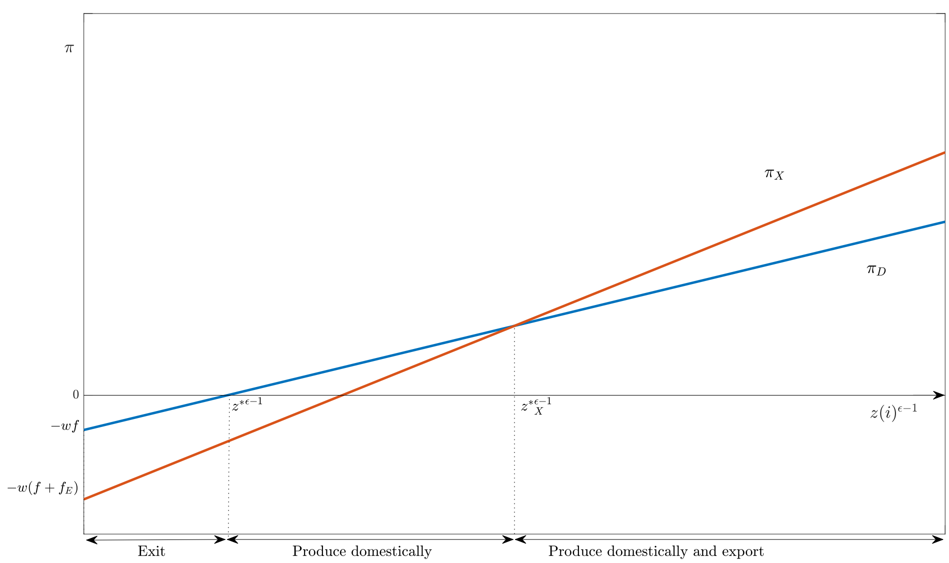


Figure 21 shows that there are now three types of firms. Firms with a productivity draw lower than z^* exit, firms with a productivity draw between z^* and z^*_X produce only for the domestic market,

[73] You can show that intersections are ordered like this iff $f_X > \tau^{1-\varepsilon} f$. Otherwise, there would be only two cases: firms either exit or sell to both markets. Then, all firms would export, which would not be in line with the data.

and firms with a productivity draw higher than z_X^* both produce for the domestic market and export. Thus, the Melitz model reproduces the empirical finding that exporting firms are more productive than non-exporting firms, and explains this by selection: only the most productive firms find it profitable to pay the fixed cost of exporting.

The cut-off productivity levels z^* and z_X^* are defined by the two equations

$$z^{*\varepsilon-1}B - wf = 0, \quad \text{and} \tag{108}$$

$$z_X{}^{*\varepsilon-1}\left(1+\tau^{1-\varepsilon}\right)B - w\left(f+f_X\right) = z_X{}^{*\varepsilon-1}B - wf. \tag{109}$$

After a little algebra, we get

$$z^{*\varepsilon-1}\frac{B}{w} = f \quad \text{and} \quad z_X{}^{*\varepsilon-1}\tau^{1-\varepsilon}\frac{B}{w} = f_X.$$

and therefore,

$$z_X{}^* = \tau\left(\frac{f_X}{f}\right)^{\frac{1}{\varepsilon-1}} z^*. \tag{110}$$

Thus, once we have the entry cut-off z^*, we know the export cut-off $z_X{}^*$. Furthermore, note that the ratio $\frac{z_X^*}{z^*}$ (which for many productivity distributions is inversely proportional to the percentage of exporting firms) is increasing in τ and f_X, but decreasing in f. This is intuitive: with higher trade costs, less firms export. Moreover, high fixed costs of production imply that only highly productive firms (with a good chance to overcome the fixed costs of exporting) are active on the market.

Entry As in autarky, in equilibrium, the entry cost wf_E must be equal to the expected profit from entering. This condition implies[74]

$$wf_E = \int_{z*}^{z_X^*}\left(z^{\varepsilon-1}B - wf\right)g\left(z\right)dz + \int_{z_X^*}^{+\infty}\left(z^{\varepsilon-1}\left(1+\tau^{1-\varepsilon}\right)B - w\left(f+f_X\right)\right)g\left(z\right)dz. \tag{111}$$

3.3.2 The equilibrium

Just as in the autarky equilibrium, the definition of the entry cut-off z^*, Equation (108), and the free-entry condition, Equation (111), form a system of two equations in two unknowns. After some

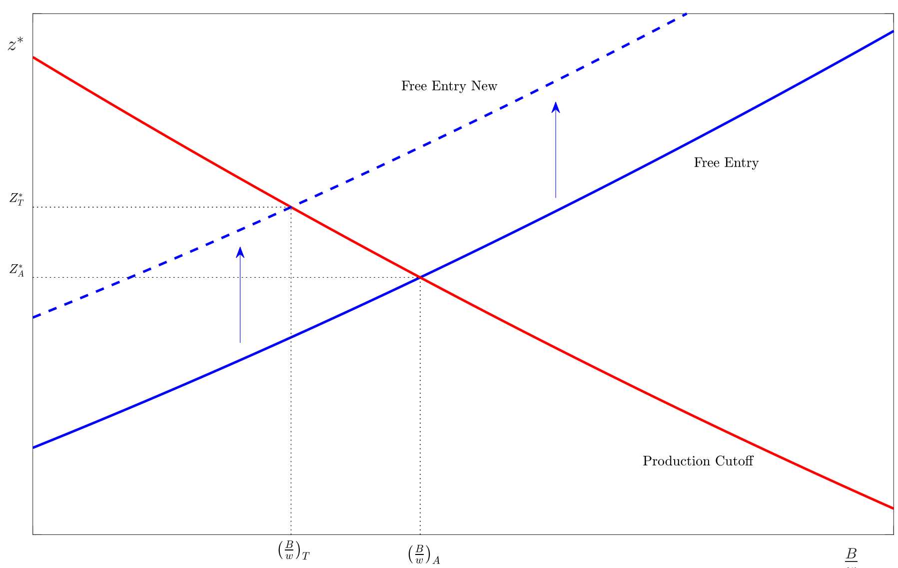
[74] This equation can be derived just as the autarky one, remarking that firms with productivity between z^* and z_X^* make a profit π_D and firms with productivity larger than z_X^* make a profit π_X.

algebra, that system writes

$$\begin{cases} z^{*\varepsilon-1}\frac{B}{w} = f. \\ \int_{z^*}^{+\infty}\left(z^{\varepsilon-1}\frac{B}{w} - f\right)g(z)\,dz + \int_{z_X^*}^{+\infty}\left(z^{\varepsilon-1}\tau^{1-\varepsilon}\frac{B}{w} - f_X\right)g(z)\,dz = f_E. \end{cases} \tag{112}$$

The entry cut-off line, defined by the first equation, is exactly the same as in autarky. However, the free entry condition now has one additional, positive term on the right-hand side, representing the expected extra profits from exporting for the most productive firms. Thus, this line is now shifted upwards with respect to autarky, as shown in Figure 22: for the same level of aggregate demand and costs (summed up by $\frac{B}{w}$), profits are higher. Therefore, the entry cut-off must be higher too, to bring expected profits back in line with the entry cost.

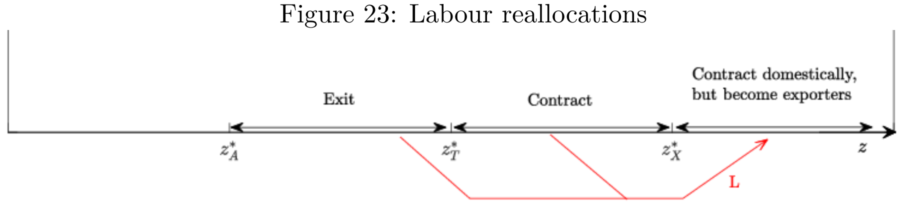
Figure 22: Trade equilibrium for z^* and $\frac{B}{w}$



Thus, trade in the Melitz model increases the entry cut-off z^*, and decreases $\frac{B}{w}$. This already enables us to understand some of the model's most important predictions.

First, trade forces the least productive firms to exit: all firms with productivity draws between z^*_{Autarky} and z^*_{Trade} were profitable in autarky, but are now forced to leave the market. Second, trade leads to a contraction of firms which remain purely domestic, that is, firms with productivity draws between z^*_{Trade} and z_X^*. Indeed, firms which only sell to the domestic market employ $z\left(i\right)^{\varepsilon-1}\left(\varepsilon-1\right)\frac{B}{w}$ production workers, and sell $z(i)^{\varepsilon}(\varepsilon-1)\frac{B}{w}$ units, as shown previously. As $\frac{B}{w}$ decreases, employment and production of these firms shrinks. Finally, trade leads to an expansion of the most productive firms, those with productivity higher than z_X^*. These firms also reduce employment and production on the domestic market, but compensate this by greater employment and production for the foreign market.[75] This reallocation of labour from the least productive firms (which either exit or contract) to the most productive ones (which start exporting and therefore expand) is driven, as in the Krugman model, by foreign competition. As foreign varieties enter the domestic market, they lower the sales of all domestic firms, driving the least productive ones out of business. However, the most productive firms more than compensate for this by taking advantage of their new export opportunities.

Figure 23: Labour reallocations



Finally, note that all this discussion for the passage from autarky to a trade equilibrium also holds for a decrease in the iceberg trade cost τ: a decrease in τ shifts the free-entry curve upwards, increases the cut-off productivity z^* and leads to a reallocation of labour towards more productive firms. This is reassuring for the model's generality, as in reality, countries rather experience progressive reduction in trade costs than radical transitions from autarky to trade.

Aggregation To conclude, we compute the equilibrium mass of varieties and the real wage.

We focus first on the mass of firms which pay the entry cost at home, M_E. This mass could be calculated as in autarky, by using the labour market clearing condition. However, to illustrate a different possibility, we will calculate it with another approach here, using firm sales. The total sales of a firm with productivity $z(i)$, on the domestic and on the foreign market, are[76]

$$\text{Sales}(i) = \varepsilon z(i)^{\varepsilon-1} B \text{ and } \text{Sales}^*(i) = \tau^{1-\varepsilon}\varepsilon z(i)^{\varepsilon-1} B,$$

showing, by the way, that firm sales are also increasing in firm productivity. Now, using a similar reasoning than in the Dornbusch-Fischer-Samuelson model (see Chapter II), we can say that the

[75] There is full employment, so the labour freed up by exiting and contracting firms must necessarily be employed by exporters. Showing that all exporters expand employment is a bit more involved, and we skip it here.

[76] These expressions can be easily obtained by combining the demand equation with firms' price-setting rules.

economy's total labour income must equal the aggregate sales of all domestic firms (including those which exit right away). Indeed, although some firms make profits, the average profit, net of entry costs, is 0. So, we have

$$wL = M_E \left(\int_{z^*}^{+\infty} \varepsilon z^{\varepsilon - 1} B g(z)\, dz + \int_{z_X^*}^{+\infty} \tau^{1-\varepsilon} \varepsilon z^{\varepsilon-1} B g(z)\, dz \right)$$

After some algebra, and making use of Equation (111), we get

$$M_E = \frac{L}{\varepsilon \left(f_E + (1 - G(z^*)) f + (1 - G(z_X^*)) f_X \right)}. \tag{113}$$

Therefore, the mass of goods produced in the domestic economy is

$$M_P = (1 - G(z^*)) M_E.$$

From this, we can directly note that, just as in the Krugman model, the mass of varieties produced in the domestic economy is lower in the trade equilibrium than in autarky. However, it is no longer true that the total mass of varieties consumed is $2M_P$, as not all varieties are exported. Instead, we have

$$M = M_P + M_P^*, \text{ with } M_P^* = \frac{1 - G(z_X^*)}{1 - G(z^*)} M_P.$$

Finally, we turn to the real wage. The definition of the ideal price index can now be written

$$P = \left(\int_0^{M_P} \left(\frac{\varepsilon}{\varepsilon - 1} \frac{w}{z(i)} \right)^{1-\varepsilon} di + \int_0^{M_P^*} \left(\frac{\varepsilon}{\varepsilon - 1} \frac{\tau w}{z(i)} \right)^{1-\varepsilon} di \right)^{\frac{1}{1-\varepsilon}},$$

taking into account both domestically produced and imported goods. After some algebra, the real wage can finally be written

$$\frac{w}{P} = \frac{\varepsilon - 1}{\varepsilon} \left(\left(1 + \frac{1 - G(z_X^*)}{1 - G(z^*)} \right) M_P \right)^{\frac{1}{\varepsilon - 1}} \left(\left(\frac{\int_{z^*}^{+\infty} z^{\varepsilon-1} \frac{g(z)}{1-G(z^*)} dz + \frac{1-G(z_X^*)}{1-G(z^*)} \int_{z_X^*}^{+\infty} \left(\frac{z}{\tau}\right)^{\varepsilon-1} \frac{g(z)}{1-G(z_X^*)} dz}{1 + \frac{1-G(z_X^*)}{1-G(z^*)}} \right) \right)^{\frac{1}{\varepsilon-1}}. \tag{114}$$

Just as in autarky, the real wage is increasing in the mass of varieties consumed, and in the average productivity with which these varieties are produced. Having solved for z^* and z_X^*, we know the equilibrium values of these variables.

To show the existence of gains from trade, it is however more convenient to use the definitions of B and P to express the real wage as

$$\frac{w}{P} = \left(\frac{1}{\varepsilon - 1} \left(\frac{\varepsilon}{\varepsilon - 1} \right)^{-\varepsilon} \frac{L}{B/w} \right)^{\frac{1}{\varepsilon - 1}}. \tag{115}$$

As we have seen previously, trade lowers the value of $\frac{B}{w}$. Therefore, it increases the value of the real wage: there are always gains from trade in the Melitz model.

3.3.3 Heterogeneity and selection

The equilibrium system defining z^* and $\frac{B}{w}$ in the Melitz model does not depend on a country's population size L. Therefore, an increase in population size does not affect these variables, but it "only" increases the mass of varieties M consumed in equilibrium, leading to higher real wages because of workers' love for variety. Thus, while we saw in the Krugman model that opening up to an identical economy had exactly the same effects as doubling country size, this is not true in the Melitz model, where opening up triggers a change in the cut-off z^* (and the appearance of a cut-off z_X^*).

The defining feature of the Melitz model that accounts for these differences is selection due to fixed trade costs.[77] Indeed, if $f_X = 0$, opening up is again equivalent to doubling population size: it just increases the mass of varieties without changing the allocation of labour among them. Positive fixed costs, on the other hand, makes trade have asymmetric effects across firms.

4 The Melitz–Chaney model and the ACR formula

4.1 A Melitz model with an analytical solution

The Melitz model presented in the previous section is very general, as it holds for any productivity distribution g. However, this generality comes at a cost, as we cannot calculate closed-form solutions for many of the model's variables. Thus, to make progress, researchers have tried to feed "realistic" productivity distributions into the model, i.e., productivity distributions which yield a distribution of firm sales and employment that closely resembles real-world data. The most successful approach is

[77] However, if preferences were not CES, increases in market size L may also affect selection.

due to Chaney (2008), who assumed that productivity follows a Pareto distribution. That is,

$$G(z) = 1 - z^{-\theta}, \text{ with } \theta \geq 1 \text{ and } \theta > \varepsilon - 1.$$

When $\theta = 1$, the Pareto distribution is the exact continuous equivalent of Zipf's law, described in Section 2. However, note that the distribution of sales and employment in the Melitz model is a transformation of the productivity distribution (sales are not proportional to z, but to $z^{\varepsilon-1}$), so that $\theta = 1$ is not necessarily the empirically most relevant choice.

We can use this distribution to calculate closed-form solutions for most relevant variables in the Melitz model. You will do this mainly in your Problem Set. For now, it is sufficient to note that (as you will prove in the Problem Set), the mass of firms paying the entry cost is not affected by trade if the productivity distribution is Pareto: we have

$$M_E^{\text{Autarky}} = M_E^{\text{Trade}} = \frac{\varepsilon - 1}{\varepsilon\theta} \frac{L}{f_E}. \tag{116}$$

Therefore, as trade increases the cut-off productivity z^*, the mass of domestic varieties (and therefore firms) falls. So, there are various effects acting on welfare: the mass of domestic varieties falls (which is negative), but there are now imported varieties (which is positive), and due to harder selection, the average firm in the market is now more productive (which is positive as well). What is the net balance of these effects?

To determine this, let us introduce a country's spending share on its own goods, and denote it by π_{nn}, in analogy to the Ricardian models covered in Chapter II. In the trade equilibrium, this spending share equals[78]

$$\pi_{nn} = \frac{\int_{z^{\text{Trade}*}}^{+\infty} z^{\varepsilon-1} g(z)\, dz}{\int_{z^{\text{Trade}*}}^{+\infty} z^{\varepsilon-1} g(z)\, dz + \tau^{1-\varepsilon} \int_{z_X^*}^{+\infty} z^{\varepsilon-1} g(z)\, dz}.$$

Using the expressions for real wages in autarky and in the trade equilibrium, as well as the fact that the mass of entrants is the same in both cases, we can now express[79]

$$\frac{(w/P)_{\text{Trade}}}{(w/P)_{\text{Autarky}}} = \pi_{nn}^{\frac{1}{1-\varepsilon}} \left(\frac{\int_{z^{\text{Trade}*}}^{+\infty} z^{\varepsilon-1} g(z)\, dz}{\int_{z^{\text{Autarky}*}}^{+\infty} z^{\varepsilon-1} g(z)\, dz} \right)^{\frac{1}{\varepsilon-1}}.$$

[78]This formula can be easily derived by using the definition of the spending share on domestic goods, pricing and demand formulas.

[79]To see this, note that we can rewrite $(w/P)_{\text{Autarky}} = \frac{\varepsilon-1}{\varepsilon} M_E^{\frac{1}{\varepsilon-1}} \left(\int_{z^{\text{Autarky}*}}^{+\infty} z^{\varepsilon-1} g(z)\, dz \right)^{\frac{1}{\varepsilon-1}}$ and $(w/P)_{\text{Trade}} = \frac{\varepsilon-1}{\varepsilon} M_E^{\frac{1}{\varepsilon-1}} \left(\left(\int_{z^{\text{Trade}*}}^{+\infty} z^{\varepsilon-1} g(z)\, dz + \tau^{1-\varepsilon} \int_{z_X^*}^{+\infty} z^{\varepsilon-1} g(z)\, dz \right) \right)^{\frac{1}{\varepsilon-1}}$, where M_E is the same in both cases.

Using the density g of the Pareto distribution to solve for the integrals, this expression simplifies to

$$\frac{(w/P)_{\text{Trade}}}{(w/P)_{\text{Autarky}}} = \pi_{nn}^{\frac{1}{1-\varepsilon}} \frac{z^{\text{Trade}*\frac{\varepsilon-1-\theta}{\varepsilon-1}}}{z^{\text{Autarky}*\frac{\varepsilon-1-\theta}{\varepsilon-1}}}.$$

Finally, considering the definition of the productivity cut-offs $z^{\text{Autarky}*}$ and $z^{\text{Trade}*}$, Equations (98) and (108), and using the definition of the auxiliary term B, we can show that both cut-offs can be expressed as simple functions of the real wage:[80]

$$z^{\text{Autarky}*} = \left(\frac{(\varepsilon-1)\left(\frac{\varepsilon}{\varepsilon-1}\right)^{\varepsilon}}{L}\right)^{\frac{1}{\varepsilon-1}} \left(\frac{w}{P}\right)_{\text{Autarky}} \quad \text{and} \quad z^{\text{Trade}*} = \left(\frac{(\varepsilon-1)\left(\frac{\varepsilon}{\varepsilon-1}\right)^{\varepsilon}}{L}\right)^{\frac{1}{\varepsilon-1}} \left(\frac{w}{P}\right)_{\text{Trade}}.$$

Therefore, we finally get

$$\frac{(w/P)_{\text{Trade}}}{(w/P)_{\text{Autarky}}} = \pi_{nn}^{-\frac{1}{\theta}}. \tag{117}$$

So, in the Melitz-Chaney model, we get the same formula for gains from trade than in the Eaton-Kortum model, a result first pointed out by Arkolakis, Costinot and Rodriguez-Clare (2012). Again, we can compute gains from trade just by looking at its import share of GDP and at one additional parameter. Although this parameter has a different interpretation in both models (the dispersion of a Fréchet distribution in Eaton-Kortum, and the shape parameter of a Pareto distribution in Melitz-Chaney), it can be estimated in the same way, as the trade elasticity in a gravity equation. Indeed, it is straightforward to show that the Melitz-Chaney model delivers the same gravity equation as Eaton-Kortum.

4.2 New trade models, same old gains?

Arkolakis, Costinot and Rodriguez-Clare published their discovery under the suggestive title "*New Trade Models, Same Old Gains*?". Their results imply that in order to calculate gains from trade in the Melitz-Chaney model, firm heterogeneity does not matter: it is enough to observe the import share of GDP, and to estimate the trade elasticity from an (aggregate) gravity equation.

Melitz and Redding (2015) disagree with this interpretation in their reply "*New Trade Models, New Welfare Implications*". They point out that the ACR formula is only true under very strong assumptions, which may not hold in reality. For instance, if the productivity distribution is not Pareto, but a truncated Pareto (with a finite upper bound for productivity), one gets a different formula, which

[80] This uses the fact that $U = \frac{wL}{P}$, which holds because of the definition of the ideal price index (see Chapter I).

depends on firm-level variables such as the cut-offs for exit and export. Melitz and Redding also provide an useful thought experiment that clarifies the welfare differences between models with homogeneous and heterogeneous firms. Consider a version of the Krugman (1980) model with two countries and homogeneous firms, but fixed and variable export costs as in Melitz. Furthermore, assume that parameters are such that the autarky real wage of these countries equals the autarky real wage of the Melitz model. When these two countries open up in the Krugman world, some firms may exit,[81] but all surviving firms behave symmetrically: they produce the same amount, and either they all export or they refrain from doing so. This symmetric reaction would have been possible in a Melitz world, too. However, it did not occur, and there was instead a change in the exit cut-off, a contraction of domestic firms, and an expansion of exporters. As the market equilibrium achieves the highest possible welfare,[82] the fact that the Melitz world does not react like the Krugman world even though it could, must mean that the Melitz world achieves a higher increase in the real wage. Thus, the welfare gains from trade are higher when firms are heterogeneous.[83]

5 Empirical evidence

5.1 Trade, reallocation and productivity

Just like the Eaton-Kortum model, the Melitz-Chaney model generates a gravity equation.[84] However, the model also has much more precise implications for the reaction of an industry to international trade, which have been the focus of most empirical studies.

These studies have typically found evidence for considerable reallocation of resources within industries after a trade liberalization. Pavcnik (2002) showed that aggregate productivity went up considerably after a trade liberalization in Chile, and estimated that two thirds of this increase were due to a reallocation of resources from less to more efficient producers. Trefler (2004) showed a similar finding for Canada: as a consequence of the Canada-US free trade agreement in 1989, the contraction of low-

[81] In Chapter IV, we had described the Krugman (1980) model without trade costs, in which the mass of varieties produced in a country is the same in autarky and with trade. However, as soon as there are trade costs, the mass of domestically produced varieties falls when a country opens up.

[82] This is not obvious, as there is a market imperfection, monopoly power. However, it can be shown that with CES preferences and inelastic labour supply, the equilibrium is efficient even in the presence of monopoly power.

[83] Several other studies have tried to quantify the gains from trade in a Melitz setup, providing further insights. For instance, di Giovanni and Levchenko (2013) find that if firm size follows Zipf's law, changes in variable trade costs do not affect the number of traded varieties very much. This is because these changes only affect the export status of firms with intermediate sizes and productivity (the largest and most productive firms were already exporting with a higher τ). However, these large firms represent the lion's share of economic activity, so that most gains from a reduction in variable trade costs come from them being able to export greater quantities.

[84] For an overview of models which can do so, see Head and Mayer (2014).

productivity plants in Canada reduced employment by 12%, but industry-level labour productivity increased by 15%. Finally, Bernard, Jensen and Schott (2006) focus on the United States, and find that in industries where trade costs have fallen most, productivity increases have been largest, and that low-productivity plants in industries where trade costs were falling a lot were more likely to exit than low-productivity plants in industries where trade costs were falling less.

5.2 Disentangling productivity and market power

While the finding that industry-level productivity increases after trade liberalizations appears to vindicate the Melitz model, it is also subject to important measurement problems. Indeed, as De Loecker and Goldberg (2014) remark in a recent survey, empirical productivity measures are based on a ratio between firm (or industry) sales to inputs. Therefore, they do not only capture productivity or product appeal, but also the market power of firms: in theory, it could be that these productivity measures increase only because firms increase their prices for the same products. To settle this question, one needs to disentangle productivity (or product appeal) from mark-ups. The recent literature has proposed three ways of doing so.

A first approach, used in Foster, Haltiwanger and Syverson (2016), focuses on homogeneous goods. The authors limit their analysis to a number of industries in which they can observe physical quantities produced, and in which physical output can be reasonably assumed to be the same across firms (for example, ready-mixed concrete, cardboard boxes, or manufactured ice). In these industries, they can directly calculate physical productivity and there are no differences in product appeal, so that all remaining differences in revenue productivity should be due to mark-ups. Using this methodology, they show that in these industries, differences in physical productivity appear to be small. Larger and older firms have higher revenue productivities, but this is only because they are able to charge higher prices than small and young ones. While they do not expicitly look at trade, their findings already suggest some caution in interpreting reallocations towards high revenue-productivity firms as necessarily beneficial. However, it is also unclear whether their results are valid beyond the very specific industries they consider.

A second approach, common in Industrial Organization studies, consists in making assumptions about demand functions and market structure, and then estimate mark-ups using this information.[85] De Loecker (2011) uses this approach in a study of the Belgian textile industry. His results show

[85] For instance, in the Melitz model with CES demand and monopolistic competition, we know that mark-ups are only a function of demand elasticity. If we can estimate the latter in the data, we would know the mark-ups of all firms.

that separating productivity and mark-ups can be very important: while revenue productivity in the Belgian textile industry has gone up by 8% after a trade liberalization, actual productivity has only increased by 2%. Thus, there are gains from selection as in the Melitz model, but they may have been overestimated (a lot) with conventional revenue productivity measures. However, while this study is very insightful, it also relies on quite a lot of assumptions: if the researcher makes a mistake in specifying the demand, market or production structure, she may get misleading results.

A third approach put forward by De Loecker and Warzynski (2012) relies on a simpler strategy, which "only" needs to assume cost minimization by producers and estimate the production function. To understand this methodology, consider a firm which produces output with the production function $y = zf(k,l)$,where k and l stand for the capital and labour inputs, and z for the productivity. Assuming that the firm takes input prices as given, its cost minimization problem is

$$\begin{array}{c} \min_{k,l} wl + rk \\ s.t.\ y = zf(k,l) \end{array}. \tag{118}$$

Setting up the Lagrangian, the first-order condition of this problem for labour is

$$w = \lambda \frac{\partial f}{\partial l}(k,l),$$

where λ is the Lagrange multiplier on the constraint. However, as we have seen before (e.g., in Chapter III), the Lagrange multiplier is equal to the marginal cost of production. This allows us to introduce the mark-up μ, which is by definition $\mu = \frac{p}{\lambda}$. So, we have

$$\mu w = p \frac{\partial f}{\partial l}(k,l).$$

Multiplying this expression by $\frac{l}{f(k,l)}$, and rearranging terms, we

$$\mu = \frac{\frac{\partial f}{\partial l}(k,l)\frac{l}{f(k,l)}}{\frac{wl}{pf(k,l)}}. \tag{119}$$

That is, the mark-up can be calculated as the ratio between the elasticity of output with respect to labour ($\frac{\partial f}{\partial l}(k,l)\frac{l}{f(k,l)}$) and the labour share of output ($\frac{wl}{py}$). The labour share is in general easy to observe in the data. The elasticity of output is more difficult to observe, but it can be estimated (for details, see De Loecker and Warzynski (2012) and the references therein).

Using this methodology, the authors find that in a dataset of Slovenian manufacturing firms, exporters on average charge higher mark-ups than non-exporters, and firms which enter export markets increase their mark-ups when doing so. This again suggests that it is important to take into account market power when analysing trade liberalizations, and that the simple world of the Melitz model, where all firms charge the same constant mark-up, is quite different from reality. Some extensions discussed in the next section have modified this assumption in order to bring the model closer to the data.

6 Applications and extensions of the Melitz model

6.1 Comparative advantage, competition and inequality

The Melitz model has had a huge influence on research in international trade, and given rise to a large number of studies which extend the basic model to study new issues. I provide here a short overview of some of the most important applications, before treating one of them in greater detail.

Classical trade theory The Melitz model can of course be combined with more classical models of inter-industry trade. Bernard, Redding and Schott (2007), for instance, propose a Melitz-Heckscher-Ohlin model. They show that the Melitz reallocation effects are strongest in comparative advantage industries. Intuitively, in these industries, there are more export opportunities, and therefore selection through exporting is stronger.

Shocks and international business cycles Ghironi and Melitz (2005) have proposed an influential dynamic extension of the Melitz model, in which there are shocks to aggregate productivity or to trade costs. Their framework can be used to analyze how international trade matters for transmitting shocks from one country to another, and links together the fields of International Trade and International Macroeconomics.

Competition and mark-ups As we had already seen in Chapter IV, CES preferences imply that firm's mark-ups do not react to international trade. Melitz and Ottaviano (2008) show that this result is not robust to other preferences. They solve the Melitz model with a different set of (quadratic) preferences, and find that this reintroduces the pro-competitive effect of trade stressed by the original Krugman (1979) model: trade liberalizations reduce mark-ups.

In recent years, there has been a lot of research on the effects of trade on competition. Early studies, such as Krugman (1979) or Melitz and Ottaviano (2008) focused on pro-competitive effects of trade due to import competition: trade lowers the market share of domestic firms, and forces them to charge lower prices. Recent studies, however, also point out a range of anti-competitive effects. For instance, Arkolakis, Costinot, Donaldson and Rodríguez-Clare (2018) point out that rising markups on imports can cancel out the falling markups of domestic firms. De Loecker, Goldberg, Khandelwal and Pavcnik (2016) and Brandt, Biesebroeck, Wang and Zhang (2017) show that in India and China, an important implication of trade liberalization was to lower the cost of imported intermediate inputs. However, firms did not fully pass on these cost savings to consumers, but instead increased their markups. Finally, in Cavenaile, Roldan-Blanco and Schmitz (2020), we argue that by stimulating innovation, lower trade costs can further increase markups: as firms innovate more to remain competitive in a global market, they also increase their relative productivity with respect to their competitors, and therefore increase their market power and markups.

Inequality and labour markets Empirical evidence shows that exporting firms pay higher wages than non-exporting firms (see Table 8, for instance). The baseline Melitz model cannot match this: as all workers are equal and the labour market operates without frictions, all workers receive the same wage w. Helpman, Itskhoki and Redding (2010) extend the Melitz model to incorporate search frictions and differences in workers' ability, generating wage dispersion. They find that trade liberalization tends to increase wage inequality, as a deeper gap opens up between the well-paid workers at successful exporting firms and those who remain in contracting domestic firms.

6.2 Trade, R&D and technology adoption

6.2.1 Empirical evidence

In the Melitz model, trade does not affect a firm's productivity: firms export because they are more productive, but do not become more productive because they export. However, in recent years, a line of studies has shown that exporting appears to lead firms to adopt better technologies and to invest more into Research and Development (R&D), which affects their productivity (for a survey of this literature, see Shu and Steinwender, 2018).

This line of research needed to overcome a difficult identification problem: exporting is not an exogenous, random event that hits a firm, but an endogenous choice, correlated with a large number of the firm's characteristics. Therefore, researchers needed to find suitable instrumental variables for an increase in firms' export opportunities. Lileeva and Trefler (2010) study the impact of the Canada-US free trade agreement of 1989. In their empirical strategy, they exploit the fact that US tariffs on Canadian goods were very diverse before the free trade agreement. However, once the agreement became operational, tariffs for all goods were set to 0%. This represented a much bigger increase in export opportunities for plants[86] which produced products which were still heavily taxed in 1988 than for plants which produced products that were already taxed at very low rates in 1988. Therefore, the reduction in tariffs between 1988 and 1989 can be used as an instrument for exports: it is correlated with plant-level exports, and it is not too implausible to assume that it is uncorrelated with a number of plant characteristics. The authors find that plants which export more due to the tariff change also increase their level of value added per worker (a measure of productivity), do more product innovation and increase their use of advanced production techniques, such as computer-based production scheduling and inventory control. Aghion, Bergeaud, Lequien and Melitz (2017) use a similar empirical strategy to construct instrumental variables for export shocks, and also find that those encourage innovation, although the effects are heterogeneous: firms which are initially more productive benefit more than those that lag behind.

Bustos (2011) presents similar evidence for the effect of the creation of the MERCOSUR (a customs union between Argentina, Brazil, Paraguay and Uruguay) in 1995 on Argentinian firms. Using a similar instrument than Lileeva and Trefler, she finds that firms in industries which saw their export tariffs to Brazil fall faster also increase their "technology spending" (R&D investment and spending on new machines) most. Aw, Roberts and Xu (2011) follow a different empirical strategy, by building a large structural model of exporting and R&D, and estimating it using data from the Taiwanese electronics industry. They also find that exporting is associated with greater R&D investment. Finally, Atkin, Khandelwal and Osman (2017) report results from an experiment in which some randomly selected members of a group of Egyptian rug producers were given access to the export market. They find that these producers increased the quality of their rugs and became better at producing them, which suggests that there has been some "learning by exporting".[87]

What are the economic mechanisms through which export opportunities (and import competition)

[86] Note that Lileeva and Trefler look at production plants, and not at firms. While many firms have only one plant, the largest firms obviously have much more.

[87] The effect of import competition on a firm's technology choices has been less studied than the effect of exports. Bloom, Draca and Van Reenen (2016) argue that import competition from China has increased innovation among European textile firms. Autor, Dorn, Hanson, Pisano and Shu (2016a) find the opposite result for the United States.

affect technology adoption and innovation choices? Bustos (2011) shows that a simple extension of the Melitz model can help to understand also this question. Her model is presented in the next section.[88]

6.2.2 The Bustos model

Bustos adds one additional element to the Melitz model: once a firm has learned its productivity $z(i)$, it can either choose to produce with this "basic" technology, or else to adopt an "advanced" technology, increasing its productivity to $\gamma z(i)$, where $\gamma > 1$. Choosing the advanced technology, however, is not free, but increases the firm's fixed costs to ηf (where $\eta > 1$). These additional costs can be interpreted as R&D investment, or as the costs of purchasing new and better machines.

In the trade equilibrium, this now leaves a firm with five different options after learning its productivity draw: it can exit, produce for the domestic market with the basic technology, produce domestically and export with the basic technology, produce for the domestic market with the advanced technology, or produce domestically and export with the advanced technology. To determine the firm's optimal choice, we need to compare the profits generated by these five options. For firms producing with the basic technology, profits are given by the same expressions as in the Melitz model: for a firm with productivity $z(i)$,

$$\pi_D^B(i) = z(i)^{\varepsilon-1} B - wf \quad \text{and} \quad \pi_X^B(i) = z(i)^{\varepsilon-1} \left(1 + \tau^{1-\varepsilon}\right) B - w(f + f_X). \tag{120}$$

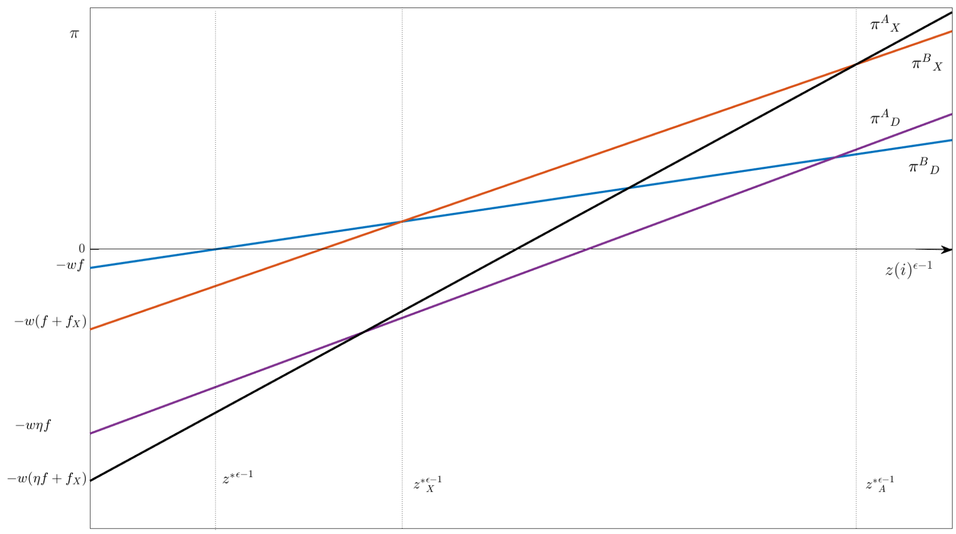
where the superscript B indicates the basic technology. It is also easy to show that the profits of firms producing with the advanced technology are

$$\pi_D^A(i) = (\gamma z(i))^{\varepsilon-1} B - w\eta f \quad \text{and} \quad \pi_X^A(i) = (\gamma z(i))^{\varepsilon-1} \left(1 + \tau^{1-\varepsilon}\right) B - w(\eta f + f_X). \tag{121}$$

These profit functions indicate that exporting and technology adoption are complementary (exporting yields higher variable profits when firms use the advanced technology, and vice-versa). Just as in the baseline Melitz model, profits are linear in $z(i)^{\varepsilon-1}$. As there are fixed costs, it is obvious that the least productive firms, with $z(i)$ close to 0, always exit. Furthermore, the slope of the π_X^A-profit line is higher than the one of all the others, so that the most productive firms will always both export and use the advanced technology. This is precisely because of the complementarity of these two activities. However, what happens in between these two extremes?

[88] In Schivardi and Schmitz (2019), we use a similar setup to study the interaction between technology adoption and management practices in Southern Europe.

Figure 24: Productivity cut-offs in the Bustos model



As in the Melitz model, several cases are possible a priori, and it depends on parameter values which ones occur. Bustos argues that for her data on Argentinian firms, there are large productivity differences even among exporting firms, and takes this as an indication for many exporting firms producing with basic rather than advanced technologies. Thus, she assumes that the π_D^B and the π_X^B-line cross earlier than the π_X^B and the π_X^A-line (so the productivity needed to start exporting is lower than the productivity needed for an exporter to start using the advanced technology). This yields the situation shown in Figure 24, separating firms into four different groups.

Firms with productivity lower than z^* exit. Firms with productivity between z^* and z_X^* produce only for the domestic market, using the basic technology. Firms with productivity between z_X^* and z_A^* export, but keep using the basic technology. Finally, the most productive firms, with productivity above z_A^*, both export and use the advanced technology.

Figure 24 also shows one more crucial result: the π_X^A and π_X^B lines cross before the π_D^A and π_D^B lines.[89] Indeed, for an exporting and for a non-exporting firm, the fixed cost of technology upgrading is the same. However, for an exporting firm, the benefit increases more steeply in productivity $z(i)$, because the upgrade can be applied to two markets rather than one. This illustrates once again the complementarity between exporting and technology adoption, and suggests that trade can increase

[89]It is possible to show this mathematically by using the preceding equations. Note, in particular, that this result hold for any values of B and w.

firm-level productivity by pushing firms over the edge of technology adoption.

The model can now be solved just as the baseline Melitz model. The exit cut-off z^* is defined by

$$z^{*\varepsilon-1}\frac{B}{w}=f. \tag{122}$$

The export cut-off z_X^* is defined exactly as in the baseline Melitz model, and we therefore have

$$z_X{}^*=\tau\left(\frac{f_X}{f}\right)^{\frac{1}{\varepsilon-1}}z^*. \tag{123}$$

Finally, z_A^* is defined by the condition $\pi_X^A\left(z_A^*\right)=\pi_X^B\left(z_A^*\right)$, which can be rewritten as

$$z_A^{*\varepsilon-1}\left(\gamma^{\varepsilon-1}-1\right)\left(1+\tau^{1-\varepsilon}\right)\frac{B}{w}=(\eta-1)\,f.$$

Dividing this expression by Equation (122), we finally get

$$z_A{}^*=\left(\frac{\eta-1}{\left(\gamma^{\varepsilon-1}-1\right)\left(1+\tau^{1-\varepsilon}\right)}\right)^{\frac{1}{\varepsilon-1}}z^*. \tag{124}$$

From this, we can derive an expression for the expected profit of a firm, and derive a free entry condition which gives us, as in the standard model, a second relationship between z^* and $\frac{B}{w}$:

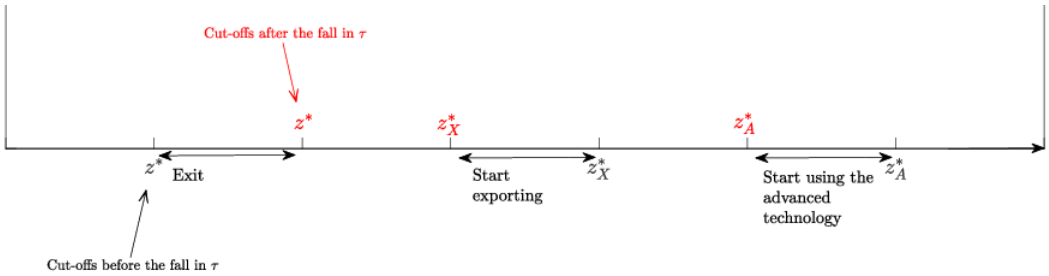
$$f_E=\Phi\left(z^*,\frac{B}{w},\tau\right), \tag{125}$$

where Φ is a complicated non-linear function which is decreasing in z^*, increasing in the demand term $\frac{B}{w}$ and decreasing in variable trade costs τ. The intersection of the curves defined by Equations (122) and (125) gives the equilibrium values of z^* and $\frac{B}{w}$, and from these, we can deduce the value of the other cut-offs z_X^* and z_A^*.

The effect of a fall in variable trade costs As in the Melitz model, a fall in variable trade costs τ shifts the free-entry curve upwards, and therefore, the productivity cut-off z^* rises and $\frac{B}{w}$ falls. Intuitively, increasing competition from foreign varieties hits the least productive firms hardest, as they cannot compensate the fall in domestic sales with greater export revenues.

What happens to the two other cut-offs, z_X^* and z_A^*? Looking at Equations (123) and (124), it is clear that both $\frac{z_X{}^*}{z^*}$ and $\frac{z_A{}^*}{z^*}$ decrease as well, meaning (for most productivity distributions) that the

Figure 25: The impact of a decrease in variable trade costs



fraction of surviving domestic firms which export and the proportion of surviving domestic firms which adopt the advanced technology increase. Furthermore, it can be shown that the absolute cut-offs z_X^* and z_A^* fall as well.[90]

Therefore, a fall in trade costs always increases the percentage of firms which invest into the advanced technology ($\frac{1-G(z_A^*)}{1-G(z^*)}$), through two effects. First, as selection gets harder, a larger percentage of firms now have productivity draws above the cut-off level necessary for using the advanced technology (the denominator decreases). Second, as exporting gets easier, there are more incentives to invest in technology (the numerator increases). As shown in Figure 25, the firms which are most likely to adopt better technologies after a trade liberalization are not the most productive ones (as those were already using the advanced technologies before), but firms with intermediate productivities. Bustos finds that this indeed occurred in Argentina.

This model, and the empirical findings on the effect of trade on technology adoption and R&D investments, provide a new perspective on gains from trade. Indeed, one of the most well-known facts in economics is that long-run GDP growth is mostly due to increases in productivity (Jones, 2016). However, productivity increases are not random, but generally the results of investments by governments, firms and private people into R&D. Therefore, if trade affects R&D decisions, it may affect a country's productivity growth rate. This suggests that apart from the static gains of trade which we have studied in this course (that is, the effects of trade on the level of GDP), there could be dynamic effects (on the growth rate of GDP). As these dynamic effects compound over the years, they are potentially more important than the static ones. Recently, there has been renewed interest in this topic, especially because, as we have seen in Chapter II, static quantitative trade model predict relatively modest gains from trade. We do not study this literature in detail here, but if you are interested, you can read a little about it in the Appendix (Chapter X).[91]

[90] This is calculation-intensive, and done in the online appendix of Bustos (2011).

[91] While we have focused on gains from R&D and technology adoption, trade can also have benefits through increasing

Part VI

Multinational Firms

1 Introduction

Large multinational firms such as McDonalds, Coca Cola, or Apple are probably the most iconic symbols of economic and cultural globalization. However, they played no role in the models we have seen so far: in all of them, firms' production took place in only one country, and the only way for a firm to be present abroad was to export its products. With this restricted focus, we were missing two essential ways in which multinational firms shape today's international economy.

First, the existence of multinationals implies that we cannot fully capture globalization by looking at trade flows. The presence of McDonalds restaurants everywhere around the globe is perhaps one of the most powerful symbols of globalization. However, the sales of a McDonalds restaurant in Milan show up neither in US exports, nor in Italian imports: their "international" aspect comes only from the fact that the Italian restaurant is a subsidiary of a US multinational. This form of international integration is the norm rather than the exception. Yeaple (2013) shows, for instance, that 75% of sales of large US firms to foreigners are done through foreign subsidiaries, and only the remaining 25% are due to exports. Thus, to better understand globalization, we must study why firms often choose to sell through foreign subsidiaries rather than export, and what drives their location choices.

Second, even when focusing on trade flows, multinationals matter. Bernard, Jensen and Schott (2009) show that they account for 90% of all US exports and imports. Even more importantly, roughly 50% of US imports are due to intra-firm trade, that is, to sales of foreign subsidiaries to their US parent firm (or of foreign parent firms to their US subsidiaries). Intra-firm trade is a relatively recent phenomenon, and shows that many firms have offshored parts of their production process around different countries of the world, creating complex supply chains. To understand the patterns of modern trade, we therefore need to understand the determinants and the consequences of this offshoring process as well.

In the remainder of this chapter, Section 2 provides some important definitions and stylized facts about the activities of multinational firms. We then turn to study sales through foreign subsidiaries in Section 3, and offshoring in Section 4.

technology diffusion. Sampson (2016) builds a Melitz model with technology diffusion to make this point, and shows that it suggests substantially larger gains from trade than static models.

2 Some stylized facts on multinational firms

2.1 Definitions

Caves (2007) provides a straightforward definition of a multinational firm: it is "*an enterprise that controls and manages production establishments - plants - located in at least two countries*". In general, the multinational firm consists of a parent firm, incorporated in the home country, and subsidiaries in foreign countries, but for various reasons (among which tax "optimization" is a prominent one), the corporate and legal structure of multinational firms is often complicated. In what follows, we will follow the international norm and consider any firm in which the parent controls more than 10% of the voting stock as a subsidiary. The acquisition of this voting stock (by buying an existing foreign firm or creating a new one from scratch) is called Foreign Direct Investment (FDI).

2.2 The most important features of multinational firms' activities

In a recent survey, Antràs and Yeaple (2014) provide a number of important stylized facts on multinational firms. Three of them are particularly relevant for our focus.

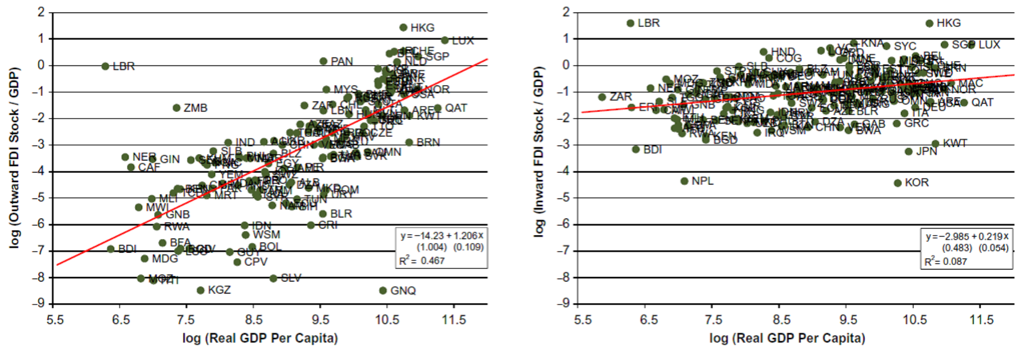
Fact 1. *Most multinational activity takes place between developed countries, is intra-industry, and is concentrated in capital and R&D-intensive industries. Developing countries receive some FDI, but their firms do little FDI themselves.*

Figure 26 illustrates this fact by showing the correlations between countries' GDP per capita and their ratio of outward and inward FDI over GDP (outward FDI is the FDI that a country's firms do abroad, while inward FDI is the FDI a country receives). In both cases, there is a positive correlation: richer countries both have more multinational firms investing abroad, and more multinational firms coming to their territory to invest there. However, the slope of the regression is much lower for inward FDI than for outward FDI, illustrating that developing countries do receive some FDI.

The outliers in these graphs are also interesting. Note, for instance, that countries such as Korea and Japan receive extremely little FDI given their level of development. Some other countries have suspiciously high levels of FDI given their level of development: these include some of the world's major tax heavens, such as Liberia, Hong Kong, and Luxembourg. It is likely that a large part of the FDI declared in these countries has no real counterpart.

Furthermore, it can be shown that at least for manufacturing, FDI is concentrated in capital- and R&D-intensive industries (motor vehicles, pharmaceuticals), whereas less capital and R&D intensive

Figure 26: FDI and Development



Source: Antràs and Yeaple (2014, p. 60).

industries (textiles, shoes...) have less multinational firms. Finally, most FDI flows are intra-industry: German car firms open subsidiaries in Italy, and Italian car firms open subsidiaries in Germany. Thus, the patterns of multinational activity are not that different from the patterns of trade in general: a large part of both is concentrated on developed countries (even though FDI in and from the South is getting more important), and involve two-way exchanges of the same things. Therefore, our models of FDI activity (see Section 3) will be similar to Krugman's and Melitz's models of intra-industry trade.

Fact 2. *Subsidiaries mainly sell locally, or to other geographically close markets.*

Table 9 illustrates this claim for the example of US multinational firms. It shows that in all industries, subsidiaries do at least around 50% of their sales in the host country. Exports to the United States represent only a minority of their sales, even though they are significant in some industries (such as, for example, textiles or transport equipment (cars, planes...)).

Table 9: Sales of US multinationals' subsidiaries

Table 2.2 Destination of Affiliate Sales (%) by Industry

	Host Country	Other Foreign	United States
Total manufacturing	55	34	11
Textile and Apparel	45	35	19
Metals and Minerals	60	32	8
Chemicals and Plastics	58	36	6
Machinery	49	36	15
Computers and Electronics	40	43	16
Electronic Equipment	47	40	13
Transport Equipment	47	35	19
Other	66	26	8

Source: 2009 Benchmark Survey of U.S. Direct Investment Abroad, BEA.

Source: Antràs and Yeaple (2014, p. 65).

Together with the first one, this second fact suggests that the main motivation for firms to become multinational is market access, i.e., the possibility to sell in the host country and its neighbours.[92] In particular, this motive is more important than offshoring: if US firms would only go abroad to reduce their production costs for the US market, then a large fraction of their subsidiaries sales would go back to the United States. Of course, this does not mean that offshoring is marginal or unimportant. It just suggests that in order to understand the existence of multinational firms, we need to focus first on the question of why serving a foreign market through a subsidiary might be better than exporting, and not only on international differences in production costs.

Fact 3. *Multinational firms (both parents and their subsidiaries) are larger, more productive, export more, and pay higher wages than non-multinational firms.*

Table 10 illustrates these points for subsidiaries of multinational firms in some European countries.

Table 10: Multinationals and the general population of firms in Europe

Table 2.1 Affiliates Relative to Local Firms (Percentage Accounted for by Affiliates)

	Finland	France	Ireland	Holland	Poland	Sweden
Enterprises	1.6	2.0	13.4	3.4	16.0	2.8
Employment	17.2	26.2	48.0	25.1	28.1	32.4
Sales	16.2	31.8	81.1	41.1	45.2	39.9
R&D Expenditure	13.1	27.4	77.3	35.8	20.9	52.0
Exports	17.5	39.5	92.3	60.0	69.1	45.8

Source: OECD (2007).

Source: Antràs and Yeaple (2014, p. 64).

This stylized fact echoes very much the differences between exporting and non-exporting firms highlighted in Chapter V. Again, just as in the Melitz model, these differences could be due to selection, if only the most productive firms become multinationals.

3 Why do firms become multinationals?

3.1 A simple model of exports and FDI

In this section, we study a simple model due to Brainard (1993, 1997) which analyses under which conditions a firm finds it advantageous to create a foreign subsidiary (and therefore, to become a multinational) and under which conditions it prefers to just export. Brainard shows that while economies of scale would suggest to concentrate production in just one country, creating a foreign subsidiary

[92] The percentage of sales realised in “Other Foreign” countries is high because multinationals often choose one country as a platform for a whole region. For example, Volkswagen has production plants in Brazil, and uses them to serve the entire Latin American market. In the next section, we will study the economic motivations for such decisions.

allows a firm to save of trade costs. Whether firms serve foreign markets through exports or FDI then depends on which side of this "concentration-proximity" trade-off is stronger.

Assumptions Brainard's model overtakes a large number of elements from the basic setup of the Krugman (1980) model with two identical countries: homogeneous firms, monopolistic competition and CES preferences. However, she introduces two novelties. First, variety creation is not free, but that a firm needs to employ f_E units of labour to create a new variety (just as in the Melitz model). Second, firms have two possibilities to sell to a the foreign country: they can either export, subject to an iceberg trade cost $\tau > 1$, or they can open a subsidiary abroad. In the latter case, they can produce abroad with the same production function than at home: they need to pay a fixed cost of f units of (foreign) labour, and a variable cost of $\frac{1}{z}$ units of (foreign) labour per unit of their variety produced abroad.

Note that a firm with a foreign subsidiary needs to pay the fixed cost of variety creation f_E only once, but needs to pay the fixed cost of production f twice. Intuitively, f_E stands for the fixed costs which are incurred at the firm level (for example, the design of a new product). f stands for the fixed costs at the plant level, which are incurred for every plant (for example, building the production plant and buying machines to produce the product).

The choice between exporting and FDI To find out whether a firm decides to export or to open a subsidiary in the foreign country, we just need to compare the profits of these two activities.[93] As usual, firms charge a constant mark-up $\frac{\varepsilon-1}{\varepsilon}$ over their marginal cost in both markets, and therefore, the profits of an exporter are

$$\pi_X = \left(1 + \tau^{1-\varepsilon}\right) z^{\varepsilon-1} B - wf, \tag{126}$$

where $B \equiv \frac{1}{\varepsilon-1}\left(\frac{\varepsilon}{\varepsilon-1}\right)^{-\varepsilon} P^{\varepsilon} w^{1-\varepsilon} U$, just as in the Melitz model.

When the firm opens a foreign subsidiary instead, its profit is (because of the model's symmetry) just the double of its home market profits:

$$\pi_S = 2\left(z^{\varepsilon-1} B - wf\right). \tag{127}$$

Equations (126) and (127) illustrate a trade-off between concentration and proximity. The fixed costs

[93] As there is no fixed cost of exporting f_X, we can be sure that firms will always either export or open a foreign subsidiary.

of a multinational are higher than those of an exporter ($2wf$ against wf), showing the benefits of concentrating production in one single plant. However, the variable profits of a multinational firm are also higher than those of an exporter (as $2 > 1 + \tau^{1-\varepsilon}$), showing the advantages of being close to the foreign market and thereby avoiding trade costs.

It is straightforward to show that $\pi_S > \pi_X$ if and only if

$$\frac{B}{w} z^{\varepsilon-1} \left(1 - \tau^{1-\varepsilon}\right) > f. \tag{128}$$

When will this condition hold in equilibrium? Assume that we are in an equilibrium where all firms become multinationals. Then, free entry implies

$$wf_E = 2\left(z^{\varepsilon-1}B - wf\right),$$

and this directly yields $\frac{B}{w} = \frac{f_E + 2f}{2z^{\varepsilon-1}}$. Replacing this expression into Equation (128) gives, after some algebra,

$$\tau^{\varepsilon-1} - 1 > \frac{2f}{f_E}. \tag{129}$$

So, an equilibrium in which all firms become multinationals can only occur when Equation (129) is satisfied (and it is straightforward to show that if it does not hold, all firms become exporters instead). This condition depends in an intuitive way on parameters: it is more likely to hold when

1. Trade costs τ are high. Indeed, in that case, the benefits of opening a subsidiary in the foreign country (and therefore avoiding the trade cost) are particularly high. Therefore, we should expect multinationals to be more present in industries where trade costs are very high. The restaurant business represents an extreme case of this: for McDonalds, it would be impossible (otherwise said, infinitely costly) to do its burgers in the United States and then ship them to Italy. Therefore, McDonalds (and Burger King, Starbucks etc.) never choose to export, but instead need to open subsidiaries to serve foreign markets.

2. Fixed costs of production f are low. This is the other side of the concentration-proximity trade-off: if fixed costs would be very high, then firms would be reluctant to open production plants in other countries and incur these fixed costs twice. So, we should expect firms in industries with very high fixed costs not to have production facilities in many different countries. This is the case, for instance, in the airline industry: all of Boeing's production facilities are in the United

States, and Airbus has final production facilities in just three countries of the world, France, China and the United States.

The role of the two other parameters is a little more complex. Equation (129) shows that becoming a multinational is more likely when entry/variety creation costs f_E are high. This comes from the fact that when f_E is high, there will be a low mass of varieties in equilibrium in every country. Therefore, the demand for every individual variety will be high, which makes variable profits (proportional to the demand for each variety) more important than fixed costs. Finally, becoming a multinational is also more likely if the elasticity of substitution ε is high. Indeed, in this case, demand reacts a lot to prices. Therefore, it is important for firms to be able to charge low prices, and therefore reduce their marginal costs from $\tau \frac{w}{z}$ in the case of exporting to $\frac{w}{z}$ in the case of opening a foreign subsidiary.[94]

3.2 Empirical evidence for the concentration-proximity trade-off

Does the concentration-proximity trade-off hold in the data? Brainard (1997) proposes to test this for US firms, by estimating the following regression

$$\ln\left(\frac{X_j^k}{X_j^k + S_j^k}\right) = \beta_0 + \beta_1 \text{Freight}_j^k + \beta_2 \text{Tariff}_j^k + \beta_3 \text{PlantSC}^k + \beta_4 \text{CorpSC}^k + \boldsymbol{\beta}_5 \text{C}_j^k + \varepsilon_j^k. \tag{130}$$

X_j^k stands for the value of exports of US firms in industry k towards country j, while S_j^k stands for the value of sales of subsidiaries of US firms in industry k located in country j. Therefore, $\frac{X_j^k}{X_j^k+S_j^k}$ is the percentage of the market in industry k and country j that US firms cover by exports. This is regressed on two measures of trade costs τ (the costs of shipping goods of industry k from the US to country j, Freight_j^k, and the tariffs that country j applies to US goods of industry k, Tariff_j^k), a measure of production fixed costs f (PlantSC^k, the number of production workers per plant in industry k. The idea here is that the higher this number, the higher are economies of scale and therefore fixed costs f) and a measure of entry/variety creation fixed costs f_E (CorpSC^k, the number of non-production workers per plant in industry k). Given the theoretical results from the previous section, we would expect to find negative estimates for β_1, β_2 and β_4, and a positive estimate for β_3.

In her 1997 paper, Brainard indeed found these results. Table 11 shows an updated version of her regression with more modern data and for more countries, done by Antràs and Yeaple (2014).

[94] Note that we have not fully solved the Brainard model: for instance, we did not solve for the equilibrium mass of varieties M. Doing so would be easy, but does not add to our main point, so we skip it.

Table 11: Results of Brainard regressions

Table 2.3 Proximity-Concentration Empirics

Dep. Var.: log $\left(\frac{x_j^i}{x_j^i+s_j^i}\right)$	(1)	(2)	(3)	(4)	(5)	(6)
Freight	−0.28**	−0.13**	−0.12**	−0.13**	−0.13*	0.01
	[0.05]	[0.04]	[0.04]	[0.04]	[0.06]	[0.25]
Tariffs	−0.23**	−0.28**	−0.27**	−0.29**	−0.38**	−0.04
	[0.06]	[0.05]	[0.05]	[0.06]	[0.10]	[0.04]
GDP/POP	0.10	0.04	0.06			
	[0.07]	[0.08]	[0.08]			
School				0.07		
				[0.09]		
KL				0.08		
				[0.06]		
GDP			0.32	0.39*		
			[0.17]	[0.17]		
PlantSc	0.09*	0.13*	0.13*	0.14*	0.18	
	[0.04]	[0.05]	[0.05]	[0.05]	[0.15]	
CorpSc	−0.18**	−0.32**	−0.31**	−0.32**	−0.35**	
	[0.03]	[0.04]	[0.04]	[0.04]	[0.14]	
Country Fixed Effects?	No	No	No	No	Yes	Yes
Industry Fixed Effects?	No	No	No	No	No	Yes
Year	1989	2009	2009	2009	2009	2009
Observations	1,762	2,315	2,315	2,315	2,482	2,482
R-Square	0.15	0.09	0.09	0.09	0.16	0.40

Standard errors are in brackets.
*At 1%.
**Significant at 5%.

Source: Antràs and Yeaple (2014, p. 77) There is a typo in their legend: $*$ means significant at 5%, and $**$ significant at 1%.

It shows that Brainard's original results still hold: all coefficients have the sign that theory predicts in the baseline regressions for 1989 and 2009 data (columns (1) and (2)). Therefore, the theoretical results for the concentration-proximity trade-off appear to be able to explain the real-world patterns of multinational activity pretty well. However, note that once industry and year fixed effects are introduced (columns (5) and (6)), results become weaker. However, this way be due to the fact that most of the variation in trade costs is at the industry level rather than at the industry-country level.

3.3 Extensions

Markusen and Venables (2000) consider a model with asymmetric countries. Then, they get the classical result from Krugman (1980) that larger countries have higher wages. Furthermore, they show that the larger the differences between countries, the less likely it gets that firms become multinationals. To get the intuition for this result, imagine that one country is very large and the other very small. Clearly, it is not interesting for firms of the small country to become multinationals: by opening a subsidiary in the large country, they would have to pay the (high) large country wage, which would greatly increase their marginal cost. Furthermore, for firms of the large country, it is also not interesting to open subsidiaries in the small country: the labour cost in that country is low, but the market is also very small, so that it does not justify paying the fixed cost of production.[95] Thus, Markusen and

[95] This is an extremely simplified summary of Markusen and Venables, taken from Antràs and Yeaple (2014). In fact, their model is much richer, and allows not only for differences in country size, but also for capital and for therefore for Heckscher-Ohlin type motives for trade.

Venables show that multinational activity is most likely between similar countries, a finding which is fully in line with the data (see Fact 1 in Section 2).

Finally, it is easy to introduce Melitz-style heterogeneity in Brainard's framework, by making productivity $z(i)$ vary at the firm level and introducing a fixed cost of serving the foreign market f_X. This is done in a paper by Helpman, Melitz and Yeaple (2004). In this case, the profit of an exporting firm becomes

$$\pi_X = \left(1 + \tau^{1-\varepsilon}\right) z(i)^{\varepsilon-1} B - w(f + f_X),$$

and the profit of a multinational becomes

$$\pi_S = 2\left(z(i)^{\varepsilon-1} B - wf\right) - wf_X.$$

Note that we assume that the fixed cost of serving the foreign market must be paid no matter whether the firm chooses to serve it through exports or through a subsidiary. From these expressions, it is straightforward to see that the most productive firms will always choose to become multinationals rather than to export, as they can cover the higher fixed costs of being a multinational through higher variable profits. This is fully in line with Fact 3 from Section 2. One can also show that in a model with more than two countries, more productive firms open subsidiaries in a higher number of countries.

4 Offshoring

4.1 The issues

While market access is probably the strongest motivation for firms to open foreign subsidiaries, it is is certainly not the only one. With the fall of communication costs and the liberalization of developing countries since the 1980s, many firms have split up their production process into different stages, performed in different countries either by their subsidiaries or by independent suppliers. There are many striking example of this, such as for instance the Barbie doll, cited by Grossman and Rossi-Hansberg (2006), themselves citing an article of the *Los Angeles Times*: "*the doll is designed in Mattel's headquarters in El Segundo, California. Oil is refined into ethylene in Taiwan and formed into plastic pellets that are used to produce the doll's body. Barbie's nylon hair is manufactured in Japan, while the cotton cloth for her clothing originates in China. The moulds for the doll are made in the United States, as are the paint pigments used to decorate it, and the cardboard used for packaging.*

Assembly takes place in Indonesia and Malaysia. Finally, the dolls are quality tested in California, and marketed from there and elsewhere around the globe." The motivation of firms to split up their production process in this way is clearly not related to market access, put instead to differences in production costs across countries (for instance, Mattel clearly assembles its dolls in Indonesia and Malaysia rather than in the US because wages are lower in the former).

In the public debate, this fragmentation of the production process is commonly referred to as "offshoring". Formally, the economic literature defines offshoring as "*the process of changing the geographic assignment of the mix of tasks needed to produce a [..] final good or service*" (Hummels, Munch and Xiang, 2016). This definition indicates that offshoring has three key elements:

- It involves intermediate inputs, not final goods.
- These intermediate inputs are imported by a domestic firm, ...
- ... even though that same firm may have produced these inputs internally.

The third requirement, in particular, indicates that not all trade in intermediate inputs can be considered offshoring. For instance, when Pirelli imports rubber for its tyres from Brazil, this cannot be considered offshoring, as Pirelli could not have produced this rubber itself in Italy.

To some extent, offshoring is just "regular" international trade (in intermediate inputs), and can therefore be analysed with the same models we have developed so far. However, offshoring also poses some new issues, both for measurement[96] and for theory. In the next section, we examine one modern model of offshoring, due to Grossman and Rossi-Hansberg (2008), emphasizing the new elements with respect to a classic model of international trade.[97] In Chapter VII, we will then study empirical research on the consequences of offshoring (and international trade more generally) on wages and employment.

4.2 A model of trade in tasks

In the models we have covered so far, production was modelled in a very simple way, as a combination of labour and capital. This abstracts from the fact that many production processes are complex, involving different stages and tasks. The starting point of Grossman and Rossi-Hansberg's model is therefore to specify a production technology involving many tasks, of which some can be offshored.

[96] There is no market (and therefore, there are no prices) for within-firm transactions, making it difficult for statisticians to assess the real volume of intra-firm trade of intermediate inputs. These difficulties for "transfer pricing" are also strategically used by many multinational firms to avoid taxes, by shifting their profits from high to low-tax countries.

[97] Other influential models of offshoring include Feenstra and Hanson (1997) and Costinot and Vogel (2010).

The production technology Consider a country (the "North"), which is populated by a mass of H high-skilled workers, and a mass L of low-skilled workers. The North produces two goods, which are also indexed by H and L, and the production of one unit of the H-good needs relatively more high-skilled labour than the production of the unit of the L-good. Both goods are produced under constant returns to scale and perfect competition. Therefore, the basic structure of the model is identical to the Heckscher-Ohlin framework.

However, the production technology is now different. The production of both goods needs a continuum of differentiated high- and low-skill tasks, both indexed on $[0,1]$. More precisely, to produce one unit of the H-good, a firm needs to perform a_H^H times all high-skill tasks and a_H^L times all low-skilled tasks. Likewise, to produce one unit of the L-good, all high-skill tasks need to be performed a_L^H times and all low-skill tasks need to be performed a_L^L times. To carry out a high-skilled task once, a firm needs to hire one unit of high-skilled labour, to carry out a low-skilled task once, it needs to hire one unit of low-skilled labour. If there were no possibility for offshoring, this would just mean that to produce one unit of the H-good, the firm needs to employ a_H^H units of high-skilled labour and a_H^L units of low-skilled labour. We have

$$\frac{a_H^H}{a_H^L} > \frac{a_L^H}{a_L^L},$$

which shows that the H-good is relatively more intensive in high-skilled tasks.

Offshoring possibilities Now, assume that low-skilled tasks can be offshored to another country.[98] This means that a firm can decide to perform any low-skilled task in another country. However, this has costs: while the firm only needs to employ one unit of (home) unskilled labour to perform a certain task i at home, it needs to employ $\beta t\left(i\right)$ units of (foreign) labour, with $\beta t\left(i\right) > 1$ for every i, to perform that task in another country. Therefore, it will only offshore a task if the foreign wage of unskilled labour is sufficiently low to compensate for this cost.

The offshoring cost is essentially an iceberg trade cost for importing the task from a foreign country. It has two components: β is common across all low-skilled tasks, and captures general obstacles to offshoring, such as for instance transport or communication costs. $t\left(i\right)$ is task-specific, reflecting the fact that some tasks are easy to offshore (e.g., typing information into a computer), while others are not (e.g., cleaning an office). We assume that tasks are ordered such that t is an increasing and continuously differentiable function.

[98] For simplicity, we abstract from the possibility to offshore high-skilled tasks. Note, however, that even though offshoring of low-skilled tasks occurs more often, offshoring of high-skilled tasks exists and is becoming more important.

To make things interesting, consider a situation where in equilibrium, the wage of low-skilled labour in the North, w_L, is higher than the wage of low-skilled labour in a foreign country, w_L^*. This naturally occurs when the foreign country (which we may call the "South") has a sufficiently higher relative endowment of low-skilled labour (so that trade does not lead to FPE). In this case, Northern firms may decide to offshore certain tasks. Precisely, a firm (no matter whether it produces the H-good or the L-good) offshores a task i to the foreign country if and only if

$$w_L \geq \beta t(i) w_L^*.$$

As t is an increasing function, we can define a cut-off task I such that all tasks indexed from 0 to I are offshored, while all tasks indexed from I to 1 are performed at home. I is given by

$$w_L = \beta t(I) w_L^*. \tag{131}$$

Therefore, the unit cost of producing one unit of the H-good at home is

$$a_H^H w_H + a_H^L \left(\beta w_L^* \int_0^I t(i) \, di + (1 - I) w_L \right),$$

and an analogous formula holds of course for the L-good. To simplify notation, denote $T(I) \equiv \int_0^I t(i) \, di$. Then, using Equation (131), the unit cost of production for the H-good becomes

$$a_H^H w_H + a_H^L w_L \left(\frac{T(I)}{t(I)} + (1 - I) \right).$$

In this model, offshoring allows specialization by comparative advantage, and therefore aggregate gains from trade. However, Grossman and Rossi-Hansberg are mostly interested in its effects on wages, and therefore study the effect of changes in general offshoring costs β on the wages of both types of workers.

The small open economy case Consider first the case where the North is a small country, and produces both goods in equilibrium. As there is perfect competition, in equilibrium the unit cost of both goods (which is also the marginal cost) must be equal to the world market prices:

$$\begin{cases} p_H = a_H^H w_H + a_H^L w_L \left(\frac{T(I)}{t(I)} + (1-I)\right) \\ p_L = a_L^H w_H + a_L^L w_L \left(\frac{T(I)}{t(I)} + (1-I)\right) \end{cases} \tag{132}$$

Together with Equation (131), this forms a system of three equations with three unknowns (w_H, w_L and I). It is however not necessary to solve for the equilibrium to analyse the effect of a change in β. As the North is small, a change in its offshoring possibilities does not affect world market prices p_H and p_L. Therefore, it will also not affect w_H and $w_L\left(\frac{T(I)}{t(I)} + (1-I)\right)$. Furthermore, it is easy to show that I is decreasing in β, which implies that the cost term $\frac{T(I)}{t(I)} + (1-I)$ is increasing in β, and that a fall in β triggers an increase in the low-skilled wage w_L.[99]

Grossman and Rossi-Hansberg call this the "productivity effect": easier offshoring implies that firms in the home country get more efficient, especially those firms which use a lot of low-skilled labour. These efficiency gains allow them to expand their output and to pay their low-skilled workers better for the remaining tasks that they do at home. While this prediction seems to be at odds with the experience of most developed countries, it is to a certain extent due to the small-country assumption: once we consider that the North is large enough to affect world prices, results may change.

Large countries and general equilibrium effects Indeed, when changes in the production patterns of the North affect world market prices, an adverse effect for low-skilled workers appears. Indeed, as the firms in the L-good sector benefit most from increasing offshoring possibilities, the North's production of L-goods increases with respect to its production of H-goods, and this turns the terms of trade against the L-good sector: the ratio $\frac{p_H}{p_L}$ will increase. Therefore, the Stolper-Samuelson theorem predicts that $\frac{w_H}{w_L}$ will rise as well. When this "terms of trade effect" is stronger than the productivity effect, it can explain that an increase in offshoring coincides with an increase in wage inequality.[100]

[99] To see this, denote $\hat{w}_L = w_L\left(\frac{T(I)}{t(I)} + (1-I)\right)$ and derive both equations in (132) with respect to β to get

$$\begin{cases} 0 = a_H^H \frac{\partial w_H}{\partial \beta} + a_H^L \frac{\partial \hat{w}_L}{\partial \beta} \\ 0 = a_L^H \frac{\partial w_H}{\partial \beta} + a_L^L \frac{\partial \hat{w}_L}{\partial \beta} \end{cases}$$

The only solution to this system is $\frac{\partial w_H}{\partial \beta} = \frac{\partial \hat{w}_L}{\partial \beta} = 0$, because we have assumed $\frac{a_H^H}{a_H^L} \neq \frac{a_L^H}{a_L^L}$.To see that I is decreasing in β, note that we have $w_L = \beta t(I) w_L^*$ and $w_L\left(\frac{T(I)}{t(I)} + (1-I)\right) = X$, where X is a constant which does not depend on β. Combining both equations and taking the derivative with respect to β directly shows $\frac{\partial I}{\partial \beta} < 0$. Combining this with the fact that $\frac{T(I)}{t(I)} + (1-I)$ is decreasing in I gives all the results mentioned in the main text.

[100] In a richer version of the model, there is even a third effect coming from changes in the factor allocation between sectors. For a simple summary of Grossman and Rossi-Hansberg's work, see http://www.economist.com/node/8559758)

Note also that once we endogeneize world market prices, offshoring causes the general price levels to fall, and tends to increase real wages, as there are gains from trade. Therefore, an increase in wage inequality may coincide with an increase in the real wage of low-skilled labour.

Summing up, three things are worth pointing out.

1. At first sight, the predictions of this model do not seem all that different from classical Heckscher-Ohlin theory: if the terms of trade effect dominates, the model predicts that trade with a country that is relatively abundant in low-skilled labour increases the wage gap between skilled and non-skilled labour at home. This is just what the Stolper-Samuelson and the Heckscher-Ohlin theorem suggest.
   However, there are important differences. First, the "productivity effect" described above was absent in the Heckscher-Ohlin theory. It opens new possibilities to gain from trade which could not be exploited when trade was limited to final goods.[101] Second, the theory predicts that the adjustment to trade takes place through within-industry changes rather than through between industry-changes as in the Heckscher-Ohlin case. Third, under certain circumstances, an increase in offshoring does not only increase the wage gap between skilled and unskilled labour in the North, but also in the South.[102] This has been the case in reality (see, for instance, Feenstra (2011) for evidence that wage inequality in China has increased during the 2000s), and therefore, the new theory is an improvement over the Stolper-Samuelson theorem, which had predicted the opposite.[103]

2. The model does not say anything about ownership: the tasks that Northern firms offshore could be performed by their subsidiaries in the South, or by independent firms which sell them these tasks. There is a vast literature on the decision of firms on whether to carry out production tasks by themselves or outsourcing them to independent suppliers. Antràs (2015) provides an overview of this literature, which we will not study here. Note, however, that for all the effects

[101] Task-based models and productivity effects have also been at the centre of the debate on the economic effects of automation (see, for instance Acemoglu and Restrepo, 2020).

[102] This is due to the same terms of trade effect which increases the wage gap in the North. Offshoring creates an efficiency gain which is stronger for the L-good sector, and therefore, the production of that sector increases, its final price falls, and the workers which are most important for that sector see their relative wage fall.

[103] Feenstra and Hanson (1996a) have developed an alternative model of offshoring, which reaches roughly the same conclusions. In their model, there is also a continuum of tasks, but all of those have a different skill intensity. In equilibrium, the North (the country with the relatively lower endowment of low-skilled labour) specializes in high-skilled tasks, and the South in low-skilled tasks. When the South increases its total factor productivity, or its capital stock, it captures a larger share of industries, so offshoring increases. Both the North and the South now produce a set of tasks which are on average more skill-intensive than before (the North has offshored its least skill-intensive tasks, and these have become the most skill-intensive tasks of the South), which increases the gap between the low-skilled and the high-skilled wage in both countries.

that we have derived so far, it is irrelevant whether the firm outsources the offshored activities or does them itself. We will see as well that empirical studies typically do not tell these two forms of offshoring apart.

3. Finally, note that by construction, offshoring has no effects on unemployment. Indeed, this model is (as all the models we have covered) a neoclassical model in which the labour market always clears and there can therefore be no unemployment. Of course, this is not the case in reality. One may think, in particular, that workers which see their task being offshored may have difficulties of adapting quickly to another low-skilled task, and therefore pass through a spell of unemployment.

Part VII

Empirical evidence on Trade, Growth and Inequality

1 Introduction

In this last chapter, we take a step back with respect to the models presented in the previous chapter, in order to consider the newest empirical evidence on some of the topics we have covered.

Identifying causal effects in macroeconomics is famously difficult (see Nakamura and Steinsson, 2018). Economics being a social science, it is impossible to run controlled experiments as in the natural sciences. For example, suppose you want to assess the effects of a 5% increase in the European Union's tariffs on steel imports. Ideally, you would want to compare two different worlds to do so: one in which European steel tariffs increase by 5%, and an exactly identical one in which steel tariffs are unchanged.[104] This is obviously impossible. Then, the best we can do is to measure the changes in output, wages and other economic variables after observed changes in tariffs. However, interpreting these changes is difficult. Suppose we observe that European steel production falls in the months after a tariff is put in place. This may be due to the tariff, but it could also potentially be due to a host of other factors: a recession that lowers production in all industries, a crisis in the steel industry (which maybe induced European steel producers to lobby for the tariff in the first place), etc.

The papers we study in this chapter try to overcome these limitations in various ways. Some of them use "natural experiments", that is, changes in trade costs that occur for non-economic reasons (such as political events or natural disasters). Others rely on more microeconomic data, comparing outcomes across different regions in a country, or across different people. These latter studies are often econometrically much cleaner than studies using aggregate data. However, as we will see later, this comes at a price: if we want to use them to make statements about aggregate outcomes (the ultimate object of interest), we need to make additional assumptions.

[104] Medical experiments are a good example for such a strategy. To test the effect of a new drug, researchers design a control and a treatment group, chosen to be identical with respect to a long list of observable characteristics. The only difference between these groups is that the treatment group gets the drug, while the control group does not. In many of these cases, the assignment of the drug is actually decided by a lottery, ensuring that it is completely independent of any personal characteristics. Such Randomized Control Trials (RCTs) have become very popular in Development Economics as well.

This chapter is divided in two parts: in Section 2, we study the relationship between trade and growth, while in Section 3, we focus on the effects of trade on wage inequality and unemployment.

2 Trade, growth and development

2.1 Estimation issues and early solutions

Perhaps the most basic prediction of trade theory is that trade should generally increase national income (although we have seen several times that there are cases where more trade actually reduces the national income of some countries). How can this prediction be tested?

It is clear that for doing this, one cannot just run a regression of the type

$$\triangle Y_j = \beta \triangle \text{Trade}_j + \varepsilon_j,$$

where Y_j stands for country j's real GDP, and Trade_j is a measure of exports or imports. Indeed, trade is correlated with many other things that also affect GDP. For instance, consider a country that discovers oil. This country will experience a boom in GDP, but also a boom in (oil) exports, and a boom in imports (as its citizens can now import more from the rest of the world). In that case, we would estimate a positive β, but this would not be because trade has caused growth.

To make progress, we need an instrumental variable: something which affects trade, but has no direct, independent effect on GDP. One of the first intents to do this is due to Romer and Frankel (1999). They proposed the average distance to all other countries in the world as an instrumental variable for trade, arguing that this distance affects the volume of trade (through higher trade costs), but should not affect GDP through any other channel. However, this latter claim was quickly disputed. For instance, Rodriguez and Rodrik (2001) argued that distance was correlated with a country's geographic position, and this was likely to have direct effects on development. For instance, countries which are close to the equator are more prone to tropical diseases, and are more likely to have been colonies in the past, both of which can adversely affect their GDP level.

2.2 Ships, planes and the Suez Canal

Recently, a couple of studies have tried to overcome the problems of the Frankel and Romer study by using different instrumental variables.

Pascali (2017) studies the effects of trade on development in the 19th century, and uses the invention of the steamship as a source of exogenous variation in trade costs. Before the invention of the steamship, most long-distance international trade was carried out with sail ships. Therefore, transport times (and therefore trade costs) between two destinations depended on wind patterns. With the invention of the steamship, wind did not matter any more. Therefore, this invention lead to an asymmetric reduction in trade costs, strongest for the countries which previously had an unfavourable wind position. Therefore, Pascali proposes to use the reduction in transport times due to the steamship as an instrumental variable for changes in trade. This reduction is strongly correlated with trade flows, and is unlikely to have affected GDP in any other way than through trade (wind positions are, in particular, not correlated systematically with a country's geographic position).[105]

Table 12: Trade and Development

PANEL A	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
			Dep. Variable = Log Per-Capita GDP					
	(OLS)	(2SLS)	(2SLS)	(OLS)	(2SLS)	(2SLS)	(OLS)	(OLS)
Log (Export/GDP)	-0.0858**	-0.194**	-0.193**					
	(0.0396)	(0.0883)	(0.0922)					
Log (Export/Pop)				0.119***	-0.240*	-0.239*		
				(0.0446)	(0.1367)	(0.1427)		
Log Predict. Export							-0.233**	-0.219**
							(0.105)	(0.104)
COUNTRY DUMMIES	YES	YES	YES	YES	YES	YES	YES	YES
YEAR DUMMIES	YES	YES	YES	YES	YES	YES	YES	YES
r2	0.0351			0.0705			0.0254	0.0239
N	324	313	313	324	313	313	313	313
F		17.34	17.29		11.00	11.02		
WEIGHTED	NO	NO	YES	NO	NO	YES	NO	YES
PANEL B								
Log Predict. Export		1.202***	1.137***		0.970***	0.918***		
		(0.256)	(0.242)		(0.259)	(0.245)		

The table reports OLS and 2SLS estimates. The unit of observation is country-year. The dependent variable is the log of per-capita GDP. "Log Predict Export" is constructed according to equation 6. Observations are un-weighted in columns 1, 2, 4, 5, and 7 and weighted by the log-population of the country in 1860 in the other columns. Panel A reports the second-stage estimates. F is the Kleiberg-Paap Wald F statistics for weak identification. Panel B reports the first-stage estimates. Standard errors (reported in parentheses) are two-way clustered (country and year) corrected to account for the fact that the instrument depends on the (estimated) parameters of the bilateral trade equation. *** significant at less than 1 percent; ** significant at 5 percent; * significant at 10 percent.

Source: Pascali (2017)

Table 12 shows the surprising result of this exercise: increases in trade, instrumented with the reduction of travel times due to steam, trigger on average a reduction in GDP per capita. The same thing is true for other measures of development, such as population density and urbanization.

What is happening? A more precise analysis of the results shows that trade has different effects on different countries: "*although trade tends to be detrimental in Africa, Central America and Asia, it*

[105]This assumption could be criticized: arguably, the reduction in travel times favours not only trade, but also migration, so results could it principle also be due to the latter.

is actually beneficial in Western Europe and North America". Pascali conjectures that this implies that institutions are important: independent countries with strong institutions tended to benefit from trade, while extractive and abusive colonial institutions did not allow many African or Asian countries to do so.[106] These findings also relate to the possibility of "bad specializations" hinted at in Chapter IV, suggesting that trade is not always positive.

Feyrer (2009) and Feyrer (2019) use similar empirical strategies to look at the effect of trade in the 20th century. In one of his papers, he actually uses an estimation strategy which is almost literally the same, looking at the reduction from trade costs due to air travel. From the 1960s onwards, air transportation started to become more and more important, and this reduced trade costs especially for countries which were more difficult to assess by sea. Using this reduction in transport time due to air travel as an instrumental variable for trade, Feyrer finds a large and positive effect of trade on GDP. In particular, he claims that "*differences in predicted trade growth can explain roughly 17 percent of the variation in cross country income growth between 1960 and 1995*." In another paper, he exploits the closing of the Suez canal between 1967 and 1975. This event was arguably exogenous to economic development and affected countries asymmetrically, depending on their distance from the Suez Canal. Again, Feyrer finds a positive and significant effect of trade on income.

3 Trade, wages and inequality

3.1 Wage inequality in developed countries

Since the middle of the 1980s, the wage gap between people with and without university education has increased (see Figure 27). Several European countries saw the same trend, although in general the increase has been weaker. This evolution has given rise to a huge debate about its causes, with the two most popular explanations being international trade and skill-biased technological change.[107]

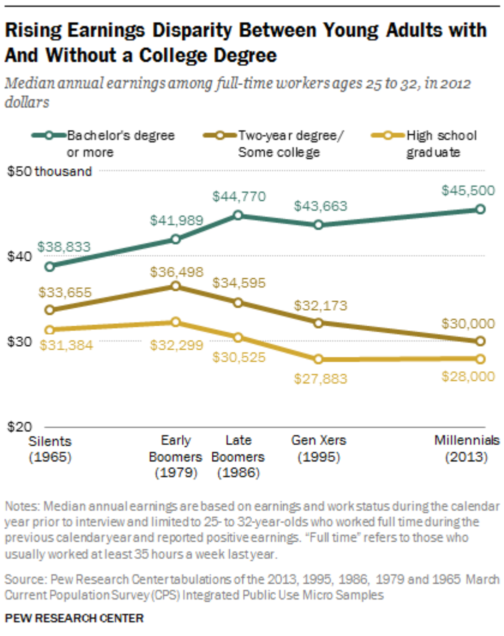
The traditional argument for international trade was based on the Stolper-Samuelson theorem. The theorem predicts that the rise of developing countries in world trade since the 1980s (with large populations, but a small percentage of university graduates) increases the returns to relatively scarce

[106] This is in line with a famous paper by Acemoglu, Johnson and Robinson (2001), which argues that European powers set up very different institutions in their colonies. Colonies where there were many European settlers set up inclusive institutions and developed fast (the United States, Australia, Canada...) while countries with few European settlers set up extractive institutions and did not manage to develop later.

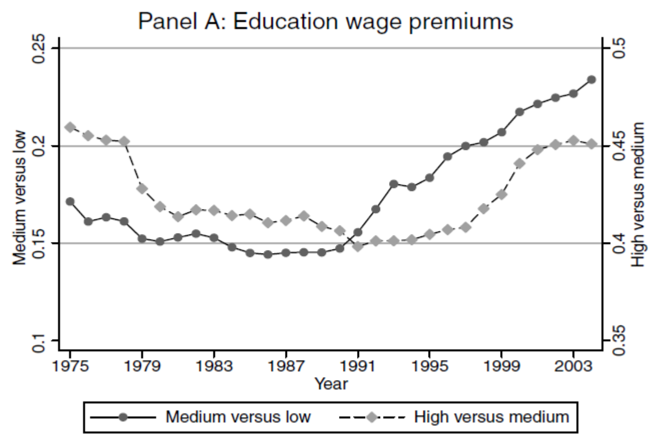
[107] Piketty and Saez (2006) and Piketty's famous book "*Capital in the 21st Century*"(Piketty, 2014) emphasize changes at the very top of the income distribution (for the richest 1 or 0.1% of the population) as a driver of overall higher inequality. While these changes in top incomes are important and interesting, they are not our main focus here.

factors in developed countries (such as people with a university education) and decreases the returns to relatively abundant factors (such as people without university education).

Figure 27: Increasing wage inequality in Germany and in the United States



United States (Pew Research Centre)



Germany (Dustmann, Ludsteck and Schönberg, 2009)

However, towards the end of the 1990s, most economists defended that international trade did not have much to do with the increase in wage inequality. Indeed, low-skilled workers wages fell with respect to high-skilled workers wages even in developing countries, where Heckscher-Ohlin theory suggested they should rise. Furthermore, Heckscher-Ohlin theory suggested that in developed countries, high-skill intensive industries should expand, while low-skill intensive industries should contract, but this was not the case. Thus, Paul Krugman wrote in 1997: "*The rise in demand for skilled workers was overwhelmingly caused by changes in demand within each industrial sector, not by a shift of the U.S.'s industrial mix in response to trade. No one can say with certainty what has reduced the relative demand for less skilled workers throughout the economy. Technological change, especially the increased use of computers, is a likely candidate; in any case, globalization cannot have played the dominant role.*" Just like many other economists, Krugman therefore focused on "skill-biased technological change", an explanation first put forward by Katz and Murphy (1992) and Berman, Bound and Griliches (1994) to explain what had happened: newly introduced computers and ICT technologies were more useful for skilled workers, and therefore firms increased their demand for them.

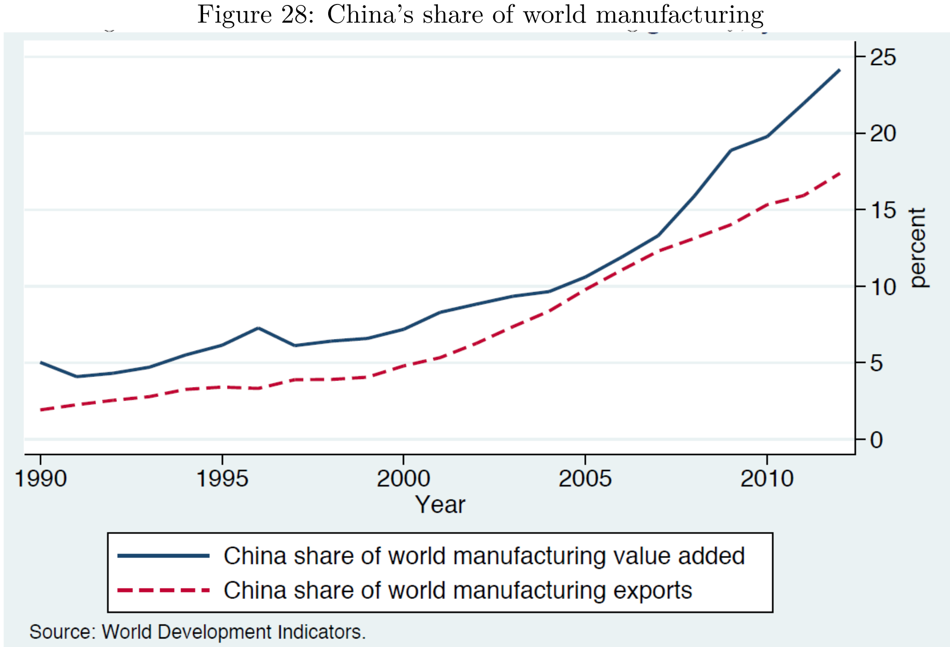
However, this consensus has shifted. In 2007, Krugman wrote that "*it's no longer safe to assert, as we could a dozen years ago, that the effects of trade on income distribution in wealthy countries are fairly minor. There's now a good case that they are quite big, and getting bigger.*"[108] Indeed, during the late 1990s and the 2000s, the patterns of world trade changed rapidly, marked by two (overlapping) developments: the rise of offshoring and the rise of China. These changes had important effects on wages, employment and inequality (see Helpman, 2018, for a book-length overview on these issues).

3.2 The "China shock"

3.2.1 China's rise

While economic reforms (such as the creation of Special Economic Zones) had started already in the 1980s, China's opening to trade (and transition to a market economy) became definitive and massive only in the 1990s.[109] Since then, China has been growing at extraordinary high rates, especially in the manufacturing sector. As Figure 28 shows, China's share in world manufacturing exports increased from 2% in 1991 to 19% in 2012.

Figure 28: China's share of world manufacturing



Source: Autor, Dorn and Hanson (2016, p. 5).

In a series of influential papers, Autor, Dorn and Hanson have analysed the effects of Chinese import

[108] Krugman's arguments can be found at http://www.voxeu.org/article/trade-and-inequality-revisited.

[109] For an overview of China's economic history in these years, see Autor, Dorn and Hanson (2016b) and Storesletten and Zilibotti (2014).

competition on the United States. They argue that the rise of China is not only important because of the country's size, but also because it constitutes an exogenous shock to the world economy, as that China's rise was arguably not triggered by any economic developments in developed countries.

3.2.2 Empirical strategy

Autor, Dorn and Hanson study the effect of Chinese import competition on different regions in the United States, by comparing regions that were more or less exposed to Chinese imports.[110]

Before looking at their results, it is worth noting that if there were a national US labour market, and American workers of a given type were homogeneous (as we implictly assumed in our models), there should be no regional differences in wages. For instance, a unique national labour market for low-skilled work implies that there should be a unique low-skilled wage, earned by all low-skilled workers, no matter where they live or who their employer is. Thus, there should not be any wage divergence between regions that are more or less affected by Chinese import competition. On the other hand, there may be movements in employment or population, as people move out of industries and/or regions affected by import competition, and into export industries.

To identify the causal effect of Chinese imports on regional outcomes, two challenges need to be overcome.

Challenge 1 The amount of Chinese imports in a region is not exogenous, but potentially correlated with many other developments which can also affect wages and employment. For instance, consider a region in which employment is dominated by one large firm. If the production facilities of this firm get destroyed by a hurricane, production capacity goes down and Chinese imports may go up, but the negative effects on employment are due to the hurricane rather than to Chinese competition.

To address this problem, Autor, Dorn and Hanson need to measure a region's exposure to Chinese competition in a way that is uncorrelated to other determinants of employment and wages. They propose to use for this a region's industry specialization in the year 1990. Precisely, the explanatory variable used in their regressions, which can be seen as an instrumental variable for Chinese imports to region j, is

$$x_j = \sum_{k=1}^{K} s_{kj}^{1990} \left(m_k^{2007} - m_k^{1990} \right) \tag{133}$$

where s_{kj}^{1990} is the share of total employment of region j represented by industry k in 1990, and m_k^{1990}

[110] Precisely, they consider "Commuting Zones", which are larger than counties, but smaller than states.

are total Chinese imports of the United States in industry k in 1990. Thus, regions affected by the China shock are regions which were specialized in 1990 in industries in which China started exporting a lot after 1990. Clearly, x_j is positively correlated with Chinese imports to region j. However, it is also exogenous to regional economic developments after 1990s (as the size of the shock is measured by national imports, not by regional ones). Furthermore, the variable is also exogenous to the China shock itself, as industry specialization was determined before it started (and the rise of China came as a surprise to most observers). As the industrial specialization of different regions in the United States is substantially different, there is enough variation in the measure x_j to run meaningful regressions.[111]

Problem 2 Using the instrumental variable x_j does not solve all problems, however. Most importantly, there could still be industry-level shocks that are correlated with Chinese imports. For instance, imagine that a new discovery in the car industry reduced the employment of US car producers, but increases their intermediate input imports from China. A regression of employment on x_j would then yield a significant negative coefficient, as the regions which are most specialized in cars are affected most by the shock. However, this reaction would in reality not be due to Chinese imports, but to the technology shock to the car industry.

Autor, Dorn and Hanson deal with this problem doing two different things. First, in Equation (133), they do not use Chinese exports to the United States, but Chinese exports to other developed economies (such as Germany, Japan and Spain). Indeed, the latter should not be correlated to any US-specific industry-level shocks. However, this alone is not enough, as some industry-level shocks (such as, most importantly, the rise of computers and ICT) affect other developed economies just like the United States. Therefore, the authors also explicitly control for these technological changes, by calculating the exposition of each industry to ICT. They find that exposition to ICT and exposition to Chinese imports are almost completely uncorrelated, and thus, that ICT-related industry-level shocks should not be a problem for their estimation.

3.2.3 Results

Employment In a first paper in 2013, Autor, Dorn and Hanson analysed the effects of Chinese import competition on employment and wages on the regional level, by regressing these outcomes on the explanatory variable x_j, and a line of control variables (Autor, Dorn and Hanson, 2013).

[111] This methodology is often used in applied work. An instrumental variable such as x_j is generally called a Bartik instrument, as it was first introduced by Bartik (1991).

Table 13: Chinese Import Competition, Employment and Wages

A. Δ Fraction of Working Age Population in Manufacturing, Unemployment, NILF

Employed in Manufacturing (1)	Employed in Non-Manufacturing (2)	Unemployed (3)	Not in Labor Force (4)
-0.60***	-0.18	0.22***	0.55***
(0.10)	(0.14)	(0.06)	(0.15)

B. Δ Log Population, Log Wages, Annual Wage and Transfer Income

Δ Log CZ Population (log pts) (5)	Δ Avg Log Weekly Wage (log pts) (6)	Δ Annual Wage/Salary Inc per Adult (US$) (7)	Δ Transfers per Capita (US$) (8)
-0.05	-0.76***	-549.3***	57.7***
(0.75)	(0.25)	(169.4)	(18.4)

N=1444 (722 commuting zones x 2 time periods 1990-2000 and 2000-2007). Employment, population and income data is based on U.S. Census and American Community Survey data, while transfer payments are based on BEA Regional Economic Accounts. All regressions control for the start of period percentage of employment in manufacturing, college-educated population, foreign-born population, employment among women, employment in routine occupations, average offshorability index of occupations, and Census division and time dummies. Models are weighted by start of period commuting zone share of national population. Robust standard errors in parentheses are clustered on state. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 13 shows that Chinese import competition had large negative impacts on the most exposed regions. In particular, an increase of 1000$ in the value of Chinese import competition per production worker reduces the fraction of the working age population active in manufacturing by 0.6 percentage points, and does not lead to an increase in non-manufacturing employment. Instead, it increases the number of unemployed and of persons out of the labour force. Furthermore, there is no evidence of outmigration (the total size of the population (Column (5)) does not react significantly). This result is therefore strongly opposed to the idea that workers in the United States are highly mobile. Pierce and Schott (2016), using a different empirical strategy,[112] also find that US manufacturing employment was strongly negatively affected by Chinese imports.

Why has there been this strongly negative effect on employment? Autor, Dorn and Hanson have argued that an important reason may be the large trade deficit of the US with respect to China. Indeed, in all the models we have seen, trade was balanced, so that imports from China would also give rise to US exports. In reality, the large trade deficit implies that losses in importing industries were not much compensated by gains in exporting industries.[113] This explanation is consistent with the fact that in economies which have not had a current account deficit with China, such as Germany, the employment

[112] Pierce and Schott use the fact that China's accession to the WTO in 2001 lead to the granting of "Permanently Normal Trade Relations" with the United States. Prior to 2001, the United States already had quite low tariffs on Chinese imports, but the US Congress had to decide every year on whether to maintain these or raise them to a high level fixed by the 1930 Smoot-Hawley law. Pierce and Schott argue that this created a lot of uncertainty which was holding back Chinese imports. Once this menace was eliminated, they show that in industries in which the Smoot-Hawley tariff was highest, Chinese imports increased most and US manufacturing employment fell most.

[113] Why does China have such a large trade surplus with the United States? There are different explanations for this fact, one being that underdeveloped financial markets in China incentivise Chinese investors to invest in the United States. Thus, China has a financial account deficit, whose mirror image is a current account surplus (see Caballero, Farhi and Gourinchas, 2008).

effects of Chinese import competition seem to have been much smaller, as has been shown by Dauth, Findeisen and Suedekum (2014).[114] It may also indicate that the employment effects could vanish in the future, as trade imbalances between the United States and China are reduced.

Wages Table 13 also shows that in regions in which Chinese import competition increased most, wages fell substantially: a 1000$ increase in Chinese import competition triggered a 550$ fall in regional wages. Note, however, that - just as for employment - these results apply to regional differences in wages, not to the absolute level of wages. That is, the correct way to read them is that if region A had a 1000$ higher increase in Chinese imports than region B, wages in region A increased by 550$ less (or fell by 550$ more) than in region B. However, these regressions do not tell us the causal effect of Chinese imports on the level of wages in region A or B.

Thus, the Autor-Dorn-Hanson results on their own do not imply that the United States as a whole lost from trade with China (and likewise, they of course do not imply that the United States as a whole gained from trade with China). What they do show - and this is already important on its own - is that the gains (or losses) from trade with China were unequally distributed across the United States, with some regions gaining much more (or losing much less) than others. In theory, the US government could have dampened these inequalities by redistributing resources between regions. However, this did not happen in practice. As column (8) in Table 13 shows, the regions most exposed to Chinese imports did see an increase in government transfers, but this increase only made up for around 10% of their relative wage losses.

Autor, Dorn, Hanson and Song (2014) take this analysis to an even more disaggregated level, looking at individual workers. Precisely, they compare workers which are equal with respect to all observable characteristics (age, experience, gender, race...), but differ by the industry in which they worked in 1991. They show that workers who worked in an industry which experienced more Chinese import competition after 1991 accumulated lower earnings for the period 1992-2007, changed jobs more frequently, and were more likely to drop out of the labour force and receive disability insurance. These effects are entirely driven by workers with initially low wages, why high-wage workers to not seem to be negatively affected.

Politics Finally, Autor, Dorn, Hanson and Majlesi (2016c) look at political outcomes, finding that an increase in Chinese import penetration also leads to more polarized political outcomes, with moderate

[114] For Italy, instead, Federico (2014) argues that the employment effects of Chinese competition were quite large.

politicians losing votes. In a note to this paper, they specifically look at the 2016 US presidential election.[115] They find that there is "*a robust positive effect of rising import competition on Republican vote share gains. The magnitude of the Republican gains is non-trivial. A counterfactual study of closely contested states suggests that Michigan, Wisconsin, Pennsylvania and North Carolina would have elected the Democrat instead of the Republican candidate if, ceteris paribus, the growth in Chinese import penetration had been 50 percent lower than the actual growth during the period of analysis. The Democrat candidate would also have obtained a majority in the electoral college in this counterfactual scenario.*"

Cross-sectional estimates and aggregate effects The studies discussed so far have many advantages. Most importantly, the use of disaggregated (regional, firm-level or even worker-level) data makes them more precise and allows them to use more convincing econometric strategies than studies relying on aggregate data.

However, this precision comes at a price: by comparing different regions or different workers, they only get at relative changes, whereas we would be really interested in absolute ones. Let us return to the two-region example introduced above. Region A had a 1000$ higher increase in Chinese imports than region B, and this caused a 550$ wage differential between the two regions. However, we cannot infer the effect of Chinese imports on the level of wages in region A from this result. Indeed, suppose the true model relating changes in wages to changes in Chinese imports in region i is

$$\Delta w_i = \alpha_1 + \alpha_2 + \beta \Delta m_i, \tag{134}$$

where Δm_i is the change in regional Chinese imports. That is, there is some common effect of Chinese imports in all regions (e.g., through changes in the prices of nationally traded goods, or through the national component of labour markets), affecting each region irrespective of the level of regional imports, and captured by α_1. However, there are also some other effects that are common through all regions (changes in monetary policy, for instance), and captured by α_2. On the other hand, regional wage changes also depend on regional imports, as captured by the slope coefficient β.

The key thing to understand is that when we estimate the cross-sectional regressions described above, we can only identify the slope β, but not the intercept α_1. More precisely, a regression software will of course give us an estimate of $\alpha_1 + \alpha_2$, but we cannot tell apart α_1 and α_2. In fact, the greatest strength

[115] The note is available at http://economics.mit.edu/files/12418.

of our empirical strategy is also its greatest weakness: we look at differences between regions precisely in order to difference out common factors that we are not interested in (such as monetary policy). But by doing this, we also difference out a common factor that we are interested in, the common effect of Chinese imports on all regions.

How can one overcome this “missing intercept” problem (Wolf, 2019)? There are essentially two possibilities. One is to bring in outside information to tell us what the intercept is, typically by assuming that for one category considered, the overall effect is zero (this is probably not a good assumption for regional outcomes, but can be a bit more defensible in other settings). Another possibility, more commonly used, is to use a model to infer the intercept. This is conceptually cleaner, but of course relies on the model being a good description of reality. One example of these approaches is Acemoglu, Autor, Dorn, Hanson and Price (2016) who estimate that for the period between 1999 and 2011, Chinese import competition is responsible for between 0.5 and 1 million lost jobs in US manufacturing, and up to 2 million jobs in the entire US economy. However, these results have been contested (Helpman, 2018), and the above discussion makes clear that they necessarily rely on more assumptions than the findings about inequalities between regions and people.

Consumer goods prices As mentioned above, one important consequence of Chinese imports which - a priori - should affect most US regions roughly equally are changes in the prices of consumer goods. In a recent paper, Jaravel and Sager (2019) try to quantify the effect of Chinese imports on US prices, using empirical strategies similar to the ones outlined above. They find a strong response of prices to Chinese import competition, and therefore large gains from trade: their headline estimate is that “*trade with China increased U.S. consumer surplus by about $400,000 per displaced job*”. If these estimates are credible, they suggest that it would be possible to compensate the losers from trade.

3.3 Offshoring

The evidence provided by Autor, Dorn and Hanson relates both to “classical” trade and to offshoring: the increase in Chinese imports which they document both comes from Chinese firms and China-based subsidiaries of US firms.

A couple of other studies have looked at the effect of offshoring more directly. For instance, Feenstra and Hanson (1996b) and Feenstra and Hanson (1997) tried to measure the incidence of outsourcing at the industry level for the United States by calculating the share of imported intermediate inputs in total intermediate inputs.[116] They have shown that increases in this share are positively correlated, over the period 1979-1990, with increases in the share of skilled workers in the total wage bill of the industry. However, the effects they found were still relatively small, and the consensus was that technological change was more important to explain increasing inequality.

In recent years, however, more detailed studies have found important effects for offshoring, too. Hummels, Jorgensen, Munch and Xiang (2014) construct instrumental variables for "offshoring shocks". Using data on the value and the source country of all products imported by Danish firms, they note that many imported products are highly specific (that is, there are only very few firms that import them). Thus, they consider a fall in the price of one of these particular good to be an offshoring shock for the Danish firm importing that product. They show that Danish firms hit by such a shock see a decrease in the average wage of low-skilled workers, and an increase in the average wage of high-skilled workers goes up. In a related paper, Hummels, Munch, Skipper and Xiang (2012) also show that workers which lose their job after such an offshoring shock have a harder time of finding a new job than workers which lose their job in other circumstances.

3.4 Taking stock

The results of the last decade of empirical studies have revealed some challenges, especially for the classic Heckscher-Ohlin theory. While their broad conclusions are very much in line with the theory (the rise of trade with developing countries appears to have been detrimental to low-skilled workers in developed countries), the more detailed aspects are not. Instead of a broad decline in low-skilled wages throughout the economy, we have seen concentrated (relative) losses for directly affected industries and directly affected workers. This suggests that production factors seem to be more differentiated than previously thought: even workers with similar education levels may have skills that are very specific to their job and employer. As these skills are not so valuable in another job or with another employer, mobility is much lower than in the ideal Heckscher-Ohlin environment, and losses from import competition are more direct.[117] Overall, the empirical evidence points to substantial distributional effects

[116] Note that they did not make a distinction whether these inputs were produced by subsidiaries of the US firms, or by independent producers.

[117] As a result of these developments, researchers have rediscovered an old model of trade, the specific factors model developed by Samuelson (1971) and Jones (1971). This model is essentially a modified version of the Heckscher-Ohlin model which we have seen in this class, with the particularity that some production factor can only work in certain industries (i.e., they are "specific" to these industries). The model predicts then that increasing imports in one industry lead to a decrease in the real compensation of the specific factors of that industry, while increasing the real compensation of the specific factors of other industries. For modern versions of the specific factor framework, see Kovak (2013), or Galle, Rodriguez-Clare and Yi (2017).

from trade, which the policies enacted in the past have not been able to attenuate.[118]

[118] We have mostly focused on the effects of trade on inequality in developed countries. For a summary of the research on trade and inequality in developing countries, see Goldberg (2015).

References

Acemoglu, D. (2009). *Introduction to Modern Economic Growth*. Princeton University Press.

—, Autor, D., Dorn, D., Hanson, G. H. and Price, B. (2016). Import Competition and the Great US Employment Sag of the 2000s. *Journal of Labor Economics*, **34** (S1), S141 – S198.

—, Johnson, S. and Robinson, J. A. (2001). The Colonial Origins of Comparative Development: An Empirical Investigation. *American Economic Review*, **91** (5), 1369–1401.

— and Restrepo, P. (2020). Robots and Jobs: Evidence from US Labor Markets. *Journal of Political Economy*.

Aghion, P., Bergeaud, A., Lequien, M. and Melitz, M. (2017). The impact of exports on innovation: Theory and evidence.

—, Bloom, N., Blundell, R., Griffith, R. and Howitt, P. (2005). Competition and Innovation: An Inverted-U Relationship. *The Quarterly Journal of Economics*, **120** (2), 701–728.

— and Howitt, P. (1992). A model of growth through creative destruction. *Econometrica*, **60** (2), 323–351.

Allcott, H. and Keniston, D. (2017). Dutch Disease or Agglomeration? The Local Economic Effects of Natural Resource Booms in Modern America. *The Review of Economic Studies*, p. rdx042.

Anderson, J. E. (1979). A Theoretical Foundation for the Gravity Equation. *The American Economic Review*, **69** (1), 106–116.

— and van Wincoop, E. (2003). Gravity with Gravitas: A Solution to the Border Puzzle. *American Economic Review*, **93** (1), 170–192.

Antràs, P. (2015). *Global Production: Firms, Contracts, and Trade Structure*. Princeton University Press.

Antràs, P. and de Gortari, A. (2020). On the geography of global value chains. *Econometrica*, **84** (4), 1553–1598.

Antràs, P. and Yeaple, S. R. (2014). Multinational Firms and the Structure of International Trade. In E. H. Gita Gopinath and K. Rogoff (eds.), *Handbook of International Economics*, *Handbook of International Economics*, vol. 4, Elsevier, pp. 55 – 130.

Arkolakis, C., Costinot, A., Donaldson, D. and Rodríguez-Clare, A. (2018). The Elusive Pro-Competitive Effects of Trade. *The Review of Economic Studies*, **86** (1), 46–80.

—, — and Rodriguez-Clare, A. (2012). New Trade Models, Same Old Gains? *American Economic Review*, **102** (1), 94–130.

Armington, P. S. (1969). A theory of demand for products distinguished by place of production. *Staff Papers (International Monetary Fund)*, **16** (1), 159–178.

Atkin, D., Khandelwal, A. K. and Osman, A. (2017). Exporting and Firm Performance: Evidence from a Randomized Experiment. *The Quarterly Journal of Economics*, **132** (2), 551–615.

Autor, D., Dorn, D., Hanson, G., Pisano, G. and Shu, P. (2016a). *Foreign Competition and Domestic Innovation: Evidence from U.S. Patents*. CEPR Discussion Papers 11664, C.E.P.R. Discussion Papers.

Autor, D. H., Dorn, D. and Hanson, G. H. (2013). The China Syndrome: Local Labor Market Effects of Import Competition in the United States. *American Economic Review*, **103** (6), 2121–68.

—, — and — (2016b). *The China Shock: Learning from Labor Market Adjustment to Large Changes in Trade*. NBER Working Papers 21906, National Bureau of Economic Research, Inc.

—, —, — and Majlesi, K. (2016c). *Importing Political Polarization? The Electoral Consequences of Rising Trade Exposure*. NBER Working Papers 22637, National Bureau of Economic Research, Inc.

—, —, — and Song, J. (2014). Trade Adjustment: Worker-Level Evidence. *The Quarterly Journal of Economics*, **129** (4), 1799–1860.

Aw, B. Y., Roberts, M. J. and Xu, D. Y. (2011). R&D Investment, Exporting, and Productivity Dynamics. *American Economic Review*, **101** (4), 1312–1344.

Axtell, R. L. (2001). Zipf Distribution of U.S. Firm Sizes. *Science*, **293** (5536), 1818–1820.

Baier, S. L. and Bergstrand, J. H. (2007). Do free trade agreements actually increase members' international trade? *Journal of International Economics*, **71** (1), 72–95.

—, —, Egger, P. and McLaughlin, P. A. (2008). Do Economic Integration Agreements Actually Work? Issues in Understanding the Causes and Consequences of the Growth of Regionalism. *The World Economy*, **31** (4), 461–497.

Balassa, B. (1963). An Empirical Demonstration of Classical Comparative Cost Theory. *The Review of Economics and Statistics*, **45** (3), pp. 231–238.

— (1966). Tariff Reductions and Trade in Manufacturers among the Industrial Countries. *The American Economic Review*, **56** (3), pp. 466–473.

Baldwin, R. E. and Robert-Nicoud, F. (2008). Trade and growth with heterogeneous firms. *Journal of International Economics*, **74** (1), 21–34.

Barjamovic, G., Chaney, T., Coşar, K. and Hortaçsu, A. (2019). Trade, Merchants, and the Lost Cities of the Bronze Age. *The Quarterly Journal of Economics*, **134** (3), 1455–1503.

Bartelsman, E. J. and Doms, M. (2000). Understanding Productivity: Lessons from Longitudinal Microdata. *Journal of Economic Literature*, **38** (3), 569–594.

Bartik, T. J. (1991). *Who Benefits from State and Local Economic Development Policies?* No. wbsle in Books from Upjohn Press, W.E. Upjohn Institute for Employment Research.

Baumol, W. J. (1967). Macroeconomics of unbalanced growth: The anatomy of urban crisis. *The American Economic Review*, **57** (3), pp. 415–426.

Berman, E., Bound, J. and Griliches, Z. (1994). Changes in the Demand for Skilled Labor within U. S. Manufacturing: Evidence from the Annual Survey of Manufactures. *The Quarterly Journal of Economics*, **109** (2), 367–397.

Bernard, A. B. and Jensen, J. B. (1999). Exceptional exporter performance: cause, effect, or both? *Journal of International Economics*, **47** (1), 1–25.

—, —, Redding, S. J. and Schott, P. K. (2018). Global Firms. *Journal of Economic Literature*, **56** (2), 565–619.

—, — and Schott, P. K. (2006). Trade costs, firms and productivity. *Journal of Monetary Economics*, **53** (5), 917–937.

—, — and — (2009). Importers, Exporters and Multinationals: A Portrait of Firms in the U.S. that Trade Goods. In *Producer Dynamics: New Evidence from Micro Data*, NBER Chapters, National Bureau of Economic Research, Inc, pp. 513–552.

—, Redding, S. J. and Schott, P. K. (2007). Comparative Advantage and Heterogeneous Firms. *Review of Economic Studies*, **74** (1), 31–66.

Bloom, N., Draca, M. and Van Reenen, J. (2016). Trade Induced Technical Change? The Impact of Chinese Imports on Innovation, IT and Productivity. *Review of Economic Studies*, **83** (1), 87–117.

—, Romer, P. M., Terry, S. J. and Reenen, J. V. (2014). *Trapped Factors and China's Impact on Global Growth*. NBER Working Paper 19951,.

Borjas, G. J. (2017). The wage impact of the marielitos: A reappraisal. *ILR Review*, **70** (5), 1077–1110.

Bowen, H., Leamer, E. and Sveikauskas, L. (1987). Multicountry, Multifactor Tests of the Factor Abundance Theory. *American Economic Review*, **77** (5), 791–809.

Brainard, S. L. (1993). *A Simple Theory of Multinational Corporations and Trade with a Trade-Off Between Proximity and Concentration*. NBER Working Papers 4269, National Bureau of Economic Research, Inc.

— (1997). An empirical assessment of the proximity-concentration trade-off between multinational sales and trade. *The American Economic Review*, **87** (4), 520–544.

Brander, J. A. (1995). Strategic Trade Policy. *Handbook of International Economics*, vol. 3, Elsevier, pp. 1395 – 1455.

Brandt, L., Biesebroeck, J. V., Wang, L. and Zhang, Y. (2017). WTO Accession and Performance of Chinese Manufacturing Firms. *American Economic Review*, **107** (9), 2784–2820.

Broda, C. and Weinstein, D. E. (2006). Globalization and the Gains From Variety. *The Quarterly Journal of Economics*, **121** (2), 541–585.

Bustos, P. (2011). Trade Liberalization, Exports, and Technology Upgrading: Evidence on the Impact of MERCOSUR on Argentinian Firms. *American Economic Review*, **101** (1), 304–40.

Caballero, R. J., Farhi, E. and Gourinchas, P.-O. (2008). An Equilibrium Model of "Global Imbalances" and Low Interest Rates. *American Economic Review*, **98** (1), 358–93.

Caliendo, L. and Parro, F. (2015). Estimates of the Trade and Welfare Effects of NAFTA. *The Review of Economic Studies*, **82** (1), 1–44.

Card, D. (1990). The impact of the Mariel boatlift on the Miami labor market. *Industrial and Labor Relations Review*, **43** (2), 245–257.

Cavenaile, L., Roldan-Blanco, P. and Schmitz, T. (2020). International Trade and Innovation Dynamics with Endogenous Markups. *Mimeo*.

Caves, R. E. (2007). *Multinational Enterprise and Economic Analysis*. Cambridge University Press.

Chaney, T. (2008). Distorted Gravity: The Intensive and Extensive Margins of International Trade. *American Economic Review*, **98** (4), 1707–21.

Clerides, S. K., Lach, S. and Tybout, J. R. (1998). Is Learning By Exporting Important? Micro-Dynamic Evidence From Colombia, Mexico, And Morocco. *The Quarterly Journal of Economics*, **113** (3), 903–947.

Coe, D. T. and Helpman, E. (1995). International R&D spillovers. *European Economic Review*, **39** (5), 859–887.

Comin, D. and Mestieri, M. (2013). *If Technology Has Arrived Everywhere, Why Has Income Diverged?* TSE Working Papers 13-409, Toulouse School of Economics (TSE).

Corden, W. M. and Neary, J. P. (1982). Booming Sector and De-Industrialisation in a Small Open Economy. *The Economic Journal*, **92** (368), pp. 825–848.

Correia, S. (2014). REGHDFE: Stata module to perform linear or instrumental-variable regression absorbing any number of high-dimensional fixed effects. Statistical Software Components, Boston College Department of Economics.

Costinot, A. (2009). An Elementary Theory of Comparative Advantage. *Econometrica*, **77** (4), 1165–1192.

— and Donaldson, D. (2012). Ricardo's Theory of Comparative Advantage: Old Idea, New Evidence. *American Economic Review*, **102** (3), 453–58.

—, — and Komunjer, I. (2012). What Goods Do Countries Trade? A Quantitative Exploration of Ricardo's Ideas. *Review of Economic Studies*, **79** (2), 581–608.

—, —, Kyle, M. and Williams, H. (2019). The More We Die, The More We Sell? A Simple Test of the Home-Market Effect. *The Quarterly Journal of Economics*, **134** (2), 843–894.

—, —, Vogel, J. and Werning, I. (2015). Comparative Advantage and Optimal Trade Policy. *The Quarterly Journal of Economics*, **130** (2), 659–702.

— and Rodriguez-Clare, A. (2014). *Trade Theory with Numbers: Quantifying the Consequences of Globalization*, Elsevier, *Handbook of International Economics*, vol. 4, chap. 0, pp. 197–261.

— and Vogel, J. (2010). Matching and Inequality in the World Economy. *Journal of Political Economy*, **118** (4), 747–786.

Dauth, W., Findeisen, S. and Suedekum, J. (2014). The Rise Of The East And The Far East: German Labor Markets And Trade Integration. *Journal of the European Economic Association*, **12** (6), 1643–1675.

Davis, D. R. and Mishra, P. (2007). Stolper-Samuelson Is Dead: And Other Crimes of Both Theory and Data. In *Globalization and Poverty*, NBER Chapters, National Bureau of Economic Research, Inc, pp. 87–108.

— and Weinstein, D. E. (2001). An account of global factor trade. *American Economic Review*, **91** (5), 1423–1453.

de Gortari, A. (2018). Disentangling Global Value Chains. *Mimeo*.

De Loecker, J. (2011). Product Differentiation, Multiproduct Firms, and Estimating the Impact of Trade Liberalization on Productivity. *Econometrica*, **79** (5), 1407–1451.

— and Goldberg, P. K. (2014). Firm Performance in a Global Market. *Annual Review of Economics*, **6** (1), 201–227.

—, —, Khandelwal, A. K. and Pavcnik, N. (2016). Prices, markups, and trade reform. *Econometrica*, **84** (2), 445–510.

— and Warzynski, F. (2012). Markups and firm-level export status. *American Economic Review*, **102** (6), 2437–71.

Dhingra, S., Huang, H., Ottaviano, G., Paulo Pessoa, J., Sampson, T. and Van Reenen, J. (2017). The costs and benefits of leaving the EU: trade effects. *Economic Policy*, **32** (92), 651–705.

di Giovanni, J. and Levchenko, A. A. (2013). Firm entry, trade, and welfare in Zipf's world. *Journal of International Economics*, **89** (2), 283–296.

—, — and Zhang, J. (2014). The Global Welfare Impact of China: Trade Integration and Technological Change. *American Economic Journal: Macroeconomics*, **6** (3), 153–83.

Dixit, A. K. and Stiglitz, J. E. (1977). Monopolistic Competition and Optimum Product Diversity. *American Economic Review*, **67** (3), 297–308.

Dornbusch, R., Fischer, S. and Samuelson, P. A. (1977). Comparative Advantage, Trade, and Payments in a Ricardian Model with a Continuum of Goods. *American Economic Review*, **67** (5), 823–839.

—, — and — (1980). Heckscher-Ohlin Trade Theory with a Continuum of Goods. *The Quarterly Journal of Economics*, **95** (2), 203–224.

Dustmann, C., Ludsteck, J. and Schönberg, U. (2009). Revisiting the german wage structure. *The Quarterly Journal of Economics*, **124** (2), 843–881.

Eaton, J. and Kortum, S. (2002). Technology, Geography, and Trade. *Econometrica*, **70** (5), 1741–1779.

— and — (2012). Putting Ricardo to Work. *Journal of Economic Perspectives*, **26** (2), 65–90.

Fajgelbaum, P. D., Goldberg, P. K., Kennedy, P. J. and Khandelwal, A. K. (2019). The Return to Protectionism. *The Quarterly Journal of Economics*, **135** (1), 1–55.

Federico, S. (2014). Industry Dynamics and Competition from Low-Wage Countries: Evidence on Italy. *Oxford Bulletin of Economics and Statistics*, **76** (3), 389–410.

Feenstra, R. C. (2011). *Offshoring to China: The Local and Global Impacts of Processing Trade*. Open Economy Lectures at the Institute for Applied International Trade, University of International Business and Economics, Beijing.

— (2016). *Advanced International Trade: Theory and Evidence*. Princeton University Press.

— and Hanson, G. H. (1996a). Foreign Investment, Outsourcing and Relative Wages. In G. M. G. Robert C. Feenstra and D. A. Irwin (eds.), *Political Economy of Trade Policy: Essays in Honor of Jagdish Bhagwati*, MIT Press, pp. 89–128.

— and — (1996b). Globalization, Outsourcing, and Wage Inequality. *American Economic Review, Papers & Proceedings*, **86** (2), 240–45.

— and — (1997). Foreign direct investment and relative wages: Evidence from Mexico's maquiladoras. *Journal of International Economics*, **42** (3-4), 371–393.

— and Weinstein, D. E. (2017). Globalization, Markups, and US Welfare. *Journal of Political Economy*, **125** (4), 1040–1074.

Feyrer, J. (2009). *Distance, Trade, and Income - The 1967 to 1975 Closing of the Suez Canal as a Natural Experiment*. NBER Working Papers 15557, National Bureau of Economic Research, Inc.

— (2019). Trade and Income - Exploiting Time Series in Geography. *American Economic Journal: Applied Economics*, **11** (4), 1–35.

Foster, L., Haltiwanger, J. and Syverson, C. (2016). The Slow Growth of New Plants: Learning about Demand? *Economica*, **83** (329), 91–129.

Gabaix, X. (2009). Power Laws in Economics and Finance. *Annual Review of Economics*, **1** (1), 255–294.

— (2011). The Granular Origins of Aggregate Fluctuations. *Econometrica*, **79** (3), 733–772.

Galle, S., Rodriguez-Clare, A. and Yi, M. (2017). *Slicing the Pie: Quantifying the Aggregate and Distributional Effects of Trade*. NBER Working Papers 23737, National Bureau of Economic Research, Inc.

Ghironi, F. and Melitz, M. J. (2005). International Trade and Macroeconomic Dynamics with Heterogeneous Firms. *The Quarterly Journal of Economics*, **120** (3), 865–915.

Goldberg, P. K. (ed.) (2015). *Trade and Inequality*. No. 15997 in Books, Edward Elgar Publishing.

Grossman, G. M. and Helpman, E. (1991). *Innovation and Growth in the Global Economy*. MIT Press.

— and — (2015). Globalization and Growth. *American Economic Review, Papers & Proceedings*, **105** (5), 100–104.

— and Rossi-Hansberg, E. (2006). The rise of offshoring: it's not wine for cloth anymore. *Proceedings - Economic Policy Symposium - Jackson Hole*, pp. 59–102.

— and — (2008). Trading tasks: A simple theory of offshoring. *American Economic Review*, **98** (5), 1978–97.

Hanson, G. H. and Slaughter, M. J. (2002). Labor-market adjustment in open economies: Evidence from US states. *Journal of International Economics*, **57** (1), 3–29.

Harrigan, J. (1996). Openness to trade in manufactures in the OECD. *Journal of International Economics*, **40** (1), 23 – 39.

Head, K. and Mayer, T. (2014). Gravity equations: Workhorse,toolkit, and cookbook. In E. H. Gita Gopinath and K. Rogoff (eds.), *Handbook of International Economics*, *Handbook of International Economics*, vol. 4, Elsevier, pp. 131 – 195.

— and Ries, J. (1999). Rationalization effects of tariff reductions. *Journal of International Economics*, **47** (2), 295–320.

Heckscher, E. (1919). *The Effect of Foreign Trade on the Distribution of Income*.

Helpman, E. (1987). Imperfect competition and international trade: Evidence from fourteen industrial countries. *Journal of the Japanese and International Economies*, **1** (1), 62–81.

— (2018). *Globalization and Inequality*. Harvard University Press.

—, Itskhoki, O. and Redding, S. (2010). Inequality and Unemployment in a Global Economy. *Econometrica*, **78** (4), 1239–1283.

— and Krugman, P. R. (1985). *Market Structure and Foreign Trade: Increasing Returns, Imperfect Competition, and the International Economy*, *MIT Press Books*, vol. 1. The MIT Press.

—, Melitz, M. J. and Yeaple, S. R. (2004). Export Versus FDI with Heterogeneous Firms. *American Economic Review*, **94** (1), 300–316.

Hsieh, C.-T. and Klenow, P. J. (2009). Misallocation and Manufacturing TFP in China and India. *The Quarterly Journal of Economics*, **124** (4), 1403–1448.

Hummels, D., Jorgensen, R., Munch, J. and Xiang, C. (2014). The Wage Effects of Offshoring: Evidence from Danish Matched Worker-Firm Data. *American Economic Review*, **104** (6), 1597–1629.

—, Munch, J. R., Skipper, L. and Xiang, C. (2012). Offshoring, Transition, and Training: Evidence from Danish Matched Worker-Firm Data. *American Economic Review*, **102** (3), 424–28.

—, — and Xiang, C. (2016). *Offshoring and Labor Markets*. NBER Working Papers 22041, National Bureau of Economic Research, Inc.

Jaffe, A. B., Trajtenberg, M. and Henderson, R. (1993). Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations. *The Quarterly Journal of Economics*, **108** (3), 577–598.

Jaravel, X. and Sager, E. (2019). *What are the Price Effects of Trade? Evidence from the US and Implications for Quantitative Trade Models*. CEP Discussion Papers dp1642, Centre for Economic Performance, LSE.

Jones, C. I. (1995). R&D-Based Models of Economic Growth. *Journal of Political Economy*, **103** (4), 759–84.

— (2011). Intermediate goods and weak links in the theory of economic development. *American Economic Journal: Macroeconomics*, **3** (2), 1–28.

— (2016). The Facts of Economic Growth. *Handbook of Macroeconomics*, vol. 2, Elsevier, pp. 3 – 69.

Jones, R. W. (1971). A Three-Factor Model in Theory, Trade and History. In J. B. et al. (ed.), *Trade, Balance of Payments, and Growth*, *1*, Amsterdam: North-Holland, pp. 3–21.

Katz, L. F. and Murphy, K. M. (1992). Changes in Relative Wages, 1963-1987: Supply and Demand Factors. *The Quarterly Journal of Economics*, **107** (1), 35–78.

Kovak, B. K. (2013). Regional Effects of Trade Reform: What Is the Correct Measure of Liberalization? *American Economic Review*, **103** (5), 1960–76.

Krugman, P. R. (1979). Increasing Returns, Monopolistic Competition, and International Trade. *Journal of International Economics*, **9** (4), 469–479.

— (1980). Scale Economies, Product Differentiation, and the Pattern of Trade. *American Economic Review*, **70** (5), 950–59.

— (1987). The narrow moving band, the Dutch disease, and the competitive consequences of Mrs. Thatcher: Notes on trade in the presence of dynamic scale economies. *Journal of Development Economics*, **27** (1-2), 41–55.

— (1993). What Do Undergrads Need to Know about Trade? *American Economic Review*, **83** (2), 23–26.

— (1995). Growing World Trade: Causes and Consequences. *Brookings Papers on Economic Activity*, **26** (1, 25th A), 327–377.

— (1997). *Pop Internationalism*. MIT Press Books, The MIT Press.

— (1999). Was it all in Ohlin? In R. Findlay, L. Jonung and M. Lundahl (eds.), *Bertil Ohlin: A Centennial Celebration, 1899-1999*, *1*, The MIT Press, pp. 3–21.

— (2008). *The Increasing Returns Revolution in Trade and Geography*. Nobel prize lecture.

— and Venables, A. J. (1995). Globalization and the Inequality of Nations. *The Quarterly Journal of Economics*, **110** (4), 857–80.

Lancaster, K. J. (1979). *Variety, Equity and Efficiency*. Columbia University Press.

Leamer, E. E. (1980). The Leontief Paradox, Reconsidered. *Journal of Political Economy*, **88** (3), pp. 495–503.

Leibovici, F. and Santacreu, A. M. (2020). International Trade of Essential Goods during a Pandemic. *Mimeo*.

Leontief, W. (1953). Domestic Production and Foreign Trade; The American Capital Position Re-Examined. *Proceedings of the American Philosophical Society*, **97** (4), pp. 332–349.

Levchenko, A. A. (2007). Institutional Quality and International Trade. *Review of Economic Studies*, **74** (3), 791–819.

Lileeva, A. and Trefler, D. (2010). Improved Access to Foreign Markets Raises Plant-Level Productivity... for Some Plants. *The Quarterly Journal of Economics*, **125** (3), 1051–1099.

Linder, S. B. (1961). *An Essay on Trade and Transformation*. Almqvist and Wiksells.

MacDougall, G. D. A. (1951). British and American Exports: A Study Suggested by the Theory of Comparative Costs. Part I. *The Economic Journal*, **61** (244), pp. 697–724.

Manova, K. (2008). Credit constraints, equity market liberalizations and international trade. *Journal of International Economics*, **76** (1), 33–47.

Markusen, J. R. and Venables, A. J. (2000). The theory of endowment, intra-industry and multinational trade. *Journal of International Economics*, **52** (2), 209–234.

Mas-Colell, A., Whinston, M. D. and Green, J. R. (1995). *Microeconomic Theory*. No. 9780195102680 in OUP Catalogue, Oxford University Press.

Matsuyama, K. (2013). Endogenous Ranking and Equilibrium Lorenz Curve Across (ex ante) Identical Countries. *Econometrica*, **81** (5), 2009–2031.

McCallum, J. (1995). National Borders Matter: Canada-U.S. Regional Trade Patterns. *The American Economic Review*, **85** (3), 615–623.

Melitz, M. J. (2003). The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity. *Econometrica*, **71** (6), 1695–1725.

— and Ottaviano, G. I. P. (2008). Market Size, Trade, and Productivity. *Review of Economic Studies*, **75** (1), 295–316.

— and Redding, S. J. (2014). Heterogeneous Firms and Trade. In E. H. Gita Gopinath and K. Rogoff (eds.), *Handbook of International Economics*, *Handbook of International Economics*, vol. 4, Elsevier, pp. 1 – 54.

— and — (2015). New Trade Models, New Welfare Implications. *American Economic Review*, **105** (3), 1105–46.

Murata, Y., Nakajima, R., Okamoto, R. and Tamura, R. (2014). Localized Knowledge Spillovers and Patent Citations: A Distance-Based Approach. *The Review of Economics and Statistics*, **96** (5), 967–985.

Nakamura, E. and Steinsson, J. (2018). Identification in Macroeconomics. *Journal of Economic Perspectives*, **32** (3), 59–86.

Nunn, N. (2007). Relationship-Specificity, Incomplete Contracts, and the Pattern of Trade. *The Quarterly Journal of Economics*, **122** (2), 569–600.

— and Trefler, D. (2014). Domestic Institutions as a Source of Comparative Advantage. In E. H. Gita Gopinath and K. Rogoff (eds.), *Handbook of International Economics*, *Handbook of International Economics*, vol. 4, Elsevier, pp. 263 – 315.

Ohlin, B. (1924). *The Theory of Trade.*

— (1933). *Interregional and International Trade.* Harvard University Press.

Pascali, L. (2017). The wind of change: Maritime technology, trade, and economic development. *American Economic Review*, **107** (9), 2821–54.

Pavcnik, N. (2002). Trade Liberalization, Exit, and Productivity Improvement: Evidence from Chilean Plants. *Review of Economic Studies*, **69** (1), 245–76.

Pierce, J. R. and Schott, P. K. (2016). The Surprisingly Swift Decline of US Manufacturing Employment. *American Economic Review*, **106** (7), 1632–62.

Piketty, T. (2014). *Capital in the 21st Century.* Harvard University Press.

— and Saez, E. (2006). The Evolution of Top Incomes: A Historical and International Perspective. *American Economic Review*, **96** (2), 200–205.

Rajan, R. G. and Zingales, L. (1998). Financial Dependence and Growth. *American Economic Review*, **88** (3), 559–86.

Ricardo, D. (1817). *On the Principles of Political Economy and Taxation.*

Rivera-Batiz, L. A. and Romer, P. M. (1991). Economic Integration and Endogenous Growth. *The Quarterly Journal of Economics*, **106** (2), 531–55.

Rodriguez, F. and Rodrik, D. (2001). Trade Policy and Economic Growth: A Skeptic's Guide to the Cross-National Evidence. In *NBER Macroeconomics Annual 2000, Volume 15*, NBER Chapters, National Bureau of Economic Research, Inc, pp. 261–338.

Romalis, J. (2004). Factor Proportions and the Structure of Commodity Trade. *American Economic Review*, **94** (1), 67–97.

— (2007). NAFTA's and CUSFTA's Impact on International Trade. *The Review of Economics and Statistics*, **89** (3), 416–435.

Romer, D. H. and Frankel, J. A. (1999). Does Trade Cause Growth? *American Economic Review*, **89** (3), 379–399.

Romer, P. M. (1990). Endogenous Technological Change. *Journal of Political Economy*, **98** (5), S71–102.

Rybczynski, T. M. (1955). Factor Endowment and Relative Commodity Prices. *Economica*, **22** (88), pp. 336–341.

Sampson, T. (2016). Dynamic Selection: An Idea Flows Theory of Entry, Trade, and Growth. *The Quarterly Journal of Economics*, **131** (1), 315–380.

Samuelson, P. A. (1949). International factor-price equalisation once again. *The Economic Journal*, **59** (234), 181–197.

— (1954). The transfer problem and transport costs, ii: Analysis of effects of trade impediments. *The Economic Journal*, **64** (254), pp. 264–289.

— (1971). Ohlin was right. *The Swedish Journal of Economics*, **73** (4), 365–384.

— (2004). Where Ricardo and Mill Rebut and Confirm Arguments of Mainstream Economists Supporting Globalization. *Journal of Economic Perspectives*, **18** (3), 135–146.

Schivardi, F. and Schmitz, T. (2019). The IT Revolution and Southern Europeâs Two Lost Decades. *Journal of the European Economic Association*, jvz048.

Sforza, A. and Steininger, M. (2020). *Globalization in the Time of COVID-19*. Tech. rep.

Shu, P. and Steinwender, C. (2018). The Impact of Trade Liberalization on Firm Productivity and Innovation. In *Innovation Policy and the Economy, Volume 19*, NBER Chapters, National Bureau of Economic Research, Inc.

Simonovska, I. and Waugh, M. E. (2014). The elasticity of trade: Estimates and evidence. *Journal of International Economics*, **92** (1), 34–50.

Smith, A. (1776). *An Inquiry into the Nature and the Causes of the Wealth of Nations*.

Solow, R. M. (1956). A Contribution to the Theory of Economic Growth. *The Quarterly Journal of Economics*, **70** (1), 65–94.

Stolper, W. F. and Samuelson, P. A. (1941). Protection and Real Wages. *The Review of Economic Studies*, **9** (1), 58–73.

Storesletten, K. and Zilibotti, F. (2014). China's Great Convergence and Beyond. *Annual Review of Economics*, **6** (1), 333–362.

Syverson, C. (2011). What Determines Productivity? *Journal of Economic Literature*, **49** (2), 326–365.

Tabellini, G. (2008). Institutions and culture. *Journal of the European Economic Association*, **6** (2-3), 255–294.

Tinbergen, J. (1962). *Shaping the World Economy: Suggestions for an International Economic Policy*. The Twentieth Century Fund.

Trefler, D. (1993). International Factor Price Differences: Leontief Was Right! *Journal of Political Economy*, **101** (6), 961–87.

— (1995). The Case of the Missing Trade and Other Mysteries. *American Economic Review*, **85** (5), 1029–46.

— (2004). The Long and Short of the Canada-U. S. Free Trade Agreement. *American Economic Review*, **94** (4), 870–895.

Tybout, J. R. and Westbrook, M. D. (1995). Trade liberalization and the dimensions of efficiency change in Mexican manufacturing industries. *Journal of International Economics*, **39** (1-2), 53–78.

Vanek, J. (1968). The Factor Proportions Theory: the N Factor Case. *Kyklos*, **21** (4), 749–756.

Ventura, J. (1997). Growth and Interdependence. *The Quarterly Journal of Economics*, **112** (1), 57–84.

— (2005). A Global View of Economic Growth. In P. Aghion and N. S. Durlauf (eds.), *Handbook of Economic Growth*, *Handbook of Economic Growth*, vol. 1, *22*, Elsevier, pp. 1419–1497.

Wolf, C. K. (2019). The Missing Intercept: A Demand Equivalence Approach. *Mimeo*.

WTO (2015). *International Trade Statistics 2015*. WTO, International Trade Statistics Section.

Yeaple, S. R. (2013). The Multinational Firm. *Annual Review of Economics*, **5** (1), 193–217.

Part VIII

Additional details for Ricardian models

1 Additional derivations for the Eaton-Kortum model

1.1 The ideal price index (Cobb-Douglas case)

This section shows that the ideal Cobb-Douglas price index P_n in the simple Eaton-Kortum model, defined by $P_n = \exp\left(\int\limits_0^1 \ln p_n(i) di\right)$, is equal in equilibrium to $\frac{1}{\gamma}\Phi_n^{-\frac{1}{\theta}}$. To derive this result, we need to use the fact that the cdf of the price distribution in country n is given by $G_n(h) = 1 - \exp\left(-\Phi_n h^\theta\right)$. Then, using the law of large numbers and the density of this cdf, we have

$$\ln P_n = \int\limits_0^{+\infty} \ln h \Phi_n \theta h^{\theta-1} \exp\left(-\Phi_n h^\theta\right) dh.$$

However, the right-hand-side expression is now just the negative of the expectation of the natural logarithm of a variable with a Fréchet distribution with parameters Φ_n and θ. Thus, applying the formula for the value of this expection, we get

$$\ln P_n = -\ln\gamma - \frac{1}{\theta}\ln\Phi_n,$$

which immediately gives the result stated in the main text.

1.2 The model with CES preferences

1.2.1 Price distributions for different source countries

As argued in the main text, in the general Eaton-Kortum model with CES preferences, the probability π_{nj} that a country n imports a given good from country j is not automatically equal to the share of country n's total spending spent on goods from country j. This is only true in the special case in which the distribution of prices would be the exact same in every destination country j. In turns out that this is actually the case (this is the magic of the Fréchet distribution), and in this section, we show the result formally.

To do this, we derive the cdf G_{nj} of prices of goods which country n imports from country j, and show that it does not depend on j. To do this, note

$$G_{nj}(h) = \mathbb{P}\left(h_{nj} \leq h \mid h_{nj} \leq \min_{j' \neq j} h_{nj'}\right) = \frac{\mathbb{P}\left(\left(h_{nj} \leq h\right) \bigcap \left(h_{nj} \leq \min_{j' \neq j} h_{nj'}\right)\right)}{\pi_{nj}}.$$

Denoting by f the density of the random variable $\min_{j' \neq j} h_{nj'}$, we then have

$$\begin{aligned} G_{nj}(h) &= \tfrac{1}{\pi_{nj}} \int_0^{+\infty} \mathbb{P}\left(\left(h_{nj} \leq h\right) \bigcap \left(h_{nj} \leq x\right)\right) f(x)\, dx \\ &= \tfrac{1}{\pi_{nj}} \left[\int_0^h \mathbb{P}\left(h_{nj} \leq x\right) f(x)\, dx + \int_h^{+\infty} \mathbb{P}\left(h_{nj} \leq h\right) f(x)\, dx\right]. \end{aligned}$$

Now, all that is left is algebra. Noting $f(x) = \theta h^{\theta-1}\left(\Phi_n - T_j\left(w_j d_{nj}\right)^{-\theta}\right) \exp\left(-\left(\Phi_n - T_j\left(w_j d_{nj}\right)^{-\theta}\right) h^{\theta}\right)$ and replacing it in the equation yields, after some calculations,

$$G_{nj}(h) = 1 - \exp\left(-\Phi_n h^{\theta}\right),$$

which is equal to G_n, and independent of any country-j variable.

1.2.2 The ideal price index in the CES case

Note that, as before, the cdf of the price distribution in country n is given by $G_n(h) = 1 - \exp\left(-\Phi_n h^{\theta}\right)$. First, let us prove the useful intermediate result that the random variable p_n^{θ} follows an exponential distribution, with parameter Φ_n. To see this, just note that

$$\mathbb{P}\left(p_n^{\theta} \leq h\right) = \mathbb{P}\left(p_n \leq h^{\frac{1}{\theta}}\right) = 1 - \exp\left(-\Phi_n h\right),$$

which is the cdf of an exponential distribution.

The ideal price index holds

$$P_n^{1-\varepsilon} = \int_0^1 \left(\left(p_n(i)\right)^{\theta}\right)^{\frac{1-\varepsilon}{\theta}} di.$$

Using the fact that we know the density of p_n^{θ}, this writes

$$P_n^{1-\varepsilon} = \int_0^{+\infty} h^{\frac{1-\varepsilon}{\theta}} \Phi_n \exp\left(-\Phi_n h\right) dh.$$

Finally, using a change of variables $u = \Phi_n h$, we directly get

$$P_n = (\Phi_n)^{-\frac{1}{\theta}} \left(\int_0^{+\infty} u^{\frac{1-\varepsilon}{\theta}} \exp(-u)\, du \right)^{\frac{1}{1-\varepsilon}},$$

which is the result in the main text. By definition of the Gamma function, the constant $\int_0^{+\infty} u^{\frac{1-\varepsilon}{\theta}} \exp(-u)\, du$ is equal to $\Gamma\left(1 + \frac{1-\varepsilon}{\theta}\right)$.

2 The Caliendo-Parro model

This Appendix presents a slightly simplified version of model of Caliendo and Parro (2015), who extend the Eaton-Kortum model to include multiple sectors, intermediate inputs and tariffs.

2.1 The model

Assumptions Assume that consumers in each country consume I final goods, over which they have Cobb-Douglas preferences. Thus, the utility of the representative consumer of country n is

$$U_n = \prod_{i=1}^{I} \left(C_n^i\right)^{\alpha^i}, \tag{135}$$

with $\alpha^i > 0$ and $\sum_i \alpha^i = 1$. In each sector i, there is a unit mass of intermediate varieties, indexed on $[0, 1]$. These intermediate varieties are assembled under perfect competition with a CES production function

$$C_n^i = \left(\int_0^1 \left(c_n^i(\omega)\right)^{\frac{\varepsilon^i - 1}{\varepsilon^i}} d\omega \right)^{\frac{\varepsilon^i}{\varepsilon^i - 1}}, \tag{136}$$

where $\varepsilon^i > 1$, so that substitution across varieties is easier than substitution across final goods. We also assume sector-specific production functions,

$$y_n^i(\omega) = z_n^i(\omega) \left(l_n^i(\omega)\right)^{\beta^i} \prod_{k=1}^{I} \left(m_n^{ik}(\omega)\right)^{\beta^{ik}}, \text{ with } \beta^i + \sum_{k=1}^{I} \beta^{ik} = 1. \tag{137}$$

In every sector, output is produced with a Cobb-Douglas production function combining labour and the I final goods. Proportions are allowed to be sector-specific, taking into account that different sectors use different inputs (for instance, we would have $\beta^{\text{Rubber,Tyres}} > 0$, but $\beta^{\text{Tyres, Rubber}} = 0$).

Note that this model collapses to the Eaton-Kortum model if $I = 1$ and $\beta^i = 1$ (with a CES utility function instead of the Cobb-Douglas one we considered above),[119] and to our toy model above if $I = 1$ and for all sectors i and i', $\beta^{ik} = \beta^{i'k}$. We maintain the Fréchet assumption for productivity, but now allow the mean parameters T_n^i to vary by country and by sector, and the dispersion parameters θ^i to vary by sector.

Finally, we continue to model distance and other trade frictions as iceberg costs d_{nj}^i, but now allow them to differ by sector. Also, we now introduce ad-valorem tariffs t_{nj}^i. As in the real world, tariffs are taxes on imports which firms need to pay to the importing country's government, and we assume that the government rebates all tariff income to consumers.

Equilibrium Using the results from Chapter 1.2, it is easy to show that for consumers in country n, the costs for importing good ω of sector i from country j are

$$h_{nj}^i(\omega) = \frac{\left(1 + t_{nj}^i\right) d_{nj}^i}{z_j^i(\omega)} \left(\frac{w_j}{\beta^i}\right)^{\beta^i} \prod_{k=1}^{I} \left(\frac{P_j^k}{\beta^{ik}}\right)^{\beta^{ik}}, \tag{138}$$

where P_j^i is the ideal price index for sector i in country j. To simplify notation, let us define $\kappa_{nj}^i = \left(1 + t_{nj}^i\right) d_{nj}^i$ and $u_j^i = \left(\frac{w_j}{\beta^i}\right)^{\beta^i} \prod_{k=1}^{I} \left(\frac{P_j^k}{\beta^{ik}}\right)^{\beta^{ik}}$. Then, we can derive the equivalents of Equations (26) and (27), which now hold for every sector, in exactly the same way than in the standard Eaton-Kortum model. Likewise, we get the same formula for expenditure shares: π_{nj}^i, the fraction of country n's total spending in sector i spent on goods from country j, is given by

$$\pi_{nj}^i = \frac{T_j^i \left(u_j^i \kappa_{nj}^i\right)^{-\theta^i}}{\Phi_n^i}, \tag{139}$$

where $\Phi_n^i \equiv \sum_{j=1}^{N} T_j^i \left(u_j^i \kappa_{nj}^i\right)^{-\theta^i}$.[120] Likewise, we can derive the formula for the price index in every sector i, which holds

$$P_n^i = \left(\Gamma\left(1 + \frac{1 - \varepsilon^i}{\theta^i}\right)\right)^{\frac{1}{1-\varepsilon^i}} \left(\Phi_n^i\right)^{-\frac{1}{\theta^i}}, \tag{140}$$

[119] In particular, it makes no difference whether we assume that firms assemble the final good and sell it to the consumer, or whether consumers directly consume the intermediates and have a utility function over them. Our basic model is identical if you assume that consumers just consume one good, which is sold to them by firms assembling it with a Cobb-Douglas production function (and selling it in equilibrium at the ideal price index).

[120] For this, we have used the result explained in Section 4.2.4 and derived in Appendix VIII, namely that the price distribution is the same for any source country (implying that the fraction of goods of sector i which country n imports from country j equals the fraction of country n's spending in sector i spent on goods from country j).

so that the aggregate price index in each country is given by

$$P_n = \prod_{i=1}^{I} \left(\frac{P_n^i}{\alpha^i} \right)^{\alpha^i}. \tag{141}$$

To close the model, we need to specify the market clearing conditions. For this, note that Y_n^i, the total expenditure on goods from sector i in country n, is equal to

$$Y_n^i = \alpha^i \left(w_n L_n + \sum_{k=1}^{I} \sum_{j=1}^{N} t_{nj}^k \frac{Y_{nj}^k}{1 + t_{nj}^k} \right) + \sum_{j=1}^{N} \frac{Y_{jn}^i}{1 + t_{jn}^i}. \tag{142}$$

To understand this formula, let us proceed in steps. Goods from sector i in country n are bought by country n's consumers, and by firms in all countries of the world.

1. The first term in the above equation captures consumer spending. As consumers have Cobb-Douglas preferences over sectors, they just spend a fraction α^i of their income on goods from sector i. Their income, in turn, is composed by labour income, and the total tariff revenue collected by their government, from imports in all sectors and from all trading partners. Tariff revenues for imports from sector k and country j are a fraction t_{nj}^k of the pre-tariff value of imports, equal to $\frac{Y_{nj}^k}{1+t_{nj}^k}$, where Y_{nj}^k is total spending of country n on goods from sector k and country j.

2. The second term captures the spending of firms, which use good i from country n as an intermediate input. Each country j spends Y_{jn}^i on good i from country n, and firms in country n receive the pre-tariff value of this expenditure, $\frac{Y_{jn}^i}{1+t_{jn}^i}$.

From this, we can get an accounting equation, just as in the Eaton-Kortum model, expressing that the total spending of country n in all sectors (excluding tariffs) must equal the total spending on country n goods. This implies that trade is balanced, so that imports are equal to exports.[121] Formally, this condition writes

$$\sum_{i=1}^{I} \sum_{j=1}^{N} \frac{Y_{nj}^i}{1 + t_{nj}^i} = \sum_{i=1}^{I} \sum_{j=1}^{N} \frac{Y_{jn}^i}{1 + t_{jn}^i}. \tag{143}$$

[121] Caliendo and Parro (2015) allow for trade deficits, from which we abstract here for simplicity. They model these as transfers, i.e., as gifts from one country to the other. This is, of course, a shortcut: in reality, trade deficits are not gifts, but the reflection of international borrowing and lending.

2.2 Relative changes

Equilibrium conditions Just as in for the basic Eaton-Kortum model, it is most convenient to solve the Caliendo-Parro model in relative changes. Let us assume that the change we want to study with our model is a modification of tariffs from $\left(t_{nj}^{i}\right)_{n,j\in[\![1,N]\!]}^{i\in[\![1,I]\!]}$ to $\left(t_{nj}^{\prime i}\right)_{n,j\in[\![1,N]\!]}^{i\in[\![1,I]\!]}$. Denote the proportional change in any variable v after the change in trade costs by $\hat{v} \equiv \frac{v'}{v}$. Our objective is to solve for these proportional changes for all of the most important variables, which is everything we are interested in. To do so, we will construct a system of four sets of equations which we can then solve using a computer.

First, from Equation (140), we have that $\widehat{P}_{n}^{i} = \left(\frac{\Phi_{n}^{\prime i}}{\Phi_{n}^{i}}\right)^{-\frac{1}{\theta^{i}}}$. Using the definition of $\Phi_{n}^{\prime i}$, this yields

$$\widehat{P}_{n}^{i} = \left(\sum_{j=1}^{N} \left(\hat{u}_{j}^{i} \hat{\kappa}_{nj}^{i}\right)^{-\theta^{i}} \pi_{nj}^{i}\right)^{-\frac{1}{\theta^{i}}}, \tag{144}$$

where $\hat{u}_{j}^{i} = \left(\widehat{w}_{j}\right)^{\beta^{i}} \prod_{k=1}^{I} \left(\widehat{P}_{j}^{k}\right)^{\beta^{ik}}$ and $\hat{\kappa}_{nj}^{i} = \frac{1+t_{nj}^{\prime i}}{1+t_{nj}^{i}}$.[122] From Equation (139), we then get

$$\widehat{\pi}_{nj}^{i} = \left(\frac{\hat{u}_{j}^{i} \hat{\kappa}_{nj}^{i}}{\widehat{P}_{n}^{i}}\right)^{-\theta^{i}}. \tag{145}$$

Finally, we can rewrite the national accounting identity in the equilibrium with the new tariffs as

$$Y_{n}^{\prime i} = \alpha^{i} \left(\widehat{w}_{n} \left(w_{n} L_{n}\right) + \sum_{k=1}^{I} \sum_{j=1}^{N} t_{nj}^{\prime k} \frac{Y_{nj}^{\prime k}}{1+t_{nj}^{\prime k}}\right) + \sum_{j=1}^{N} \frac{Y_{jn}^{\prime i}}{1+t_{jn}^{\prime i}}. \tag{146}$$

Solution algorithm In order to solve the model, we first need to set some parameter values using data. We can set the βs using data from national accounts and input-output tables: according to our model, β^{ik} is equal to the percentage of the total sales of industry i that get spent on input k, which is observable in the data. Likewise, we can set the αs using data on consumer spending. t and t' are set using tariff data and specifying the policy change we want to study, and the trade elasticities θ^{i} can be estimated using gravity equations, as discussed in Chapter II. Finally, we set the spending shares π_{nj}^{i} using trade data, and set $w_n L_n$ equal to national incomes. These last choices are conceptually different from the previous ones, because here we do not set a parameter value, but the value of an

[122] To derive Equation (144), note that $\frac{\Phi_{n}^{\prime i}}{\Phi_{n}^{i}} = \sum_{j=1}^{N} \frac{T_{j}^{i}}{\Phi_{n}^{i}} \left(u_{j}^{\prime i} \kappa_{nj}^{\prime i}\right)^{-\theta^{i}} = \sum_{j=1}^{N} \pi_{nj}^{i} \left(\frac{u_{j}^{\prime i} \kappa_{nj}^{\prime i}}{u_{j}^{i} \kappa_{nj}^{i}}\right)^{-\theta^{i}}$.

endogenous variable (a model outcome). By doing so, we assume that our model is the true model of the world: otherwise, this substitution would not make sense.

We are now ready to solve the model, using the following algorithm.

1. Make a guess for the vector of nominal wage changes in all countries, $(\widehat{w}_1, .., \widehat{w}_N)$.

2. Using this guess, solve the Equation system defined by (144). This is an equation system in $N \times I$ equations, and $N \times I$ unknowns (the changes in price levels for every sector in every country), and can be solved by a computer.

3. From the solution of this system, deduce the values of $\widehat{\pi}^i_{nj}$ using Equation (145). This immediately gives π'^i_{nj}, the values of spending shares in the new equilibrium.

4. Using the Equation system defined by (146), another system in $N \times I$ equations, and $N \times I$ unknowns (spending on every sector in every country), we solve for the Y'^i_ns. For this, we need to use the fact that $Y'^i_{jn} = \sum_{k=1}^{I} \beta^{ki} \pi'^i_{jn} Y'^k_j$. To understand this equation, note that in each country j, the sales of firms producing k are Y'^k_j. Because of their Cobb-Douglas technology and perfect competition, firms spend a fraction β^{ki} of these sales on intermediate inputs from sector i, and a fraction π'^i_{jn} of these comes from country n.

5. We have now solved for all relevant endogenous variables. However, we still need to check whether our initial guess was correct. To do so, we plug our results into the trade balance equation (143), and check whether the trade balance holds with equality for every country. If this is not the case, we need to update our guess (increasing wages for surplus countries and decreasing them for deficit countries) and start again from 2. If this is the case, the algorithm has converged and we have found the equilibrium.

Again, solving the model in relative changes allows us not to make any assumptions for a large number of model parameters, such as the average productivities, the elasiticities of substitution, and the level of non-tariff trade costs.

The ACR formula in the Caliendo-Parro model From the definition of u, we know that changes in the wage of country n hold, for every sector i,

$$\widehat{w}_n = \left(\frac{\hat{u}_n^i}{\prod_{k=1}^{I} \left(\widehat{P}_n^k \right)^{\beta^{ik}}} \right)^{\frac{1}{\beta^i}} .$$

Furthermore, the change in the home spending share in sector i is $\widehat{\pi}_{nn}^i = \left(\frac{\hat{u}_j^i}{\widehat{P}_n^i} \right)^{-\theta^i}$. Combining these two equations, we get an expression for $\widehat{w}_n / \widehat{P}_n^i$. To get to the change in real wages, $\widehat{w}_n / \widehat{P}_n$, we need to use the ideal aggregate price index of Equation (141). We then get, after some algebra, that

$$\frac{\widehat{w}_n}{\widehat{P}_n} = \underbrace{\prod_{i=1}^{I} \left(\widehat{\pi}_{nn}^i \right)^{-\frac{\alpha^i}{\theta^i}}}_{\text{Classic ACR}} \underbrace{\prod_{i=1}^{I} \left(\widehat{\pi}_{nn}^i \right)^{-\frac{\alpha^i}{\theta^i} \frac{1-\beta^i}{\beta^i}}}_{\text{Input multiplier}} \underbrace{\prod_{i=1}^{I} \left(\left(\widehat{P}_n^i \right)^{\frac{1-\beta^i}{\beta^i}} \left(\prod_{k=1}^{I} \left(\widehat{P}_n^k \right)^{\beta^{ik}} \right)^{-\frac{1}{\beta_i}} \right)^{\alpha^i}}_{\text{Sectoral linkages}} . \tag{147}$$

When there is just one sector and there are no input-output linkages, only the first term remains, and we are back in the baseline ACR case. With one sector and intermediate inputs which are used in the same way in every sector, the first and the second term remain, and we recover the ACR formula from our previous toy model in Equation (52). Finally, in the full model, one additional term appears, which takes into account the exact linkages between sectors, showing for instance that wage gains will be larger if prices fall for important inputs in important sectors.

Note, however, that the real wage is not a sufficient statistic for welfare any more in this model. This is because there is now also revenue from tariffs: to assess whether a country is better or worse off after a change in trade costs, we need to consider the entire national income, and not just wage income.

Part IX

Appendix: Trade, increasing returns, and divergence

1 Introduction

It is often argued that international trade may be detrimental to countries which choose a "bad" specialization, that is, focus on exporting goods which have a low potential for productivity growth and/or increasing returns. Therefore, international trade is seen as a reason for the divergence between Northern and Southern countries during the 19th century, when the former concentrated on manufacturing production and the latter on agricultural and primary products.

However, is this argument really convincing? In the models we have seen so far, international trade makes the gains from increasing returns and higher productivity spill over to all countries of the world: if one country can produce more and more of a given good, the latter's price must fall, turning the terms of trade in favour of the importing countries.[123]

None of these models can explain why ex-ante identical countries (as the global North and South were, more or less, before the 18th century) could have very different fortunes after opening up to trade. A line of research starting in the 1990s argues that agglomeration effects are this missing element that links increasing returns and divergence. The basic idea is that many goods are produced not only from capital and labour, but also use a lot of intermediate inputs. For the producers of these intermediate inputs, it has been historically very important to be close to their customers, and this importance could even overturn the desire to minimize costs. For example, in its early years, the US car industry was quite concentrated around the city of Detroit. Even if it may have been cheaper for intermediate input producers to produce in, say, Denver, they did not do this, as it would have been quite expensive to then ship these goods to Ford and General Motors, which were in Detroit.

These agglomeration effects, coupled with increasing returns, can lead to divergence between ex ante identical countries. The remainder of this section illustrates this with a model similar to the ones we have seen so far (a simplified version of Matsuyama, 2013). That model is in many aspects very similar to a famous model by Krugman and Venables (1995), with which they aimed to explain both

[123]This holds in the Krugman model, and also in Ricardian models (see, for instance, Problem Set 2).

the North-South divergence in the 19th century and the beginning convergence nowadays (their paper was initially entitled "History of the World, Part 1"). The basic story of these models, which consider two ex ante exactly identical countries, goes like this:

1. Initially, prohibitively high transport costs force the two countries to live in autarky, where they have the exact same income.

2. In a second phase (the 19th century), transport costs fall and make final goods tradable. However, intermediate inputs remain non-tradable. Therefore, agglomeration effects appear: one country may specialize in industries which are more intensive in intermediate inputs (manufacturing), and therefore, intermediate input producers want to locate there. Because of increasing returns, this country (called the North) develops a greater variety of intermediate inputs, allowing it to get competitive in many other industries as well, and to capture a larger share of world income. The two countries diverge, even though they were initially exactly the same.

3. In a third phase (end of the 20th/beginning of the 21st century?), intermediate goods become tradable. Thus, intermediate input producers start to move production to the cheaper South, and the two countries start to converge again.

Note that the optimistic conclusion on the eventual convergence of the two countries depends on the assumption that agglomeration effects were due only to trade costs. If they were do to some other factors as well (that is, some non-material advantages for industries to cluster in one point), then even the disappearance of transport costs cannot eliminate divergence completely.

2 A model of divergence

2.1 Assumptions

The Matsuyama model is essentially a mix of two models we have covered already. The basic structure is the one of the Dornbusch-Fischer-Samuelson model. There are two exactly equal countries, called North (N) and South (S). Both countries are populated by a mass L of identical consumers which have Cobb-Douglas preferences on a continuum of industries indexed by $[0, 1]$ and inelastically supply one unit of labour.

The goods of these industries are produced by a continuum of firms operating under perfect competition. However, and here we deviate from Dornbusch-Fischer-Samuelson, the production of these goods does not only use labour, but also a continuum of intermediate inputs, indexed on $[0, M_j]$. Precisely, the production function for industry i is

$$y_j(i) = L_j(i)^{1-\beta(i)} \left(\left(\int_0^{M_j} (q_j(i,m))^{\frac{\varepsilon-1}{\varepsilon}} dm \right)^{\frac{\varepsilon}{\varepsilon-1}} \right)^{\beta(i)}.$$

This production function implies that all industries use the same inputs. However, they do so in different proportions. Industries with a low value of $\beta(i)$ use a lot of labour and little intermediate inputs, while the reverse is true for industries with high values of $\beta(i)$. We will assume that β is continuous and strictly increasing in i (that is, industries ordered from the least to the most reliant on intermediate inputs).

Intermediate input production takes place under monopolistic competition and is subject to increasing returns, just as in the basic Krugman model. Thus, each intermediate input producer has a monopoly on the production of one variety m and the same production technology as in the basic Krugman model: to produce $y_j(m)$ units of intermediate inputs, it needs to hire $f + \frac{1}{z} y_j(m)$ units of labour.

Note that when $\beta(i) = 0$ in all industries (and we introduce some productivity differences), we get the Dornbusch-Fischer-Samuelson model. If instead $\beta(i) = 1$ in all industries, and intermediate inputs are tradable, we get the Krugman model, just that the gains from greater variety arise now at a different level (instead of consumers consuming a greater number of varieties, producers will now produce with a greater number of varieties, which will lower their costs and allow them to produce more). In the general case, where $\beta(i)$ varies across industries, we will get agglomeration effects (if intermediate inputs are not traded).

However, we start by studying the autarky equilibrium, in which neither final goods nor intermediate inputs are tradable.

2.2 The autarky equilibrium

As North and South are identical, the autarky equilibrium is the same in both countries. Therefore, to ease notation, I drop the country index j.

The problem of final goods producers As a first step, we need to determine how much labour and how much of every intermediate input is demanded by the final goods producers of each industry. Although their production function may look complicated, it is actually a combination of things we know: you can recognize that it is just a Cobb-Douglas production function in labour and intermediate inputs (and therefore, we will get the classical result that a share $\beta(i)$ of the revenue of every industry will flow to intermediate input producers and a share $1-\beta(i)$ to labour, as in Chapter III). Intermediate inputs are aggregated with a CES aggregator, and this yields a demand function with a constant price elasticity for the output of every intermediate input producer.

One can show that the total cost function of final goods producers of industry i is

$$TC(y(i)) = \left(\frac{w}{1-\beta(i)}\right)^{1-\beta(i)} \left(\frac{1}{\beta(i)}\left(\int_0^M p(m)^{1-\varepsilon}\,dm\right)^{\frac{1}{1-\varepsilon}}\right)^{\beta(i)} y(i). \tag{148}$$

Furthermore, it is also possible to show that the demand function of a final goods producer for one particular intermediate input has price elasticity ε.

Intermediate input producers As the price elasticity of demand is ε for every intermediate input producer, it chooses a price

$$p(m) = \frac{\varepsilon}{\varepsilon-1}\frac{w}{z}.$$

Thus, all intermediate inputs producers charge the same price p, and therefore also produce the same quantity y. This quantity can be obtained, just as in the basic Krugman model, by using the zero-profit condition implied by free entry, which gives

$$y = fz(\varepsilon-1).$$

Aggregating up As a last step, let's determine the mass of intermediate inputs produced by every country. This is a little more difficult than in the basic Krugman model, because we do not get it directly from the labour market clearing any more (as labour is now split between final and intermediate goods production, making calculations more difficult).

However, we can remark that the total revenue obtained by all intermediate input producers is Mpy. Furthermore, because of the constant cost shares implied by the Cobb-Douglas production function, the spending of industry i on intermediate inputs is $\beta(i)\,wL$.[124] Therefore, it must be that

$$Mpy = \int_0^1 \beta(i)\,wLdi.$$

Using the expressions for p and y derived above, this gives the autarky mass of varieties M_A as

$$M_A = \frac{\Gamma_A}{f\varepsilon}L,$$

where $\Gamma_A \equiv \int_0^1 \beta(i)\,di$ is a positive constant.

We have not yet fully solved for the values of all variables in equilibrium. To proceed, we would need to determine how much labour is used in the final production of every industry, and this would give us the equilibrium production of each good, and the equilibrium wage. However, going to this full solution is not necessary to make the main points of this model. The important take-away from this autarky equilibrium is just that as the two countries are identical, they will have identical equilibrium values for all aggregate variables.

2.3 Trade in final goods

Now, assume we are in the "second phase" of the Krugman-Venables story: final goods can be traded, but intermediate inputs are nontradable. In this case, it can be shown that the model has multiple equilibria, generated by agglomeration effects. Here, we are interested in showing that there is one equilibrium in which the North develops more varieties ($M_N > M_S$) and therefore gets a higher national income. So, let's solve for the equilibrium under the assumption that $M_N > M_S$, and see whether the equilibrium we find verifies that initial assumption.

Solving for an asymmetric equilibrium The behaviour of consumers, final goods producers and intermediate input producers remains essentially the same as in autarky. In particular, the price charged by intermediate input producers is still given by $p = \frac{\varepsilon}{\varepsilon-1}\frac{w}{z}$. Therefore, using the total cost function given by Equation (148), the marginal cost of an industry i firm in country j is given by

$$MC_j(i) = w_j\left(M_j^{-\frac{\beta(i)}{\varepsilon-1}}\left(\beta(i)\frac{\varepsilon-1}{\varepsilon}\right)^{-\beta(i)}(1-\beta(i))^{\beta(i)-1}\right).$$

[124]This uses the fact that because of the Cobb-Douglas preferences of consumers, the total spending of consumers on industry i is wL.

Now, I need to determine which country produces which final good in equilibrium. Just as in the Dornbusch-Fischer-Samuelson model, the North will produce in industry i if it has a lower production cost, i.e.,

$$MC_N(i) \leq MC_S(i)$$
$$\iff \frac{w_N}{w_S} \leq \left(\frac{M_N}{M_S}\right)^{\frac{\beta(i)}{\varepsilon-1}} \tag{149}$$

This is an exactly analogue expression to the Dornbusch-Fischer-Samuelson model, with $T(i) = \left(\frac{M_N}{M_S}\right)^{\frac{\beta(i)}{\varepsilon-1}}$. Whereas specialization was given by productivity differences in the Ricardian model, it is now driven by differences in the varieties of intermediate inputs available in each country. Note that because we have assumed $M_N > M_S$, the right-hand side of this equation is increasing in i. Let's denote by i^* the industry for which Equation (149) holds with equality, that is,

$$\frac{w_N}{w_S} = \left(\frac{M_N}{M_S}\right)^{\frac{\beta(i^*)}{\varepsilon-1}}. \tag{150}$$

Then, the North produces all industries indexed from $i^*$to 1, and South all industries from 0 to i^*. We can note already that because $M_N > M_S$, in equilibrium, it must be that $w_N > w_S$. Just as in the Ricardian model, the less productive country needed to compensate its lower productivity with a lower wage, here, the country with a lower variety of intermediate inputs needs to compensate that disadvantage with a lower wage. This already implies that the North will be richer than the South in equilibrium, as national income is just given by $w_j L_j$, and $L_N = L_S = L$.

However, we have yet to verify that our initial conjecture $M_N > M_S$ holds true. For this, we determine the mass of variety produced in every country just as in autarky. The revenue of the North's intermediate input producers is now given either by $M_N py$, or by the aggregation of their revenues $\beta(i)(w_N L + w_S L)$ across all the industries the North produces. Therefore,

$$M_N py = \int_{i^*}^{1} \beta(i)(w_N L + w_S L) di. \tag{151}$$

Furthermore, we can remark that the Cobb-Douglas structure on final goods implies that $w_N L = (1 - i^*)(w_N L + w_S L)$, that is, that a country which produces a fraction $1 - i^*$ of the world's final goods also captures a fraction $1 - i^*$ of world income. Replacing this in the above expression, we get

$$M_N = \frac{\Gamma_N(i^*)}{f\varepsilon} L \tag{152}$$

where $\Gamma_N(i^*) = \frac{1}{1-i^*}\int_{i^*}^{1} \beta(i)\, di$. In the same way, we get for the South

$$M_S = \frac{\Gamma_S(i^*)}{f\varepsilon} L \tag{153}$$

where $\Gamma_S(i^*) = \frac{1}{i^*}\int_0^{i^*} \beta(i)\, di$. Now, the only thing left to do is to verify that $\Gamma_N(i^*) > \Gamma_S(i^*)$. This is true because Γ_S is strictly increasing in i^*[125], and therefore, $\Gamma_S(i^*) < \Gamma_S(1) = \Gamma_A$, and in the same way, $\Gamma_{N,1}$ is strictly decreasing in i^*, so $\Gamma_N(i^*) > \Gamma_N(0) = \Gamma_A$.

Interpretation Relative incomes after final goods trade is given by

$$\frac{w_N L}{w_S L} = \frac{w_N}{w_S} = \frac{1-i^*}{i^*}$$

and the cut-off industry is defined by the equation

$$\frac{i^*}{1-i^*} = \left(\frac{\Gamma_N(i^*)}{\Gamma_S(i^*)}\right)^{-\frac{\beta(i^*)}{\varepsilon-1}}$$

We can note several things about this model. First, as there exists an equilibrium with $M_N > M_S$, there obviously exists as well the exactly analogue equilibrium with $M_N < M_S$.[126] The model explains why divergence is possible even between ex ante identical countries, but it cannot explain why it was the North that managed to grow rather than the South.

Why is there divergence? Matsuyama explains that there is a two-way causality:

1. A greater variety of intermediate inputs gives a country a comparative advantage for industries which rely on intermediate inputs more heavily (sectors with a high $\beta(i)$).

2. And a country producing in sectors with a high $\beta(i)$ will develop a larger variety of intermediate inputs.

---

[125] Indeed, $\Gamma_S'(i^*) = \frac{1}{i^*}\left(\beta(i^*) - \frac{1}{i^*}\int_0^{i^*} \beta(i)\, di\right) > 0$, because β is a strictly increasing function.

[126] There is also a symmetric equilibrium, with $M_N = M_S$, and half of the industries produced in every country. However, this equilibrium is unstable: if a small mass of intermediate input producers would move to another country (say, the North), industries with a high $\beta(i)$ would immediately want to move to the North as well, and this would set off a feedback loop that would bring the economy again to the asymmetric equilibrium.

This feedback loop generates agglomeration effects, and therefore, divergence between countries.

2.4 Trade in intermediate inputs: back to convergence

In the equilibrium above, Northern intermediate input producers could have produced more cheaply in the South (as $w_S < w_N$). However, they did not move to the South because intermediate inputs were not tradable, and therefore, moving to the South would have meant losing the ability to sell to their clients. These linkages are key to generate the divergence in this model.

Once this restriction is lifted and intermediate inputs become tradable, intermediate input producers move South to take advantage of cheaper production costs. This drives up labour demand and wages in the South, until $w_S = w_N$, and the two countries again have the exact same income.

Is this optimistic vision of globalization justified? We do observe some of its elements in reality: many firms have taken advantage of lower transport costs to split up production chains and produce some intermediate inputs in the South. This has clearly been a factor behind the rapid growth of countries such as China. On the other hand, it is unclear that transport costs are the only sources of industry agglomeration. For example, it is unlikely that the US IT industry is concentrated in Silicon Valley just in order to avoid transport costs. Other agglomeration forces, linked to externalities and knowledge spillovers, are unlikely to disappear with transport costs, and may therefore be more durable forces for divergence.

Part X

Appendix: Trade and growth

1 A (very) short introduction to endogenous growth models

In the long run, most of the growth in GDP per capita in developed countries is due to increases in productivity. For instance, for the United States, Jones (2016) estimates that increases in TFP explain 80% of the increase in GDP per capita between 1948 and 2013.[127]

Productivity increases are mostly due to innovations, that is, the creation of new products and pro-

[127] These numbers are based on a simple "growth accounting" methodology, which attributes all increases in GDP which are not due to increases in capital, labour and human capital to a residual term which is generally interpreted as representing productivity. For details, see Jones (2016).

duction processes. The fundamental idea of endogenous growth theories, which became popular in the early 1990s, is that these innovations are not random events, but respond to economic incentives, at least for the substantial part of inventions and applications made by the business sector of the economy. Endogenous growth theory stresses, however, that the production of ideas and innovations has some special characteristics. Most importantly, ideas and innovations are generally **non-rival**. This means that if one firm makes a discovery (for example, a way to produce aluminium using less electricity), other firms can use that discovery at the same time than the original inventor, without (or almost without cost). As Romer (1990) puts it, "*once the cost of creating a new set of instructions has been incurred, the instructions can be used over and over again at no additional cost.*" This property makes ideas particular production factors (in contrast to them, capital or labour are both rival: two firms cannot use at the same time the same machines or the work of the same workers). This implies that no firm would ever invest into innovation under perfect competition: an innovating firm would be alone to pay the costs of innovation, but the benefits would be freely available to any other firm on the market, eroding any potential profits.

Therefore, models of endogenous growth always feature imperfect competition, and some permanent or temporary monopoly power for inventing firms. This allows firms to earn profits from their innovations, and therefore explains that there are investments into innovation in equilibrium. In practice, firms manage to achieve temporary monopoly power by using the patent system, or trying to keep their innovations secret. When we discussed the Bustos model, we implicitly assumed that firms have a patent on their variety and on any improvements of that variety, giving them a monopoly. Otherwise, no firm would have been willing to pay the cost of entry (which can be interpreted as the cost of inventing a new variety) and/or the cost of upgrading technology.

Models of endogenous growth (pioneered by Romer (1990), Grossman and Helpman (1991)), and Aghion and Howitt (1992)) have enabled economists to analyse more precisely the determinants of economic growth, and stressed the importance of innovation incentives. As we have seen in the Bustos model, international trade also affects innovation incentives, and therefore, we can use models of endogenous growth to analyse more precisely how trade affects the rate of productivity growth.[128]

[128] Endogenous growth models are not the only possible framework to study the effects of trade on economic growth. Alternatively, we could introduce growth into the classical models for instance by allowing for capital accumulation (as in the classical Solow (1956) model). In a closed economy, capital accumulation runs into decreasing returns: the more capital there is, the lower is its marginal product and therefore the rental rate of capital. Eventually, the economy reaches a steady state in which growth stops. In an open economy, structural transformation can avoid these increasing returns during a certain time, as suggested by the Rybzcynski theorem (see Part III).

2 How does trade affect innovation incentives?

2.1 Four important mechanisms

In a recent review paper, Grossman and Helpman (2015) list four mechanisms through which international trade can affect innovation (and thereby productivity growth).

1. **Market size effects**. International trade allows innovating firms to sell their new or improved product not just to their home market, but also to many other countries. This obviously increases the profits from innovation and therefore gives firms an incentive to innovate more.[129] This channel is also the most important reason for why trade (or a decrease in trade costs) increase innovation investment in the Bustos model.

2. **Competition**. International trade exposes firms to more competition. We have seen this most clearly in the Krugman and Melitz models, where the appearance of foreign varieties causes local variety sales to fall. What effect does higher competition have, generally speaking, on innovation incentives? It turns out that this is a very tricky question, as there are two effects working in opposite directions:

   (a) On the one hand, competition reduces the incentives to innovate, as it reduces the profits which can be made by a successful innovator. Indeed, as we have seen in the previous section, under perfect competition, no innovation at all takes place. This channel is present in Bustos's model (just as in the Melitz model, opening up decreases the demand for domestic varieties $\frac{B}{w}$, and this lowers innovation incentives) but it turns out to be weaker than the market size channel.

   (b) On the other hand, more fierce competition increases the incentives to escape competition by innovating and recovering temporary monopoly power. This motive is not present in Bustos's model. However, it is the main explanation that Bloom *et al.* (2016) provide for their finding that Chinese competition seems to have increased innovation by European firms. This seems to be the case for example for "*footwear, a classic low-tech sector that conventional wisdom says should have all been offshored to China. Many Western shoe*

[129] This market size effect has also been mentioned by some researchers to stress the importance of international trade for the Industrial Revolution (see slides for Lecture 1). A quote from Matthew Boulton, the business partner of James Watt (the inventor of steam power) illustrates this idea well. Boulton wrote to Watt that "*[i]t is not worth my while to manufacture your engine for three countries only, but I find it very well worth my while to make it for all the world*" (quoted in Acemoglu, 2009).

manufacturers have disappeared, but some are innovating in designs that serve parts of the market where China is less able to compete". [130]

These two opposed effects make it difficult to predict in general how increased competition affects innovation. Aghion, Bloom, Blundell, Griffith and Howitt (2005) have studied this question in greater details and find evidence for an inverse U-shaped relationship: there is no innovation when there is perfect or no competition, and innovation is highest for "intermediate" levels of competition.

3. **Knowledge spillovers**. International trade also affects the costs for coming up with new ideas or innovations, and the possibilities of domestic inventors to build on the foreign state of knowledge. This has been a source of greater innovation and growth for many centuries, as trade enabled, for example, Europeans to become conscious of Arab and Chinese discoveries in the fields of mathematics and engineering.

4. **Technology diffusion**. Technology diffusion is linked to knowledge spillovers, but it is not exactly the same thing. Grossman and Helpman speak of technology spillovers when trade enables a country to import the latest products and production processes from other countries (for example, by giving the opportunity to an Italian firm to import the latest Microsoft software). These become knowledge spillovers only if the Italian firm uses this knowledge to do some innovation itself. Thus, technology diffusion is a more "passive" process than knowledge spillovers. Neither of these two channels is present in the Bustos model which we studied before.

In general, market size effects, knowledge spillovers and technology diffusion suggest that trade should have a positive impact on innovation. However, the competition channel is more ambiguous, and therefore, the exact effect of trade on innovation depends on the details of the model. In the next section, we study a simple model which will further illustrate the role played by market size, competition and knowledge spillovers.

[130] This quote is taken from a summary of their research available on http://www.voxeu.org/article/who-s-afraid-big-bad-dragon-how-chinese-trade-boosts-european-innovation. The authors also provide an example for an innovating firm: "*Massai Barefoot Technology (MBT), which makes posture-correcting shoes, began when Karl Muller, a Swiss engineer with a bad back, relieved his condition by walking barefoot on Korean grass. He patented a design to emulate the effect, which has gone on to great success [...]*"

2.2 A simple model of endogenous growth

This section presents a very simple endogenous growth model in the tradition of Romer (1990).[131] Romer's original endogenous growth model is almost identical to a closed-economy version of Krugman (1980). Indeed, one can already consider Krugman (1980) as a model of (product) innovation, when interpreting the creation of a new variety as an innovation. Note, in particular, that the model assumed that the inventor of every new variety is granted monopoly power forever (which one could interpret as a perpetual patent). Continuous economic growth would imply that over time, more and more varieties are created.

Let us, as a first step, introduce time in the autarky version of the Krugman model. To do so, assume that time is discrete ($t \in \mathbb{N}$) and that there is one representative consumer every period[132], supplying inelastically L units of labour and choosing consumption to maximize the utility function

$$C_t = \left(\int_0^{M_t} c_t\left(i\right)^{\frac{\varepsilon-1}{\varepsilon}} di \right)^{\frac{\varepsilon}{\varepsilon-1}} .$$

If all other assumptions are kept unchanged, sustained growth in the mass of varieties is impossible. Why?

The most important obstacle for economic growth in this model are the fixed production costs, f. Indeed, if the mass of varieties M_t would grow at a constant rate g (that is, $\frac{M_t - M_{t-1}}{M_{t-1}} = g$), the total amount of labour needed to cover the fixed costs of all producers, $M_t f$, would also grow at rate g. This means it would at some point become larger than the total labour supply L, which is obviously a contradiction. Thus, for sustained growth (in this model), we need to get rid of the fixed cost of production. Following Romer, let us assume that instead of it, there is a fixed cost for creating a new variety in period t, of $n\left(M_{t-1}\right)$ units of labour. Crucially, this cost must only be paid once (at the moment the variety is created) and not in every period. Moreover, it may depend on the mass of varieties previously invented, M_{t-1}. With this, we create a role for knowledge spillovers, which may be positive or negative: when n decreases in M_{t-1}, a large number of varieties makes it easier to create a new variety, when n increases in M_{t-1}, it makes it harder.

Finally, we make one additional assumption, for technical reasons: each firm's monopoly to produce

[131] The section is loosely based on the model of Baldwin and Robert-Nicoud (2008). They consider firm heterogeneity as in the Melitz model, but I abstract from this, as it is not necessary to make the main points.

[132] I assume that there is one consumer per period (i.e., that people live one period and do not care about future generations) to keep the problem as simple as possible and to not introduce a consumption-saving decision.

its variety only lasts one period (you may say that the firm is granted a patent which lasts one period). Afterwards, all firms are allowed to produce the variety, which means that there is effectively perfect competition.[133]

With these new assumptions, what is the model's equilibrium?

Now, in period t, there are two categories of firms. Firms which created their varieties in earlier periods (a mass M_{t-1}) operate under perfect competition, and therefore charge a price equal to marginal cost,

$$p_t^{\text{Comp}} = \frac{w_t}{z}.$$

On the other hand, firms which have just created their variety (a mass $M_t - M_{t-1}$) are monopolists and charge a price

$$p_t^{\text{Monop}} = \frac{\varepsilon}{\varepsilon - 1}\frac{w_t}{z}.$$

As there is free entry for the creation of new varieties, in equilibrium, the cost to create a new variety must be equal to the profits which can be made by selling that variety, that is,

$$w_t n\left(M_{t-1}\right) = \left(p_t^{\text{Monop}} - \frac{w_t}{z}\right) c_t^{\text{Monop}} = \frac{1}{\varepsilon - 1}\frac{w_t}{z} c_t^{\text{Monop}}.$$

From this condition, we can deduce the quantity produced by a monopolist in equilibrium,

$$c_t^{\text{Monop}} = z\left(\varepsilon - 1\right) n\left(M_{t-1}\right).$$

Knowing the quantity produced by a monopolist also allows us to calculate how much a firm produces under perfect competition. Indeed, from the consumers' demand function, we know that for any varieties i and $i^{'}$, $\frac{c(i)}{c(i^{'})} = \left(\frac{p(i)}{p(i^{'})}\right)^{-\varepsilon}$. Therefore, we have

$$c_t^{\text{Comp}} = \left(\frac{\varepsilon - 1}{\varepsilon}\right)^{-\varepsilon} \varphi\left(\varepsilon - 1\right) n\left(M_{t-1}\right).$$

Finally, the labour market clearing condition gives, as in the standard Krugman model, the mass of varieties M_t. To derive it, note again that there are two categories of firms: every monopolist demands $n\left(M_{t-1}\right)$ units of labour to create its variety and $\frac{c_t^{\text{Monop}}}{z}$ units of labour to produce it, while every

[133] This simplifies the firm's problem at the moment of entry: the firm has to compare the entry cost to a one-period profit, and not to an infinite stream of profits (which would need to be discounted at some appropriate interest rate).

firm operating under perfect competition demands $\frac{c_t^{\text{Comp}}}{z}$ units of labour for production. Therefore, the labour market clearing condition is

$$\left(M_t - M_{t-1}\right)\left(n\left(M_{t-1}\right) + \frac{z\left(\varepsilon-1\right)n\left(M_{t-1}\right)}{z}\right) + M_{t-1}\frac{\left(\frac{\varepsilon-1}{\varepsilon}\right)^{-\varepsilon} z\left(\varepsilon-1\right)n\left(M_{t-1}\right)}{z} = L. \tag{154}$$

Equation (154) defines M_t as a function of M_{t-1}, and thereby closes the model. In the following, we assume that the function n takes the convenient form $n\left(M_{t-1}\right) = \frac{f_E}{M_{t-1}}$ (which implies, in particular, that there are positive knowledge spillovers: the more varieties there are, the less costly it gets to invent a new one). With this, the labour market clearing condition becomes

$$\frac{M_t - M_{t-1}}{M_{t-1}}\varepsilon f_E + \left(\frac{\varepsilon-1}{\varepsilon}\right)^{-\varepsilon}\left(\varepsilon-1\right)f_E = L.$$

Therefore, we get an expression for the growth rate of varieties, g ($g \equiv \frac{M_t - M_{t-1}}{M_{t-1}}$), which is

$$g = \frac{L}{\varepsilon f_E} - \left(\frac{\varepsilon}{\varepsilon-1}\right)^{\varepsilon-1}. \tag{155}$$

Two comments are in order.

1. How does the growth rate of varieties relate to the growth rate of consumption? Using the results above, one can show that $C_t = \left(M_{t-1}c_t^{\text{Comp}\frac{\varepsilon-1}{\varepsilon}} + \left(M_t - M_{t-1}\right)c_t^{\text{Monop}\frac{\varepsilon-1}{\varepsilon}}\right)^{\frac{\varepsilon}{\varepsilon-1}}$. After some algebra, this simplifies to $C_t = M_t^{\frac{1}{\varepsilon-1}} z\left(\varepsilon-1\right)f_E\left(1 + g\left(\frac{\varepsilon-1}{\varepsilon}\right)^{1-\varepsilon}\right)^{\frac{\varepsilon}{\varepsilon-1}}$. From this expression, it is clear that there is a one-to-one relationship between the mass of varieties and aggregate consumption, and that we have

$$\frac{C_t - C_{t-1}}{C_{t-1}} = \left(1+g\right)^{\frac{1}{\varepsilon-1}} - 1. \tag{156}$$

   So, if the growth rate of the mass of varieties increases, the growth rate of aggregate consumption increases as well.

2. How does the growth rate depend on the parameters of the model? From Equation (155), it is easy to see that growth increases with L. Larger economies grow faster for the exact same reasons that larger economies have more varieties in the Krugman model: they have a larger market, making innovation more attractive.[134] Furthermore, growth is obviously decreasing in

[134] This is sometimes called a "scale effect", and is common to many endogenous growth models. It has lead to many

the innovation cost parameter f_E. The effect of changes in ε is a priori indeterminate.

Now that we know the autarky equilibrium, we are finally ready to analyse the effects of trade on the growth rate in this model.

2.3 The effects of trade on growth

We will consider here a case of costless trade with an identical country (just as in the original Krugman model), and two sub-cases: one with and one without international knowledge spillovers.

No international knowledge spillovers In this first case, we assume that the n function is not affected by opening up to trade, and remains $n(M_{t-1}) = \frac{f_E}{M_{t-1}}$. That is, for domestic producers, invention gets easier when there is already a large mass of domestic varieties, but the mass of foreign varieties has no impact at all. In that sense, there are no international knowledge spillovers: inventions in one country do not make inventions in the other more or less costly.

In an equilibrium with costless trade, the prices charged by firms operating under perfect competition and by firms operating as monopolists are still given by the autarky formulas. However, the free entry condition for a new variety is now

$$w_t n(M_{t-1}) = 2\frac{1}{\varepsilon - 1}\frac{1}{z}c_t^{\text{Monop}},$$

because every monopolist exports, and as trade is costless and the two countries are identical, profits abroad are equal to profits at home. This implies that

$$c_t^{\text{Monop}} = \frac{1}{2}z(\varepsilon - 1)n(M_{t-1}).$$

and

$$c_t^{\text{Comp}} = \left(\frac{\varepsilon - 1}{\varepsilon}\right)^{-\varepsilon}\frac{1}{2}z(\varepsilon - 1)n(M_{t-1}).$$

From these equations, we can directly see that the total production of any domestic firm is exactly the same as in autarky (for instance, a monopolist now produces $2c_t^{\text{Monop}} = z(\varepsilon - 1)n(M_{t-1})$). As

discussions following a famous article by Jones (1995), who remarked that even though developed economies have become larger over time, productivity growth rates have not accelerated. Jones himself and others have proposed different changes to endogenous growth models in order to eliminate this scale effect.

a consequence, labour demands and the labour market clearing condition do not change, so that the growth rate of varieties is still given by

$$g = \frac{L}{\varepsilon f_E} - \left(\frac{\varepsilon}{\varepsilon - 1}\right)^{\varepsilon - 1}.$$

In this baseline model, opening up has therefore had no effect at all on the growth rate. This is the case because there are two effects at work, the market size effect (opening up makes it possible to export the variety and earn profits from this, which should stimulate innovation) and a competition effect (opening up lowers the profits of firms on their domestic market, which should disincentivise innovation). In this application, they cancel out exactly.

Note that this result depends on the specific assumptions that we made. With different preferences, or with trade costs, results change. Moreover, we have assumed that in order to create new varieties, firms need labour. Therefore, more trade does not create more resources for innovation (total labour supply is L before and after opening up), so changes in the growth rate can only be due to a change in the allocation of labour between production and innovation. An alternative model would be one in which new varieties could be created using units of consumption goods. In such a model, opening up does increase the resources available for innovation (because there are gains from trade), which makes it more likely that trade is growth-enhancing.[135]

Knowledge spillovers Now, let us change the assumptions on n, and analyse how this affects the equilibrium. In particular, assume $n(M_{t-1}) = \frac{f_E}{M_{t-1} + \lambda M^*_{t-1}}$, where M^*_{t-1} is the amount of foreign varieties consumed in the domestic economy and $\lambda \in [0, 1]$ is a parameter. Thus, now, there are international knowledge spillovers: a larger mass of foreign varieties makes domestic innovation less costly. The parameter λ governs how strong these spillovers are: when λ is close to 0, they are almost absent, but when λ is close to 1, foreign varieties reduce the cost of innovation just as much as domestic ones.

As there are no trade costs and the equilibrium is symmetric, $M^*_{t-1} = M_{t-1}$. Going through the same steps as in the previous paragraphs, we get

$$g = \frac{(1+\lambda) L}{\varepsilon f_E} - \left(\frac{\varepsilon}{\varepsilon - 1}\right)^{\varepsilon - 1}. \tag{157}$$

[135] Rivera-Batiz and Romer (1991) were the first to systematically study the differences between models in which innovation is produced with labour and models in which innovation is produced with consumption goods.

Therefore, with international knowledge spillovers, the growth rate increases when opening up, and the increase is stronger the higher the international knowledge spillover parameter λ is. In particular, if $\lambda = 1$, opening up has the same effect on growth than doubling the size of the economy.

This simple model has given us a better understanding of some of the mechanisms which can create a link between trade and growth rates. However, it cannot answer the question whether trade will increase or decrease growth in practice: by changing some assumptions of the model, its predictions may change completely. Therefore, the question how trade affects growth is ultimately an empirical one: we need to look at the data to see how strong the different mechanisms are.

3 The empirical evidence

Unfortunately (or fortunately for future researchers!), there is little empirical evidence on the link between trade and growth. In their recent survey, Grossman and Helpman write that "*[t]he theoretical literature identifies a number of different potential links between globalization and growth. Unfortunately, the empirics have not kept pace. We still know relatively little about which mechanisms are operative and what are their quantitative significance. There are several reasons for this. Empirical work on trade and growth is hampered by a dearth of natural experiments and by the limited number of observations we have on what might be considered "the long run." The cross-country regression methodology is flawed in this context, not only because there are many endogenous variables and few instruments, but also because trade implies that countries' experiences cannot meaningfully be treated as independent observations.*"

Because of these limitations, there is only very partial evidence so far. Pascali (2017), Feyrer (2009) and Feyrer (2019) propose instrumental variables to assess the effect of trade on national income, but they do not distinguish between static and dynamic effects. That is, they estimate that an x% increase in trade leads to a y% increase (or decrease) in income, but they cannot say whether this is due to static gains from trade or changes in growth rates.

Micro-level studies such as Lileeva and Trefler (2010), Bustos (2011) and others discussed in Chapter V, do show a positive effect of trade on innovation behaviour. However, they do not try to tell apart the different theoretical channels, and most importantly, they do quantify their results for the whole economy. Therefore, these papers tell us that some firms increase innovation after opening up, but not by how much we should expect overall innovation to go up (or whether there are general equilibrium mechanisms that overturn that result) and by how much we should eventually expect growth rates to increase. An exception from this is Bloom, Romer, Terry and Reenen (2014), which tries to quantify the impact of the entry of China into the WTO on the productivity growth rates of Western countries. They find that Chinese entry alone has increased the yearly level of growth rates in OECD countries by 0.2 percentage points, essentially because of market size effects.

Finally, there is some work on international knowledge spillovers, starting with Coe and Helpman (1995), who found a positive correlation between foreign R&D spending and domestic productivity, which they interpret as evidence for such spillovers. On the other hand, research on patent citations (Jaffe, Trajtenberg and Henderson, 1993; Murata, Nakajima, Okamoto and Tamura (2014)) tends to find that knowledge spillovers are very localized: in their patent applications, firms mainly build on past patents of firms located close by.