

C THE CAUSES OF TRADE

From an economic perspective, the case for freer trade rests on the existence of gains from trade and most economists typically agree that there are gains from trade. In recent years, however, free trade has increasingly come under fire and it is not uncommon to hear trade sceptics say that economists' arguments in favour of free trade and in particular comparative advantage may have been valid at the time of Ricardo (in the early 19th century) but that they are no longer valid in today's globalized world. This section critically assesses the relevance of economic theories of international trade in today's global trading environment. Most trade models are designed to answer two closely related questions: what goods do countries trade and why. While the main focus of this section is on the causes of trade, the discussion often touches upon the question of the patterns of trade.

This assessment of the relevance of trade theories is based on an overview of the theoretical models as well as of the empirical literature. This section begins by examining how robust the theories are and how far they can be generalized. This is an important part of the discussion – in particular, when the traditional approach is considered. This is because the traditional case for gains from trade is largely theoretical. In fact, it could even be argued, as Leamer and Levinsohn (1995) do, that “though obviously important and theoretically robust, the existence of gains from exchange is fundamentally a premise of economics, not a testable implication of a particular model”. Bearing this in mind, this section also reviews empirical work that tests trade theories and that attempts to estimate the relative importance of different types of gains from trade.

The idea that there are gains from trade is the central proposition of normative trade theory.¹ The gains-from-trade theorem states that if a country can trade at any price ratio other than its domestic prices, it will be better off than in autarky – or self-sufficiency.² More generally, the basic gains from trade propositions are that:³ i) free trade is better than autarky; ii) restricted trade (i.e. trade restricted by trade barriers) is better than autarky; and, iii) for a small country (i.e. a country too small to influence world prices) free trade is better than restricted trade.

Samuelson (1939) showed that there are potential gains from trade for small countries provided world

prices diverge from autarky prices. Kemp (1962) showed that restricted trade is better than no trade. He also extended the argument to the large country case, proving that free trade is potentially superior to autarky, in the case when there are many commodities and factors and with variable factor supplies. As noted by Deardorff (2005a), most treatments of the gains from trade say that if trade could potentially benefit all members of a country's population (assuming their preferences and income were identical), it is regarded as benefiting the country because some form of income redistribution among the country's consumers is assumed to be feasible. Beyond the feasibility of income redistribution in the form of lump-sum transfers (which is necessary to avoid market distortions associated with taxes), these results are based on a number of other key assumptions, notably constant returns to scale,⁴ perfect competition,⁵ no other market distortions, such as externalities,⁶ and the flexibility in the prices of factors of production (principally capital and labour) that ensure full employment. While the main message of the gains-from-trade theory remains valid when some of those assumptions are relaxed (for example, feasibility of lump-sum transfers), attempts to relax others (such as constant returns to scale) introduce significant complexities (Corden, 1984).

These basic propositions about the gains from trade, however, are not the end of the story. First, as pointed out by Corden (1984), the divergence between autarky and free trade prices is only an approximate explanation of the gains from trade. A full explanation of those gains should link them to the causes of trade – that is, to the elements that give rise to divergence between autarky and free trade prices. Those elements are the ones that lie behind the sources of comparative advantage. They would include differences in technology or differences in endowments. Second, economic theory points at other forms of gains from trade that are not linked to differences between countries. In particular, countries trade to achieve economies of scale in production⁷ or to have access to a broader variety of goods. Also, if the opening-up of trade reduces or eliminates monopoly power or enhances productivity, there will be gains from trade additional to the usual ones. Finally, trade may have positive growth effects.

This section covers the traditional gains from trade and their underlying causes, the gains from trade highlighted in the more recent trade theories, and the dynamic gains from trade. Each subsection starts with a brief presentation of a theory focusing on these specific gains from trade. The robustness of the theories to changes in their main assumptions is examined. Finally, the empirical evidence concerning the proposed rationales for international trade is reviewed.

Before considering the simplified theoretical frameworks (models) which focus on any particular source of gains from trade, it is important to emphasize that patterns of international trade typically reflect the interaction of several different causes. International trade theories and specific applications of the theories (models) should not be seen as mutually exclusive. This is of particular importance when trying to assess their relevance. The validity of a particular theory should be assessed on the basis of its capacity to explain trade in its limited domain. North-South trade might be explained by models which link trade patterns to differences between countries, while a model of monopolistic competition may best characterize trade between similar countries.

1. THE TRADITIONAL APPROACH: GAINS FROM SPECIALIZATION

Until recently, most trade models explained the commodity pattern of trade in terms of the law of comparative advantage.⁸ Before turning to particular models, such as the Ricardian model or the Heckscher-Ohlin model, which focus on particular product and/or country characteristics that determine the relative autarky price, it may be worth restating what comparative advantage means, and what it does and does not imply.

Comparative advantage is one of the most basic ideas in economics. Deardorff (1998) usefully distinguishes between the definition of comparative advantage and two versions of the law of comparative advantage. Comparative advantage can be defined as the “low relative cost of a good compared to other countries in autarky”. The double comparison across both goods and countries is the critical element. It indicates that it is impossible by definition for a country to have a comparative disadvantage in every good. In practice, every country will have a comparative advantage in something. There are two

laws of comparative advantage: one “positive” which predicts what countries can be expected to do and one “normative” which suggests what they should do. The positive version predicts that if permitted to trade, a country will export goods in which it has a comparative advantage. The normative version suggests that if permitted to trade, a country will gain through specialization.

Focusing on the normative side, the main contribution of the law of comparative advantage is to point to the fact that there are many more circumstances under which international trade is beneficial than most people appreciate. This can be illustrated using the example of an engineer and a nanny. Assume that the engineer is a good mother, better than the nanny at taking care of her child. The engineer, however, earns US\$ 500 an hour in her professional capacity while the nanny charges US\$ 12 an hour. Excluding from the question what is best for the child and fun for the mother, it makes economic sense for the engineer to pay the nanny to watch her child. As mentioned, the idea of comparative advantage is incorporated in several theories which are now considered.

(a) Differences in technology

As already mentioned, differences between countries are one of the main reasons why they engage in trade. The Ricardian model and its extensions point to technological differences as the source of comparative advantage. In order to keep the model as simple and the focus as clear as possible, a number of assumptions are typically made. One of these, i.e. that labour is the only factor of production, is specific to the Ricardian model. Most of the others, such as perfect competition, no trade costs, constant returns to scale, fixed endowments and international immobility of factors are standard in traditional trade models. With labour the only factor of production, differences in technology are modelled as differences in the amount of output that can be obtained from one unit of labour.

Using an example with two countries and two goods, Ricardo showed that even when one of the two countries has an absolute advantage in both lines of production, i.e. it can produce more output with one unit of labour in both sectors, there is scope for mutually beneficial trade if both countries specialize according to their pattern of comparative advantage. A country has a comparative advantage

in the production of good X if it is relatively more productive in the production of this good. More precisely, a country has a comparative advantage in the production of steel, for example, if the opportunity cost of steel in terms of the other good is less than in the other country.⁹ See Box 1 for a more detailed presentation of the Ricardian model.

The main results from the simple Ricardian model have been summarized by Deardorff (2005b):

“[...] comparative advantage can be usefully defined in terms of a comparison of relative autarky prices, which also represent marginal

opportunity costs in autarky. A difference in relative autarky prices, and thus the presence of comparative advantage, implies the potential to increase world output by reallocating resources within the two countries. Combined with market structures of perfect competition, comparative advantage also implies that unless policies interfere with market incentives, countries stand to gain from trade in the sense that at least one country will gain and neither will lose. And this gain from trade is achievable only if countries each export the good in which they have comparative advantage.”

Box 1
A numerical presentation of the Ricardian model

Along the lines of Ricardo’s own presentation of his model in 1817, a simple numerical example with two countries (A and B), two goods (logs and iron bars) and one single input (labour) can be used to illustrate how countries can gain from trade through specialization according to comparative advantage based on differences in technology.

Technology in each of the two countries A and B is summarized by labour productivity in the production of logs and iron bars. Labour productivity is expressed in terms of unit labour requirements. Labour productivity in the log industry in Country A, for example, noted a_{AL} , is the number of hours of labour required to produce one unit of log. The table below illustrates unit labour requirements in countries A and B.

Unit labour requirements

	Logs	Iron bars
Country A	$a_{AL}=1$	$a_{AI}=3$
Country B	$a_{BL}=4$	$a_{BI}=4$

In this example, unit labour requirements for both industries are lower in Country A than in Country B, which means that labour productivity is higher in A in both industries. Thus, Country A has an absolute advantage in both industries. From looking at absolute advantage, it can be concluded that there is no scope for mutually beneficial trade between A

and B. How could producers in B compete with those in A if they are less efficient ? Ricardo suggested that what matters is not absolute but comparative advantage. In this example, the ratio of the labour required to produce one iron bar to that required to produce one log in Country A ($a_{AI}/a_{AL}=1/3$) is lower in Country A than in Country B ($a_{BI}/a_{BL}=1$). This amounts to saying that Country A has a comparative advantage in the logging industry. The flipside of this is that Country B has a comparative advantage in the production of iron, as the ratio of the labour required to produce one bar of iron to that required to produce one log in Country B ($a_{BI}/a_{BL}=1$) is lower than in Country A ($a_{AI}/a_{AL}=3$).

Comparative advantage can also be established using the notion of opportunity cost. A country is said to have a comparative advantage in the production of a particular good if the opportunity cost of producing that good in terms of the other good is lower in that country than it is in the other countries.¹⁰ The opportunity cost of one log is defined as the number of iron bars the economy would have to give up producing in order to produce an extra log. Producing an extra log would require $a_{AL}=1$ unit of labour, which could have been used to produce $1/a_{AI}=1/3$ of an iron bar. The opportunity cost of iron in terms of logs in Country A is thus $a_{AL}/a_{AI}=1/3$, compared with $a_{BL}/a_{BI}=1$ in Country B. With constant unit labour requirements, these opportunity costs are constant.

In the absence of trade, the relative prices of logs and iron bars in each country would be determined by the relative unit labour requirements. In Country A, the relative price of logs would be $P_{AL}/P_{AI} = a_{AL}/a_{AI} = 1/3$. In B it would be $P_{BL}/P_{BI} = a_{BL}/a_{BI} = 1$. Opening up for trade between A and B allows producers in A to sell logs at a higher price in B, while producers in B start selling iron bars in A. If $P_{AL}/P_{AI} > a_{AL}/a_{AI}$, wages in the logging industry will be higher than in the iron industry, that is $P_{AL}/a_{AL} > P_{AI}/a_{AI}$. Workers will wish to work in the higher-wage industry and thus the economy will specialize in the log industry. Eventually trade will equalize the relative prices in A and B. It can be shown that the normal result of trade is that the price of a traded good relative to that of another good ends up somewhere between its autarky prices. In this case, the new relative price of logs will be in the range between $1/3$ of an iron bar and 1 iron bar.

This pattern of specialization and trade produces gains from trade. Trade can be seen as an indirect method of production that is more efficient than the direct method. In direct production of iron in A, one hour of labour produces $1/3$ of an iron bar. If, for instance, the after trade relative price of a log in A is $1/2$ of an iron bar, the same hour of labour can be used to produce 1 log which can then be traded against $1/2$ of an iron bar. Similarly, in B, one hour of labour would produce $1/4$ of a

log in direct production, while the same hour could be used to produce $1/4$ of an iron bar which could then be traded against $1/2$ of one log. Both countries clearly gain from trade.

It is interesting to examine what this simple example tells us about relative wages. After specialization, which happens to be complete in this simple model, Country A produces only logs. Hourly wage in A must be 1 log, as one hour of labour produces one log in A. Similarly, hourly wage in B must be $1/4$ of an iron bar since it takes four hours of labour to produce one bar. Assuming that the price of logs is 10 dollars per unit while that of an iron bar is 20 dollars per unit, which corresponds to the terms of trade, then hourly wage in A is 10 dollars while hourly wage in B is 5 dollars ($1/4$ of 20 dollars). The relative wage of workers in A is $10/5 = 2$. Note that this result only depends on the level of productivities and the relative prices. It does not depend on the absolute price of a log or of a bar. The fact that the relative wage lies between the ratio of the two countries' productivities in logging (where A is twice as productive as B and the same ratio in the iron industry, where A is only about 1.3 times more productive) explains why trade is profitable for both countries. In logs, A can compensate its higher wage with its higher productivity while in iron, B can compensate its lower productivity with its lower wage rate.

The simplified two-goods, two-countries presentation of the Ricardian model often fails to convince non-economists who ignore how far it can be generalized and who question its validity in today's world. It is, therefore, worthwhile to examine the robustness of its main results to changes in some of the underlying assumptions. A distinction needs to be made between the robustness of the law of comparative advantage on the one hand, which is not an exclusivity of the Ricardian model, and the idea that comparative advantage is rooted in technological differences on the other hand.

One of the main differences between the Ricardian model and other trade models is the assumption that marginal costs do not change with the level of production. An important ramification of the constant costs assumption is that it implies complete specialization of the trading partners, which is not necessarily realistic. The problem is that with non-constant marginal costs, comparative costs are not uniquely defined. As the discussion of other trade models will show, however, with non-constant costs, countries' behaviour is not very different except that trading countries can continue to produce both goods.

Box 2 Ricardian models with more goods and more countries

This box examines how the Ricardian model functions when the two-goods and two-countries only assumption is relaxed.

With N goods and two countries, the Ricardian model can be generalized using the concept of chain of comparative advantage. Numbering the goods in order of Country A's relative labour requirements:

$$\frac{a_{A1}}{a_{B1}} < \frac{a_{A2}}{a_{B2}} < \dots < \frac{a_{AN}}{a_{BN}}$$

where a_{Ci} is Country C's unit labour requirement for good i , it can be shown that Country A will produce all of the goods for which:

$$\frac{a_{Ai}}{a_{Bi}} < \frac{w_B}{w_A}$$

while B will produce all the ones with a ratio of unit labour requirement larger than the ratio of wage rates.¹¹

The Ricardian model can also easily be generalized to more than two countries if the number of goods is kept at two. A ranking similar to the one in the multi-goods case can be constructed over the countries' relative labour requirements:

$$\frac{a_{11}}{a_{12}} < \frac{a_{21}}{a_{22}} < \dots < \frac{a_{M1}}{a_{M2}}$$

where a_{ij} is Country i 's unit labour requirement for good j , where $i=1..M$ and $j=1,2$. In this case, all exporters of good 1 will lie to the left of all exporters of good 2.

The Ricardian model, however, does not generalize as easily to more than two goods and more than two countries simultaneously. Deardorff (2005b) discusses a number of attempts to come up with strong generalizations in the multi-goods, multi-country case. His conclusion is that comparative advantage only predicts trade patterns in simple cases.

To move closer to reality, it is important to consider how the Ricardian model functions when the two-goods-only assumption is relaxed. As explained in Box 2, with multiple goods and multiple countries, the Ricardian model only predicts trade under strong simplifying assumptions. In models with more realistic assumptions, such as trade barriers, intermediate inputs, and large numbers of both countries and goods, it fails to do so.¹² This does not mean that the law of comparative advantage is useless under realistic assumptions. In the more realistic models, comparative advantage continues to predict and explain gains from trade. Even if, as discussed in the introduction, part of the gains from trade results simply from perfect competition, comparative advantage also plays a role. While the basic gains-from-trade theorem indicates that free trade improves a country's welfare if the prices it faces with trade diverge from autarky prices, comparative advantage provides the reason why prices with trade differ from those in autarky and thereby ensures positive gains from trade.

Comparative advantage may not allow strong generalizations under more realistic assumptions,

but it may allow weak generalizations. Indeed, instead of indicating whether any particular good will be exported or imported by any particular country, comparative advantage can provide average relationships such as, for instance, that the trade-weighted average of the country's autarky prices of goods it exports, relative to world prices, is less than the trade-weighted average of the relative prices of its imports (Deardorff, 2005b). Along the same lines, Deardorff (1980) formalizes such average relations in the form of correlations. For instance, he derives a negative correlation between autarky prices and quantities of net exports across all goods and countries.

Having derived the more general correlations, it is interesting to examine how robust they are. Deardorff (2005b) discusses a number of assumptions, distinguishing between those that are consistent with comparative advantage correlations, including gains from trade, and those that are not. Starting with the ones that are consistent, he notes that both the gains from trade and the average relationships continue to hold in the presence of restrictive trade policies as well as with transport

and other real trade costs. The correlations also hold for all types of goods (final, intermediate or both) and even for services. Differentiated products can be accommodated as long as markets are perfectly competitive. Also, the correlations remain valid for all sorts of preferences. The two main assumptions on the other hand that cause problems for the theory of comparative advantage, both as a source of gains from trade and as a predictor of patterns of trade, are domestic distortions caused by externalities or market power, for instance, and increasing returns. These assumptions do not reverse the story but rather complicate it.

(b) Differences in resource endowments

The Ricardian model assumes that labour is the only factor of production. Under this assumption, the only possible source of comparative advantage is differences between countries in labour productivity. Clearly, differences in labour productivity are not the only source of comparative advantage. Differences in resource endowments must play a role. Countries that are relatively better endowed with fertile land than others are likely to export agricultural products. The idea that international trade is driven by differences between countries' relative factor endowments is at the heart of the Heckscher-Ohlin model. This model, named after the two Swedish economists – Eli Heckscher and Bertil Ohlin – who developed it, is probably the most influential model of international trade. The Heckscher-Ohlin model provides an alternative explanation of trading patterns. Because it takes into account more than one factor, it also has implications for the internal distribution of income. The gains from trade in the Heckscher-Ohlin framework, however, are of the same nature as in the Ricardian model. They are gains from specialization that arise because of differences between countries. The Heckscher-Ohlin model only focuses on another source of comparative advantage.

The standard version of the Heckscher-Ohlin model assumes that there are two countries, two goods and two factors of production. It also assumes that technologies and tastes are identical across countries, that factor endowments differ and that factors are mobile between industries but not between countries. Under those assumptions, four core propositions can be derived:¹³

1. The Heckscher-Ohlin theorem states that a country has a production bias towards, and hence tends to export, the good which uses intensively the factor with which it is relatively well endowed.
2. The Stolper-Samuelson theorem states that an increase in the relative price of one of the two goods raises the real return of the factor used intensively in producing that good and lowers the real return of the other factor.
3. The Rybczynski theorem states that if goods prices are kept constant, an increase in the endowment of one factor causes a more than proportionate increase in the output of the commodity which uses that factor relatively intensively and an absolute decline in the output of the other commodity.
4. The factor-price equalization theorem states that, under certain conditions, free trade in final goods is sufficient to bring about complete international equalization of factor prices.

Using a simple example with two countries – A (assumed to be well endowed with labour) and B (assumed to be relatively rich in capital) – and two goods (automobiles, the production of which is assumed to require relatively more capital, and clothing, that requires more labour), the four propositions can be illustrated in the following way. The Heckscher-Ohlin theorem tells us that A exports clothing and imports automobiles. The Stolper-Samuelson theorem tells us that a tariff on clothing (more likely in B, which imports clothing) would raise real wages and reduce real return on capital. The Rybczynski theorem tells us that immigration would raise the output of clothing more than proportionately and reduce the output of cars. Finally, the factor-price equalization theorem tells us that even without allowing for international mobility of labour and capital, trade alone would, under certain conditions, equalize wages in A and B and rates of return on capital in A and B.

Again, the question arises whether the core propositions that have been derived in the standard basic model can be generalized. This question is important because together with the law of comparative advantage, the four core propositions can be seen as the central body of international trade theory. Among the extreme assumptions

which underpin the core results are that of low and even “dimensionality”. The sensitivity to higher dimensions of the basic propositions, because it is a key issue for the practical relevance of the dominant trade theory, has been an area of active research since the 1940s. The two-goods, two-factors model is special not only because of the assumption regarding the number of goods but also because this number of goods equals the number of factors.

Economists have analyzed all possible cases: those with an even number of goods and factors, those where the number of goods is larger than the number of factors, those where the number of factors is higher, and finally the general case with N goods and M factors. Several authors have surveyed this large volume of theoretical work.¹⁴ Their conclusions are relatively nuanced. In general, dimensionality matters in the sense that many of the results from the basic 2×2 model are lost with higher dimensions. Generalizations run into difficulties in all cases, even or uneven. Ethier (1984) nevertheless optimistically concludes that the basic messages of elementary theory still come through to a relatively large extent. Like the law of comparative advantage, the Heckscher-Ohlin theorem survives as a correlation or in an average sense, while the Stolper-Samuelson and Rybczynski theorems survive in undiluted strength but they only apply to some factors or goods but not necessarily to all.

As mentioned above, a number of other assumptions underpin the Heckscher-Ohlin theory. The ramifications of those regarding intersectoral and international factor mobility as well as of those regarding the nature of the products traded are discussed below. Models with economies of scale, imperfect competition and differentiated products are considered in sub-sections 2 and 3. Another important assumption of the model is that factor markets are perfect. Realizing that factor market imperfections can be significant in some countries, economists have examined the effect of three major types of distortions: wage differentials, generalized “sticky”¹⁵ wages and sector-specific sticky wages. These distortions introduce various types of “pathologies”.¹⁶ Brecher (1974), for instance, shows that the minimum wage country levels of employment and welfare may be less with free trade than with no trade. This would be the case if trade leads the minimum wage country to export the capital-intensive good under incomplete specialization. If, on the contrary, free trade leads

the home country to export the labour-intensive good, employment and welfare increase regardless of the degree of specialization.

(c) Empirical evidence

In the introduction to their 1995 review of empirical evidence on international trade theory, Leamer and Levinsohn (1995) note that “international microeconomics is primarily a theoretical enterprise that seems little affected by empirical results”. In their view, the reason for this is neither a lack of empirical work by economists, nor a lack of appropriate data. Rather, their review is premised on the idea that economists “have not done the job right”. Why is that? In his earlier survey of empirical tests of trade theories, Deardorff (1984) identifies the difficulty of constructing sound theoretical tests of trade theories as the major obstacle to their testing. This difficulty, in his view, arises from the nature of the theories themselves, which “are seldom stated in forms that are compatible with the real world complexities that empirical research cannot escape”. It is not clear what the Heckscher-Ohlin model in its standard form with two goods, two countries and two factors tells us about the real world where there are many of all three and it therefore has been difficult to agree on a valid test.

While progress has been relatively limited with regard to the testing of trade theories, there have been some improvements in empirical applications of these theories. The available evidence, though it does not prove much, sheds some light on the factors that contribute most to the understanding of international trade. This sub-section provides a brief overview of empirical work on the traditional models of international trade. It first considers evidence regarding gains from trade and comparative advantage and then summarizes the main results of empirical tests of the Ricardian and Heckscher-Ohlin models.

Very little is known about the empirical magnitudes of the gains from international trade and the mechanisms that generate these gains. In particular, very limited evidence is available on how much specialization according to comparative advantage can contribute to an economy’s overall income. This may come as a surprise given the flurry of estimates of gains from trade liberalization obtained through the use of Computable General Equilibrium (CGE) models. However, while CGE models can be a very

useful tool for policy analysis, they do not provide hard evidence on the gains from trade. This is because CGE models are typically “theory with numbers” in the sense that they rely on a number of behavioural and other assumptions and offer assessments of potential gains from trade.

A relatively recent study by Bernhofen and Brown (2005) provides the first piece of hard evidence on the magnitude of the static gains from trade resulting from comparative advantage. The specificity of Bernhofen and Brown’s study is that it embeds the analysis of the gains from trade within a theoretical framework that also identifies the underlying cause of international trade. They use Japan’s 19th century trade liberalization as a natural experiment to estimate the effects of trade on national income. They first provide supportive evidence that Japan’s trading pattern after its opening up was governed by the law of comparative advantage and then take the next step and estimate the gains from trade resulting from comparative advantage. They estimate that at most the gain in real income was 8 to 9 per cent of GDP.

Irwin (2001) uses another of the few historical examples where a country has moved from self-sufficiency – or autarky – to free trade or vice versa rapidly enough to allow the use of time series data to estimate the gains from trade. He calculates that the welfare cost to the United States of the nearly complete embargo imposed by the US Congress on international trade between December 1807 and March 1809 was some 5 per cent of GDP. This cost, however, does not represent the total gains from trade because trade was restricted in the pre-embargo situation.

Bernhofen and Brown’s work on Japan is remarkable because it provides the first and to our knowledge only direct test of the theory of comparative advantage. Direct testing of the theory of comparative advantage is notoriously difficult because it involves relating trade flows and specialization patterns to autarky prices which, by their nature, are almost always unobservable. Bernhofen and Brown (2004) test a weak formulation of the law of comparative advantage using the natural experiment of Japan’s opening up to trade in the 1860s.¹⁷ They carefully verify that Japan in the mid-19th century met the requirements needed to apply the theory. In particular, they show that before 1854 Japan was completely closed to trade while by the late 1860s it had fairly free trade and no export subsidies.

Their results provide a strong empirical case for the prediction of the theory.¹⁸

If direct tests of the law of comparative advantage are so difficult, what about testing the theories that explain comparative advantage? As explained above, the Ricardian model attributes comparative advantage entirely to differences in labour requirements of production. Unfortunately, testing the Ricardian model turns out to be as problematic as testing the law of comparative advantage. The main problem is that the Ricardian link between trade patterns and relative labour costs is much too sharp to be found in any real data set. Because of the complete specialization that the model implies, for instance, relative labour requirements ought to be unobservable. Deardorff (1984) discusses tests of a weaker link and concludes that they are deficient. Overall, while the Ricardian model can be seen as an important reminder that technological differences can be a source of comparative advantage, the one-factor model is too simple to study the impact of technologies on trade flows (Leamer and Levinsohn, 1995).¹⁹

The literature on testing and estimating Heckscher-Ohlin models is both voluminous and complex. While an exhaustive and systematic overview of this literature clearly falls beyond the scope of this Report, the following provides a quick summary of its main results.

Leontief (1953) is the earliest and probably the best known attempt to confront the Heckscher-Ohlin model with data. Given the United States’ relatively high capital-labour endowments ratio compared with other countries, in particular in the late 1940s, the Heckscher-Ohlin model would predict that the United States exported capital-intensive goods and imported labour-intensive goods. Surprisingly however, comparing the amount of factors of production used to produce US\$1 million worth of exports with the amount used to produce the same value of US imports, Leontief found that US exports were less capital intensive than US imports. This result, which contradicted the Heckscher-Ohlin theorem, came to be known as the Leontief paradox. A wide range of explanations were offered for this paradox, of which several concerned the fact that Leontief focused only on two factors of production, ignoring land and human capital.²⁰ In the following years, a number of studies redid the analysis, taking into account those factors.

The paradox persisted in the data from the earlier decades but seems to have disappeared since the early 1970s (Deardorff, 1984). Leamer (1980) provided the definitive critique of the Leontief paradox. He showed that Leontief had performed the wrong test. Even if the Heckscher-Ohlin model is true, the capital/labour ratios in exports and imports need bear no particular relationship to relative factor endowments if trade is unbalanced.

Leontief (1953) may be interpreted as an application of the so-called “factor content” version of the Heckscher-Ohlin theorem. Empirical application of the theorem has been of two forms, corresponding roughly to two versions of the theorem. The “commodity version” says that countries tend to export those goods which use relatively intensively their relatively abundant factors of production. The “factor content” version developed by Vanek (1968) (also termed the Heckscher-Ohlin-Vanek theorem), says that countries will tend to export the services of their abundant factors, embodied as factor content in the goods they trade. The test performed by Leontief was a partial test of the “factor content” version (Feenstra, 2004).

The first complete test of the “factor content” version of the Heckscher-Ohlin theorem was by Bowen et al. (1987). For a sample of 27 countries and 12 factors of production, they showed that the test failed. Their negative result was confirmed by other authors. Researchers then began to examine which parts of the theory were causing the problems.²¹ Building on this work, Davis and Weinstein (2001) show that with a few simple modifications, the Heckscher-Ohlin-Vanek model is consistent with data from ten OECD countries and a rest-of-the-world aggregate. These modifications include, in particular, the introduction of cross-country differences in technology, a breakdown of factor price equalization, the existence of non-traded goods, and costs of trade.

A number of issues have been left unresolved by Davis and Weinstein (2001). First, researchers are currently looking into extending the range of countries used for the tests (Feenstra, 2004). Second, trade in intermediate products needs to be adequately distinguished from trade in final goods. Third, technological differences have been shown to be a major determinant of trade patterns and their underlying causes should be identified. Fourth, researchers are investigating the role of the

integrated equilibrium assumption and factor price equalization (Davis and Weinstein, 2000).²²

In summary, most of the empirical work that attempted to test or estimate Heckscher-Ohlin models used inappropriate methods and is therefore largely irrelevant. Complete tests failed under the conventional assumptions of identical tastes and identical technologies with factor price equalization across countries. In recent years, however, studies using appropriate methods have shown that if technological differences and home bias are included in the model and if the assumption of an integrated world is relaxed, there appears to be a substantial effect of relative factor abundance on the commodity composition of trade. As pointed out by Feenstra (2004), recent work has been more about accounting for global trade flows than about testing hypotheses related to trade but it certainly has the merit to highlight the fact that there are multiple causes for trade. As the next sub-sections will show, economies of scale, product differentiation, or imperfect competition all play important roles.

(d) Intermediate inputs, services, tasks and fragmentation

As discussed in more detail later in this section, the most important development in world trade in the last few years has been the acceleration of the fragmentation of production of both goods and services and the associated development of foreign outsourcing and offshoring. Because the fragmentation of production involves trade in intermediate products and services, their role in international trade is viewed as increasingly important. This sub-section considers whether the principal results of the traditional theory of trade still hold in the presence of fragmentation, outsourcing and offshoring involving intermediate inputs and services.

i) Intermediate inputs

Deardorff (2005c) examines the role of comparative advantage in a Ricardian trade model with intermediate inputs. He finds that only an average relationship between comparative advantage and trade seems to be at all robust. The gains from trade, however, are unambiguous in these Ricardian models, with imported inputs actually providing an additional source of gain from trade. Deardorff (1979) shows that similar results hold in the

Heckscher-Ohlin case. With intermediate inputs, a trade barrier on an input that raises its price can make production of the corresponding final good too costly to survive, even though the country might otherwise be a relatively low-cost producer of the final good. Kemp (1964) shows that the Stolper-Samuelson and the Rybczynski theorems still hold in the presence of traded intermediate products. In a model where each final good can be used as intermediate input in the production of the other final good, Schweinberger (1975) shows the conditions under which the Heckscher-Ohlin theorem holds.

ii) Services

Hindley and Smith (1984) consider the question of the applicability of the normative theory of comparative cost to the services sector. They discuss two potential difficulties in applying this theory to trade in services: the pervasiveness of regulations and licensing in services industries and the fact that services can be traded in different modes. They argue that none of these potential difficulties appears to yield any prior reason to suppose that the theory does not apply. In their words, “services are different from goods in ways that are significant and that deserve careful attention, but the powerful logic of the theory of comparative advantage transcends these differences”. In other words, there is no reason to have any doubt on the potential for countries to gain from free trade in services.

Deardorff (1985) focuses on the positive issue of whether trade in services conforms to a pattern that is explainable by comparative advantage. He looks at three different characteristics of trade in services and considers in each case what they suggest for the validity of the principle of comparative advantage. The first of these characteristics is that traded services often arise as a by-product of trade in goods. The second is that trade in services frequently requires or is accompanied by international direct investment. The third is that while goods can be produced elsewhere from where they are consumed, services cannot. He argues that while the first two of those characteristics do not undermine the usefulness of the law of comparative advantage in explaining trade, the third raises a number of issues.²³ In the third case, he uses a model that is like the standard Heckscher-Ohlin model except that one of the two goods is a service that must be produced where it is consumed and one of the factors

is "management" which can contribute to services production "in absentia". In this case, no version of the principle of comparative advantage is generally valid. Depending on the specific assumptions, weak versions of the law may apply.

Melvin (1989) includes capital services as a tradable in a Heckscher-Ohlin framework and shows that, contrary to the view of Hindley and Smith (1984), the introduction of services does require a different approach, which necessitates the reinterpretation of the law of comparative advantage.²⁴ If the tradable commodity uses the mobile factor service intensively, the country well endowed with capital will import the capital-intensive good, even though the relative price of this good was lower in this than in the other country in autarky. This result, at first glance, seems to contradict the law of comparative advantage and the Heckscher-Ohlin theorem. However it conforms with comparative advantage, as interpreted by Deardorff (1980), in the sense that it predicts that the country well endowed with capital exports capital services and imports the labour-intensive commodity. The Heckscher-Ohlin theorem also holds, for while the country which is well endowed with capital imports the capital-intensive commodity Y, it exports capital services, which are more capital-intensive than any good. One important implication of this model is that a service-exporting country will be observed to have a merchandise trade deficit. Such deficits, the author argues, would just reflect the country's comparative advantage in the service sector.²⁵

Deardorff (2001) argues that for many services, the benefits from liberalization extend beyond the traditional gains from trade liberalization. Many services play a critical role of facilitating international trade in goods and other services. Trade liberalization for those services can yield benefits by facilitating trade in goods that are larger than might be expected from analysis of the services trade alone. Deardorff's paper explores this idea using simple theoretical models to specify the relationships between services trade and goods trade. Services industries, such as transportation, insurance and finance, provide inputs needed to complete and facilitate international transactions in goods. Measures that restrict trade in those services create costs that limit the international flow of trade in goods. By reducing these costs, liberalization can stimulate international trade of goods.

Supportive evidence is provided by Blyde and Sinyavskaya (2007). They match goods data from the United Nations Commodity Trade Statistics Database (COMTRADE) with International Monetary Fund (IMF) Balance of Payments services data to investigate empirically the relationship between trade in services and trade in goods. They find that trade in services is important to facilitate trade in goods in all the 2-digit SITC goods categories. Investigating which types of trade in services are more important for international trade in goods, they find that trade in transportation and communication services generate the largest impact on trade in goods. Insurance, business and travel services are found to generate positive impact on the international trade of only certain types of goods. Lennon (2006) finds some evidence of complementarity between trade in goods and trade in services. Bilateral trade in goods explains bilateral trade in services: the resulting estimated elasticity is close to unity.²⁶ Likewise, bilateral trade in services has a positive effect on bilateral trade in goods: a 10 per cent increase in trade in services raises traded goods by 4.58 per cent.

iii) Trade in tasks and fragmentation

Revolutionary advances in transportation and (especially) communications technology have enabled an historic break-up of the production process by making it increasingly viable and profitable for firms to undertake different production stages in disparate locations.²⁷ This has resulted in offshoring of both services and manufacturing sector jobs and rapidly growing trade in intermediate products or tasks (see Box 3). This phenomenon has variously been called fragmentation, unbundling, offshoring,

vertical specialization, slicing-up of the value-added chain or trade in tasks. It will be considered in more detail in Section D. This sub-section only discusses how fragmentation has been integrated in traditional trade models and how it affects the main results of those models.

Two main approaches to the modelling of fragmentation can be distinguished. The first approach is to model fragmentation as trade in intermediates based on comparative advantage. The main insight is that offshoring is similar to technical progress in the production of final goods. Consider a world with two nations, Home (H) and Foreign (F), one final good (X) and one single production factor (labour). The production of X involves two tasks, 1 and 2, which are produced with labour.²⁸ Assume H has a comparative advantage in task 1 and F has a comparative advantage in task 2. With free trade in tasks, H specializes in the production of task 1, F specializes in the production of task 2. Specialization allows more of the final good to be produced (and consumed) in both countries (standard static gains from production and consumption efficiency). Since more of the final good can be produced with the same amount of primary factors, fragmentation is akin to technological progress in the final good. In other words, offshoring increases labour productivity, expressed as output of the final good per hour worked, in both nations.

Deardorff (2005a) examines in more detail the effect of fragmentation on traditional gains from trade in this first approach. He models fragmentation as the possibility to split a productive activity into parts that can be performed in different locations, much like a new technological possibility that becomes available

Box 3

Tasks, services and intermediate goods

It is important to point out that trade in tasks is potentially encompassing both trade in services and/or trade in intermediate goods.

Tasks can be classified as follows:

1. analytical tasks;
2. interactive tasks;
3. routine cognitive tasks;
4. routine manual tasks;
5. non-routine manual tasks.

If there is trade in tasks 1-3, this will be classified as trade in services. If there is trade in tasks 4-5, this will be classified as trade in goods, because it implies the sourcing of physical inputs (intermediates) produced abroad. Hence, trade in tasks can involve both trade in services and trade in goods.

Source: Spitz-Oener (2006).

to a country or to the world. Fragmentation, as he understands it, involves offshoring and thus trade of services. His conclusions about the gains from fragmentation are similar to the conclusions of trade theory about the gains from trade. Cases can be identified where fragmentation lowers the welfare of particular countries. If, for instance, fragmentation causes a change in relative world prices, it is possible that one country's terms of trade worsen to such an extent that it is made worse off, despite the new technological ability that fragmentation represents. Similarly, if fragmentation interacts negatively with existing distortions, such as tariffs, it can lower the welfare of particular countries and even of the world as a whole. However, on average, fragmentation is likely to expand world welfare because it will systematically expand what the world is able to do potentially with its given resources.

The second approach to fragmentation has been introduced recently by Grossman and Rossi-Hansberg (2006b). They present a theory of offshoring, or trade in tasks, which they refer to as a "new paradigm". Because their main contribution relates to the effect of fragmentation/offshoring on wages and distribution, it will be discussed in more detail in Section E of this Report.²⁹ The discussion here focuses on the linkages between the "new paradigm" and traditional trade theory. The main result is that in addition to comparative advantage gains from trade, fragmentation has a welfare-enhancing productivity effect on wages in the offshoring country, according to Grossman and Rossi-Hansberg.³⁰ A main difference between their approach and the first approach of fragmentation is that they factor in that a firm with better technology can use this technology abroad. There are also task-specific offshoring costs that are best understood as the communication and organizational costs that a firm pays when it sources the performance of a task abroad. The advantage of offshoring a task is that the firm combines its superior technology with cheap foreign labour when the task is performed abroad.³¹

To understand the thinking behind the model, consider two countries, North and South. Firms in North have superior technology. Wages are higher in North than in South because they are tied to technologies. North firms are interested in combining their superior technology with cheap labour in South. They will offshore a task if the initial wage gap is larger than the offshoring costs. The wage in South is assumed to remain

constant. The reason for this is that South firms are assumed to continue producing the final good using South technology which keeps the wage at the low level. The wage in North will increase because productivity increases. Productivity increases because offshoring releases domestic workers who can focus on the tasks where they have a trade-cost-adjusted comparative advantage. This productivity effect is independent of comparative advantage based on tasks. For the offshoring country, it comes in addition to the Ricardian gains from trade that existed in the first approach.

(e) Factor mobility

So far, models have been considered where the factor(s) of production are assumed to be mobile between industries but not between countries. In this sub-section, these assumptions are relaxed and consideration is given to how the gains from trade and comparative advantage results are affected. The assumption that there is no movement of factors of production between countries is maintained but the assumption of perfect factor movement between industries is further qualified. This sub-section ends by considering how traditional trade models take account of international mobility.

i) Internal mobility

The specific factors model assumes that an economy produces two goods using three factors of production in a perfectly competitive market. Two of the three factors of production, typically land and capital, are assumed to be sector-specific, which means that they can be used only in the production of a particular good, while the third, typically labour, is common to both sectors. Since mobility of factors in response to any economic change is likely to rise over time, the specific factors model can be interpreted as capturing medium-term effects and the models with perfect movement between industries as representing the long-term effects.

A number of interesting results – in particular, regarding the distributional effects of trade – can be derived from the specific factors model, which was used extensively prior to the ascendancy of the Heckscher-Ohlin model. Because there is only one factor that is used in both sectors, the allocational problem in the specific factors model is relatively simple. The wage rate and the equilibrium allocation of labour can be found by setting the

sum of labour demand in each sector equal to the available supply of labour. The wage rate can then be used to determine the rental rate of the two specific factors.

While the gains from trade result remains valid in the specific factors model, there are some issues with the law of comparative advantage and the effect of changes in prices or endowments that are different here from what they are in the Heckscher-Ohlin model.

First, trade produces overall gains in the limited sense that those who gain could in principle compensate those who lose while still remaining better-off than before.³² Second, as already mentioned, in a two-sector, multi-factor world, comparative advantage will not be an infallible predictor of a country's trade pattern. As demonstrated by Falvey (1981), however, while the statement that “a country will export those commodities in which it has a comparative advantage” is no longer a theorem, it appears to be a useful presumption, even in a multi-factor world. Third, the implications of the specific factors model are quite different from those of the Heckscher-Ohlin model. In the specific factors model, an increase in the price of a good raises the real return to the specific factor in that sector, lowers that to the other specific factor, and has an ambiguous effect on the real return to the mobile factor.³³ An increase in the endowment of a factor specific to a sector leads to a less than proportionate increase in the output of that sector and a decline in the output of the other sector. The return of the mobile factor rises, while those to sector-specific factors decline. An increase in the endowment of the mobile factor lowers the return to that factor and increases those to specific factors. Outputs of both sectors rise.

The specific factors model has been much neglected empirically (Leamer and Levinsohn, 1995). Grossman and Levinsohn (1989) provide some evidence suggesting that capital is sector-specific while Kohli (1993) finds that a sector-specific structure is broadly consistent with data for the US economy.

ii) International factor mobility

From an economic point of view, trade in factors is much like trade in goods. It is driven by international differences in resources and is beneficial in the sense

that it increases world production. The focus here, however, is not on explaining factor movements but rather on the interactions between trade in goods and factor mobility. A major and strong assumption in the models discussed so far is that factors of production cannot move between countries. In this sub-section, this assumption is relaxed and consideration is given to how this affects the law of comparative advantage and the validity of some of the main trade theorems. Trade literature has focused on capital movements, probably because labour is considered less mobile at least in the short term. However, some of the results would in principle apply to any factor.

The idea that trade is a substitute for factor movements dates back to the early 20th century and has been expressed by a number of eminent economists. This idea is based on the factor endowment theory of international trade elaborated by Heckscher and Ohlin. According to this theory, trade in goods is caused by differences in factor endowments between countries. Thus, on the one hand, movements of factors between countries that tend to equalize resources reduce incentives to trade. On the other hand, as already mentioned, exports of goods can be viewed as indirect exports of factor services. Trade in goods tends to equalize factor prices and thus to reduce incentives for factors to move.

Mundell (1957) laid out the argument that trade and factor movement can substitute for each other in a model where both trading countries share the same technology. When factor-price equalization holds, free trade implies commodity price equalization and a tendency towards factor price equalization even when factors are immobile while perfect factor mobility implies factor price equalization and a tendency towards commodity price equalization even when trade in goods is not allowed. When factor prices are not equalized, goods trade and factor movement are nevertheless substitutes but in a weaker sense (Wong, 1995).

Wong (1995) shows how the law of comparative advantage can be generalized to cover the movements of goods and capital. The general law of comparative advantage, however, is so general that it cannot be used to predict the direction of movement of a particular good or capital even if all the autarkic prices are known. Wong thus discusses the conditions under which patterns of trade and direction of international capital movements are

predictable. He shows that perfect capital mobility between countries preserves most of the core trade theorems in a Heckscher-Ohlin setting with two goods, two immobile factors and internationally mobile capital. He also shows that without the assumption of identical technologies, the analysis can become quite complicated. Comparative advantage and absolute advantage, defined in terms of price ratios in the countries, are no longer a fixed concept. In the presence of capital movement, they depend on the direction and level of capital movement. Reversal of comparative advantage and the transformation of absolute into comparative advantages are possible.

Norman and Venables (1995) investigate both the direction of trade and the question of which goods or factors are traded. They let goods be tradeable and factors of production be internationally mobile. Since goods trade alone does not equalize factor prices, there is an incentive for international factor mobility. From this general model, they are able to derive conditions on factor endowments and trade costs with the result that the equilibrium has no trade; has trade in goods only; has factor movements only; or has both trade in goods and factor movements.

The substitutability relationship between trade and factor movements is closely associated with the Heckscher-Ohlin endowments driven trade theory. Markusen (1983) demonstrates that factor movements and trade in goods can be complements in models where trade is driven by differences in technologies or by other factors.³⁴ To do this, he uses a simple model with two goods and two factors and assumes that both countries have the same factor endowments but that one of the countries is more efficient in the production of one of the goods. In this setting, the more efficient country exports the good that he produces more efficiently. In the initial trading equilibrium, factor prices are not equalized and if factors are allowed to move, there will be an inflow of the factor used intensively in the production of the export good. This will add a factor proportions basis for trade that will complement the differences in technology basis. Factor mobility will thus lead to an increase in the volume of trade.

Another interesting effect of international factor mobility is that it makes it important to distinguish between domestic and national welfare. Bhagwati and Brecher (1980) shows that in the traditional

Heckscher-Ohlin model of trade theory, a shift from autarky to free trade may reduce national welfare while it increases domestic welfare. Assume for instance that the importable good is labour intensive, labour is wholly national but capital is all foreign. A change from autarky to free trade will lead to exports of the capital-intensive good, which will reduce the real income of labour and increase the real income of capital. Free trade in this case would reduce national welfare.

2. “NEW” TRADE THEORY: GAINS FROM ECONOMIES OF SCALE, PRODUCT VARIETY AND INCREASED COMPETITION

This sub-section discusses the “new” trade theory, motivated to a large extent by the observed importance of intra-industry trade and of trade between similar countries (in terms of technology or resources) that traditional models had difficulties in explaining. Even in the absence of differences, countries gain from trade, since consumers have a wider choice of products at lower prices and firms can exploit economies of scale when having access to a larger market. Of course, the rationalization of production also implies that some firms go out of business. The size and relative importance of these effects have been subject to empirical investigation of pre- and post-liberalization episodes in a range of countries.

(a) Intra-industry trade and the volume of trade between similar countries

Perhaps one of the earliest and best-known studies on the importance of intra-industry trade has been by Balassa (1966) on the formation of the European Economic Community (EEC). He made a number of observations that have triggered the search for an alternative explanation of international trade beyond country differences and comparative advantage. He showed that the trade share of the dominant suppliers in an industry during the implementation of the EEC decreased in practically all industries in the 1958-63 period. This contrasts with the predictions of traditional trade theory, according to which inter-industry specialization in line with comparative advantage would be expected, with the largest supplier within each industry taking the lion's share in the expansion of trade.

Rather than a concentration in traditional export sectors and increasing imports in sectors where countries were at a comparative disadvantage, Balassa observes that EEC countries reduced their reliance on industries in which they had been leading exporters before the establishment of the common market and began to exhibit an increasing uniformity in export patterns. As a consequence, in the absence of declining industries, the need for structural adjustment was limited with little evidence of resulting unemployment and the number of bankruptcies even falling following European economic integration.

Grubel (1967) confirms these results in the case of the EEC, showing that exports and imports within sectors exhibit a tendency towards equalization rather than national specialization. He also notes that the increase in trade between EEC members was mainly due to trade in manufactured goods rather than trade in raw materials. In response to criticisms that the importance of intra-industry trade was a function of the definition of industrial sectors, Grubel and Lloyd (1975) have shown that

significant intra-industry trade is also present at finer levels of statistical aggregation.

These observations are still valid today. For many countries, a large part of international trade takes place within the same sector, even at high levels of statistical disaggregation. Table 3 below shows the Grubel-Lloyd index,³⁵ which is a measure of the importance of intra-industry trade within a given industry, for various German and US sectors. A value of 0.97 for railway/tramway equipment, for example, means that German exports and imports of such products are almost identical. Such a result does not square with traditional trade theory, which predicts that a country is either an exporter or an importer in an industry, not both. If such was the case, the index should be low, zero at the extreme, as in the US footwear industry (0.11), where the United States has substantial imports, but hardly any exports. Looking at the top ten and bottom ten industries for each of these countries, it appears that the former comprise technologically more advanced products, while the latter industries involve comparatively “low-tech” activities.

Table 3
Grubel-Lloyd indices of intra-industry trade, 2006

United States		Germany	
Product (SITC-2)	Grubel Lloyd Index	Product (SITC-2)	Grubel Lloyd Index
Top 10 products			
Metalworking machinery	0.9980	Crude fertilizer/mineral	0.985
Dairy products & eggs	0.9941	Leather manufactures	0.975
Leather manufactures	0.9915	Railway/tramway equipment	0.970
Power generating equipment	0.9876	Sugar/sugar prep/honey	0.966
Electrical equipment	0.9740	Non-ferrous metals	0.953
Perfume/cosmetic/...	0.9479	Meat & preparations	0.947
Crude fertilizer/mineral	0.9405	Furniture/furnishings	0.946
Animal/veg oils processed	0.9393	Coffee/tea/cocoa/spices	0.946
Industry special machine	0.9186	Animal feed	0.937
Plastics non-primary form	0.9009	Organic chemicals	0.935
Bottom 10 products			
Cork/wood manufactures	0.2876	Dyeing/tanning/...	0.55
Furniture/furnishings	0.2830	Metalworking machinery	0.54
Gas natural/manufactured	0.2727	Fixed veg oils/fats	0.47
Petroleum and products	0.1798	Industry special machine	0.45
Travel goods/handbag/etc	0.1612	Vegetables and fruit	0.45
Hide/skin/fur, raw	0.1590	Pulp and waste paper	0.44
Oil seeds/oil fruits	0.1384	Petroleum and products	0.40
Apparel/clothing/access	0.1135	Gas natural/manufactured	0.24
Footwear	0.1110	Oil seeds/oil fruits	0.18
Manufactured fertilizers	0.0789	Coal/coke/briquettes	0.13

Note: Results are similar at the SITC-3 level.

Source: Calculation by authors based on UN Comtrade Database (2007).

Fontagné and Freudenberg (1997) revert to the issue of sectoral aggregation, recalling that the more products are considered as forming part of one industry, the more trade will be of the intra-industry type. In addition, they observe that the Grubel-Lloyd index, even for more disaggregated categories, lumps together trade in intermediate goods (e.g. engines) and final goods (e.g. cars) and would qualify such exchanges as “intra-industry”. The authors, therefore propose to use the most disaggregated trade classification that is available and to distinguish “vertical” two-way trade owing to the international fragmentation of the production chain from “horizontal” intra-industry trade. For the latter to actually capture trade in similar products only, they propose that the export and import unit values should differ by less than 15 per cent and that the lower trade flow (e.g. imports) amounts to at least 10 per cent of the higher trade flow (exports).

If these criteria are applied to a country’s bilateral trade relationships, trade is broken down into: (i) two-way trade in similar products (significant overlap and low unit value differences), i.e. “horizontal” intra-industry trade; (ii) two-way trade in vertically differentiated products (significant overlap and high unit value differences), i.e. “vertical” intra-industry trade; and (iii) one-way trade (no or no significant overlap). Using this methodology, it is evident that intra-industry trade remains important, but bilateral intra-industry intensities vary quite substantially in terms of the trading partners concerned.

Table 4 shows that over half of Germany’s trade with a number of European countries is of the “narrowly” defined intra-industry type. With other countries,

such as Malaysia and a number of other emerging economies but also some industrialized countries, a large part of trade is of the vertical kind, whereas one-way trade still dominates trade relations with a range of developing countries. These patterns indicate that countries share more intra-industry trade with each other the more similar they are in terms of economic size. For example, Chart 4 shows for Germany the positive relationship between intra-industry trade (here defined as “overlap” trade, i.e. both horizontal and vertical two-way trade) and a country similarity index developed by Helpman (1987).³⁶

The chart features high shares of intra-industry trade for many other industrialized countries of similar economic size, but also for emerging economies that are rapidly catching up in terms of GDP. While Germany has an almost balanced trade, for instance, in road vehicles with rapidly developing countries, such as Korea, some of the lowest intra-industry indices are found in relation to other developing countries, where Germany acts as either an importer (e.g. of oil from Azerbaijan or of apparel and clothing from Bangladesh) or as an exporter (e.g. of cars to Sudan) (not shown in the chart). These observations appear to suggest that the theories of comparative advantage remain valid for certain sectors and trading partners, where country differences in technology and resources continue to play a role. However, it is astonishing in view of the importance of intra-industry trade in other sectors and countries (as demonstrated for Germany, which features a 52 per cent share of horizontal intra-industry trade with France as one of its most important trading partners) that such large trade flows remained unexplained until the late 1970s. It was only at that point

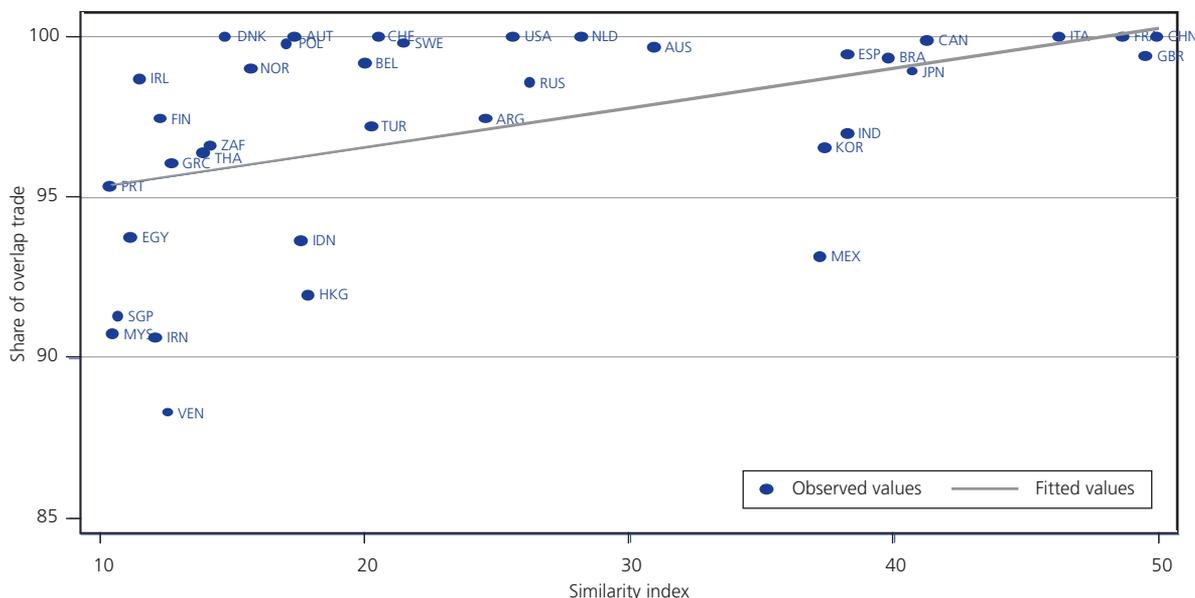
Table 4
Fontagné-Freudenberg indices of intra- and inter-industry trade of Germany, top ten trading partners per type of trade

Partner	Horizontal	Partner	Vertical	Partner	One way
United Kingdom	0.56	Malaysia	0.49	Bangladesh	1.00
Switzerland	0.53	Italy	0.41	Zimbabwe	0.99
France	0.52	Spain	0.39	Madagascar	0.98
Austria	0.51	Belgium	0.38	Algeria	0.98
Netherlands	0.49	Portugal	0.37	Nigeria	0.97
Denmark	0.49	Netherlands	0.37	Macao, China	0.97
Czech Republic	0.47	France	0.36	Panama	0.97
US	0.47	Slovenia	0.35	FYROM	0.97
Belgium	0.45	Sri Lanka	0.34	Iran	0.96
Singapore	0.44	Hong Kong, China	0.34	Ghana	0.96

Notes: Data for Switzerland includes Liechtenstein, Belgium includes Luxembourg. “Horizontal” denotes the share of horizontal two-way trade, “vertical” the share of vertical two-way trade and “one-way” the share of one-way trade.

Source: Calculation by authors based on CEPII BACI database (2007).

Chart 4
Intra-industry trade and similarity in economic size, selected trading partners, Germany, 2004
 (Percent)



Notes: Share of overlap trade is defined as the sum of exports plus imports in sectors characterized by two-way trade divided by the sum of total exports and imports (in per cent). GDP refers to real GDP.
 Source: Calculation by authors based on UN Comtrade Database (2007) and World Bank (2007) World Development Indicators.

that, a complementary theoretical approach was developed that could explain trade in similar goods (e.g. in terms of skill-intensity) between similar countries (e.g. at similar levels of development and technological achievement).

(b) Imperfect competition and trade

This sub-section introduces Krugman’s monopolistic competition model as the best-known way of explaining the gains from intra-industry trade and from trade between similar countries. It also mentions the reciprocal dumping model, which highlights that, under certain conditions, even trade in identical products may be beneficial.

i) Monopolistic competition

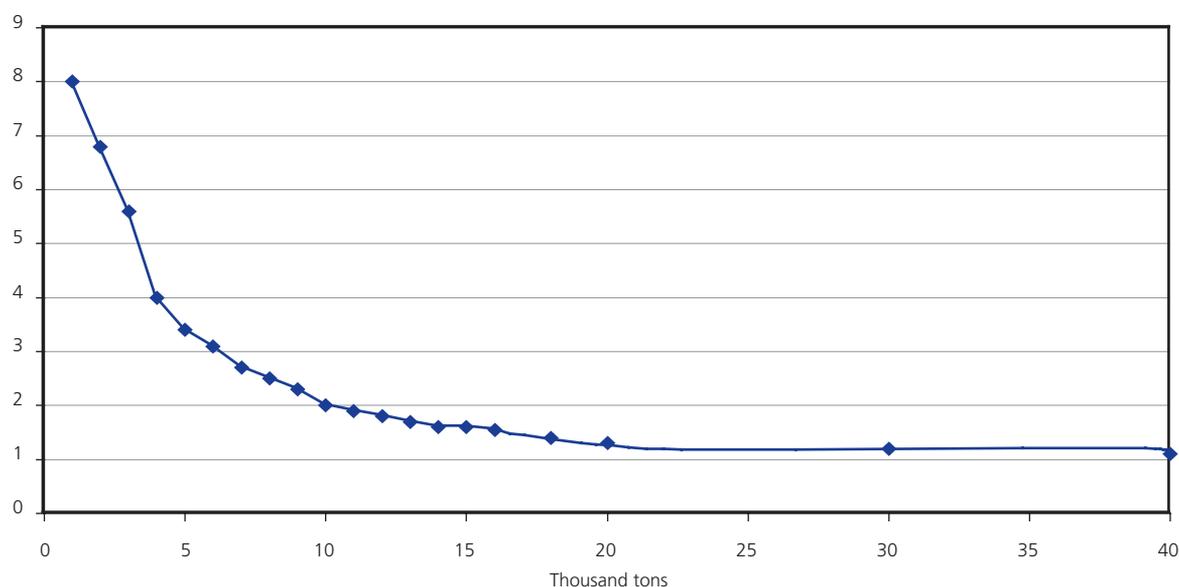
Since traditional trade models seemed unable to explain the above phenomena, a “new” trade theory was needed. Krugman’s monopolistic competition model (Krugman, 1979) is perhaps the best known approach, providing a simple but convincing theory of why similar (in terms of technology, endowments) countries gain from trading with each other and why a significant part of that trade may take place within the same industries. Two basic assumptions, both of which can readily be observed in the real world, are fundamental to Krugman’s model: “increasing returns to scale” and “consumers’ love of variety”.

In the presence of increasing returns to scale (also called “economies of scale”), firms that double their inputs more than double their output.³⁷ Such situations are quite common. In order to start a business (or maintain operations), firms typically face so-called “fixed” costs, i.e. they have to pay for certain goods or services independently of how much they ultimately produce. Such costs may relate to the time of employees spent on administrative issues or to investment in machinery and equipment. In addition, a firm incurs variable costs that increase proportionally to the level of output – for instance, a worker can only produce a given number of units per hour and any increase in production requires the hiring of additional workers at the going wage rate. Marginal costs, i.e. the costs of producing an additional unit of output, are therefore constant, but when the overall level of output rises, the fixed costs get distributed over a larger number of units, and, hence, the firm’s average costs of production decline.

Chart 5 is based on data from a study on slaughterhouses in Norway (van den Broek et al., 2006). It shows that such facilities benefit from economies of scale owing to the presence of significant fixed costs, notably from investments in infrastructure, insurance and personnel to oversee hygiene standards. The larger a facility, the lower its average costs.³⁸ Particularly capital-

Chart 5
Increasing returns to scale and average costs of production

(Norwegian kroner per kilo)



Note: The chart shows the unit cost curve for slaughterhouses in Norway as a function of volume.

Source: Van den Broek et al. (2006).

Box 4 **Economies of scale at Wal-Mart**

The retail company Wal-Mart exploits economies of scale in many ways. Wal-Mart has invested continuously in technology to match its stock with customer demand. For example, it introduced a bar code system early on and a software that connects its stores, distribution centres and suppliers, providing detailed data on the availability of stock. Such fixed investment in stock management is unlikely to pay off for smaller-sized operations. According to Basker (2007), the use of that technology has allowed Wal-Mart to grow and lower its operating costs further through the exploitation of increasing returns to scale.

Basker and Van (2007) estimate that a 10 per cent increase in total sales volume has decreased Wal-Mart's average cost by 2 per cent. The authors also point out that the availability of these technologies, by reducing the costs of tracking stock, has increased Wal-Mart's incentives to add new product lines to its stores, such as pharmacies and auto services. They note that the fact that Wal-Mart is a big retailer gives it a competitive edge in any activity involving a fixed cost, such as contracting with foreign suppliers, which allows Wal-Mart to import at a lower average cost than other retailers.

intensive industries, such as aircraft manufacturing or electronics production, tend to have large fixed costs and economies of scale that often lead to only a few producers worldwide. Increasing returns to scale can also be reaped in the area of services, especially when digital transmission allows for a centralization of certain activities. Box 4 provides a case study of the well-known retailer Wal-Mart.

Since goods can be produced more and more cheaply (i.e. for the same costs, more and more output can be produced), it is certainly economically efficient to produce at a larger scale. The reason why, at

the extreme, there is not only one firm producing a single type of product is that consumers prefer to choose from different varieties for each product they buy rather than buy the same one each time, i.e. they have a "love of variety". Taking the example of food, this means that consumers prefer a selection of different restaurants over one pizza restaurant. Consumers' love of variety favours the existence of many small firms, each producing a somewhat differentiated product, while the exploitation of economies of scale makes it worthwhile to organize production in larger firms.³⁹

Krugman has built these two opposite tendencies into a simple framework of “monopolistic competition”. With larger firms having a cost advantage over smaller ones, the market may cease to be perfectly competitive.⁴⁰ In order to abstract away from the complex issue of firm interaction in such a setting, the “monopolistic competition” market structure assumes that each firm produces a product “variety” that is “differentiated” from the varieties produced by other firms. Therefore, each firm has some leeway to set prices without having to fear that consumers immediately switch to a competing supplier for small differences in price. Since a firm has a “monopoly” in its particular variety within the industry, it can set its own price, and since each firm is small compared with the entire market, it does not take into account the impact of its own price on the prices of other firms. At the same time, while these varieties are not exactly the same, they are substitutes for one another, and each firm continues to face competition from other producers in the industry. In fact, the more varieties that exist (i.e. the lower each firm’s market share), the lower the price that a firm can charge. By the same token, the more firms there are, the less each firm sells (for a given size of the market) and the higher a firm’s

average costs. In market equilibrium, price must equal average costs, which, in turn, determines the total number of firms. If price exceeds average costs, new firms would enter the industry as long as profits can be made; conversely, if price is less than average costs, some firms would exit the market.⁴¹

What happens if two (identical) countries, each with a monopolistically competitive industry, open up to trade?⁴² According to traditional models on country differences (see Section C.1), there would not be any trade. By contrast, with differentiated goods and increasing returns to scale, trade opening enables firms to serve a larger market (and reduce their average costs) and gives consumers access to an increased range of product varieties. However, as consumers can choose among more varieties, they also become more price-sensitive. Hence, while each firm can produce a larger quantity than before (selling to both the domestic and the foreign market), they can do so only at a lower price. As total sales in the integrated market stay the same, and any individual firm is larger, some firms will go out of business. These effects are best illustrated with a hypothetical example (see Box 5).

Box 5 Gains from market integration

This hypothetical example is taken from Krugman and Obstfeld (2006). Assume that two countries, Home and Foreign, each have a monopolistically competitive car sector. Before trade opening, 900,000 cars per year are sold in Home and 1.6 million cars in Foreign. Apart from their different market sizes, the two countries are identical in terms of technology, resources and consumer preferences. Assuming certain fixed and variable costs of production as well as a given elasticity of substitution between varieties, Krugman and Obstfeld determine that there are six firms in Home and eight in Foreign. Given the respective market sizes of the two countries, each firm in Home sells 150,000 cars, while sales per firm in Foreign are 200,000.

Exploiting increasing returns to scale, Foreign firms can produce at a lower average cost and charge a lower price, which the authors determine to be US\$ 8,750 as opposed to US\$ 10,000 in Home. When Home and Foreign open up to trade, the size of the integrated market is 2.5 million cars (the sum of the two national markets in autarky). Each firm serving this larger market now produces more units (250,000 per firm) and the market price for a car has come down to US\$ 8,000. However, this also means that the integrated market can only support ten firms in total. In other words, while consumers have a wider range of choice (ten different car varieties instead of six in Home and eight in Foreign before trade opening), the total number of firms (ten) after market integration is less than the sum of firms in autarky (14).

Example	Home market, before trade	Foreign market, before trade	Integrated market, after trade
Total sales of cars	900,000	1,600,000	2,500,000
Number of firms	6	8	10
Sales per firm	150,000	200,000	250,000
Average cost	10,000	8,750	8,000
Price	10,000	8,750	8,000

In sum, the gains from trade in such a scenario are threefold. Firms produce larger quantities and better exploit their economies of scale (“scale effect”). Consumers in both countries can choose from a wider variety of products in a given industry (“love-of-variety” effect). At the same time, in an integrated market, they pay a lower price (“pro-competitive effect”). Because of these gains, it makes sense that similar countries trade with each other and export/import different varieties of the same good.

While consumers and producers win, those producers that go out of business “lose”. It is impossible to know from the above framework who these producers are and in which country the surviving firms will be located. It may be that each country specializes in producing a narrower range of product varieties under free trade than before (while, of course, all varieties are traded and available for consumption in both countries). Yet, firms may also decide to locate predominantly in one market. For instance, if trade is costly, production may concentrate in the larger domestic market (Krugman, 1980), even if there is some demand abroad. By producing near its largest markets, firms can realize economies of scale, while minimizing transport and other trade costs. Thus, the larger country will produce more varieties and be a net exporter in that industry (the so-called “home market effect”).⁴³

More about the expected trade patterns following liberalization can be gleaned if the basic Krugman model is combined with the traditional approaches concerning country differences discussed in Section C.1 above (Helpman and Krugman, 1985). As in the Heckscher-Ohlin model, one country may be relatively abundant in labour and the other country may be capital-abundant, and one of the two goods may be labour-intensive (e.g. food) and the other capital-intensive (e.g. manufacturing). However, unlike in the Heckscher-Ohlin model, one of the industries, manufacturing, has economies of scale,

with firms producing differentiated varieties in a monopolistically competitive market.

As stated earlier, in the absence of increasing returns to scale, the capital-rich country would export manufactured goods and import food, and vice-versa for the labour-rich country. If manufacturing is a monopolistically competitive sector, the capital-abundant country will still be an importer of food and a net exporter of manufactured goods. The other country, with a comparative advantage in food production, will export both food and manufactured goods, since it produces different varieties of manufactured goods, which some consumers in the capital-abundant country will appreciate. The trade patterns are depicted in Chart 6, with both countries exporting and importing manufactured goods, but with the capital-rich country running a trade surplus in manufacturing.

Ethier (1982) provides another approach to explain trade patterns on the basis of Krugman’s framework. His variant of the model focuses on trade in intermediate inputs, the production of which is subject to economies of scale that are internal to each firm. The cost of producing the final product is lower the larger the bundle of intermediate varieties used. In turn, the larger the production of the final manufactured good, the larger the number and scale of production of the intermediate. If trade in intermediate inputs is free, it does not matter where in the world the production of manufactured goods is located in order to realize these economies of scale. However, if trade in intermediate inputs is restricted, producers of final goods would still need to use all the intermediate varieties available in order to manufacture their products at the same cost. In the presence of trade costs, this will only be possible if all the intermediate and final good production is concentrated in the same country. Hence, this variant of the model can explain the existence of an industrial complex in certain countries. It implies that a reduction of trade costs reduces the need for a concentration of production in any one country.

Chart 6
Trade patterns with country differences in resources and a monopolistically competitive sector



Source: Krugman and Obsfeld (2006: 127).

Before looking at the empirical evidence on the predicted gains from liberalization and related trade patterns, when economies of scale play a role, another model of imperfect competition will briefly be reviewed. This model shows that, in view of certain market imperfections, trade may even be beneficial when countries exchange absolutely identical products.

ii) Reciprocal dumping

As described above, the monopolistic competition model highlights economies of scale as a rationale for trade in similar products and between similar countries. It recognizes that imperfect competition is a necessary consequence of increasing returns to scale at the level of the firm, but disregards most of its consequences. However, imperfect competition, notably the power of firms to price-discriminate between exported and domestically sold products, can itself give rise to international trade between similar countries.

Brander (1986) and Brander and Krugman (1983) describe a situation in which the same good is produced by a monopolist in each of two identical countries. In order to maximize profits, monopolists artificially restrict supply and set prices that are higher than under competitive conditions. While a monopolist could expand sales by reducing its price, it would receive a lower mark-up on all products sold and, therefore, make less profit than at the profit-maximizing price.⁴⁴ If the monopolist firm in each country charges the same price, no international trade will take place. However, if the foreign and domestic market can be segmented effectively, i.e. if a firm can charge a different price on the export than domestic market and domestic residents cannot easily buy goods designated for export, each

monopolist may decide to price-discriminate and enter the foreign market. This decision depends on whether the firm perceives its sales in the foreign market to be more responsive to price reductions than in the domestic market. In the presence of trade costs, this is likely to be the case, as each firm is a lower cost producer at home (where it does not incur the transport costs to ship the good abroad, for example) and will have a lower market share abroad than in the domestic market.

With lower market shares, a firm may double its sales for a given price reduction, but it would need to cut its price much further to double its sales when it has a high market share; hence, a firm is likely to see itself as having less monopoly power abroad and has a higher incentive to keep prices low for exports.⁴⁵ If trade costs are not prohibitive, it makes sense for both firms to “dump”, i.e. charge a lower price for exports than it charges domestically. By selling in the foreign market, each firm makes additional sales and, hence, profits, even if the price is lower than domestically, while the negative effect on the price of existing sales are imparted on the other firm, not on itself.⁴⁶

In this model, reciprocal dumping leads to two-way trade in the same product, even though trade is costly and, initially, prices have been equal.⁴⁷ With the monopoly being replaced by a duopoly situation,⁴⁸ consumers in each country benefit from a larger amount of the product in question at a lower average price. While the increased competition represents a benefit, it is, of course, wasteful to spend resources on the shipping of identical products (or close substitutes), and, depending on transport costs, the overall welfare effect may well be negative.

(c) Empirical evidence

While the “new” trade theory provides a persuasive account of why similar countries may find it beneficial to trade with each other, its usefulness ultimately depends on the actual evidence of the predicted gains from liberalization and its performance relative to competing explanations of trade flows. As far as the gains from intra-industry trade are concerned, most studies have focused on either one of the variety, scale or pro-competitive (price) effects of trade opening. Each effect will be discussed in turn before presenting some evidence about the explanatory power of the models presented in this sub-section compared with other approaches in regard to the observed patterns in international trade.

i) Gains from increased variety

Attempts to measure consumer gains from increased variety are quite recent (due to the detailed data and large computing power needed) and the few studies that now exist have found these gains to be substantial. Broda and Weinstein (2004) compute the welfare gains to consumers as a reduction in the overall price index due to the availability of new varieties, a method developed by Feenstra (1994).⁴⁹ The higher the share of total spending on a new variety, when it appears on the market, and the higher its degree of differentiation compared with existing varieties, the higher the reduction of the overall price index, i.e. the greater the gains to consumers.⁵⁰

Looking at highly detailed import data for the United States, Broda and Weinstein (2004) note a dramatic increase in imported varieties, from about 75,000 varieties in 1972 (or an average of 7,731 varieties from an average of 9.7 countries) to almost 260,000 varieties in 2001 (or about 16,400 varieties from an average of 15.8 countries). The authors divide the sample in two time periods and find that the variety-adjusted unit price for imports fell by 22.5 per cent compared with the unadjusted price over the 1972 to 1988 period (or about 1.6 per cent per year). For the 1990 to 2001 time period they calculate a variety-related price reduction of 5 per cent or about 0.5 per cent annually. Converting these price changes to real income changes, the authors find that welfare has increased by almost 3 per cent solely as a result of the increase in available product varieties.

In ranking US trading partners over time according to the number of exported products, Broda and Weinstein (2004) find evidence that countries do not simply export more of existing products but also supply a greater range of differentiated products as they develop and liberalize. In particular, during the time before 1990, the United States realized important gains from increased variety in imported goods from East Asia, notably the Republic of Korea. More recently, following the North American Free Trade Agreement (NAFTA), the number of varieties imported from Canada and Mexico have risen sharply, and China has continued to play a more and more important role as a supplier of differentiated products.

Feenstra and Kee (2007) examine the effects of trade liberalization on export variety more thoroughly for Mexican and Chinese exports to the United States. Constructing sectoral export variety indices, they find some indication that export variety increased more in sectors where trade liberalization was more pronounced. For example, large tariff reductions by the United States *vis-à-vis* Mexico in the NAFTA context took place in the electronics sector, whereas reductions in agriculture were much smaller. Accordingly, the variety of Mexican exports increased most in electronics and least in agriculture. However, the authors also show that by 2001, China’s export variety exceeded Mexico’s in sectors such as electronics, where Mexico had an initial market access advantage. Estimating that every 1 per cent increase in the export variety of China reduces export variety of Mexico by 0.5 per cent, the authors find evidence that the expected gains from trade liberalization in terms of increased variety must take into account simultaneous liberalization with other trading partners.

ii) Gains from increased competition

A number of empirical studies (examining liberalization in goods and, to a lesser extent, services) have focused on the effect of foreign competition on firms’ pricing decisions. Overall, it appears that trade liberalization has indeed reduced mark-ups of price over costs, although it has proven difficult to disentangle the effects of other relevant factors. Harald (2007) examines the effect of the creation of the European Union (EU) single market (announced in 1985 and implemented in 1993) on price over cost mark-ups using data on 10 EU member states and 18 sectors from 1981 to 1999. Taking cyclical

and technological factors into account, he finds that mark-ups went down in manufacturing by 31 per cent following integration, in particular in the chemicals, rubber and plastic products, metals and metal product sectors as well as parts of the machinery and equipment sector, such as electronic and optical equipment. Conversely, for services mark-ups have risen again slightly since the early 1990s despite the regime shift, which the author attributes to the comparatively weak state of the single market for services and the persistence of anti-competitive strategies in certain services sectors.

Evidence on the significant pro-competitive impact of trade liberalization is also available from developing country case studies. Krishna and Mitra (1998) find important decreases in price-cost margins for most industries in response to a range of liberalization measures undertaken by India in 1991. Harrison (1990) obtains similar results for Côte d'Ivoire following the implementation of a comprehensive trade reform in 1985. Both studies take other factors into account, such as the influence of technological progress and business cycles. Using data on almost 300 firms, Harrison even accounts for the possibility of variations in mark-ups not only across sectors but also across firms. Roberts and Tybout (1991) have put together a collection of developing country case studies (Chile, Colombia, Mexico, Morocco and Turkey), which examine the relationship between the exposure to trade and price-cost margins at both the industry and plant levels, taking the usual factors into account plus a measure of existing domestic competition. Owing to the latter, it becomes apparent that the pro-competitive effects of increased import penetration are particularly strong in highly concentrated industries, i.e. that the impact of trade liberalization is strongest where firms have a degree of market power prior to trade opening.

Finally, Hoekman et al. (2004a) undertake a cross-country analysis of 42 developed and developing countries in order to examine to what extent country differences may explain why trade opening has a more pronounced effect on mark-ups in some countries, taking other differences into account, such as a country's level of economic development or institutional environment.⁵¹ The authors find that both tariff cuts and reductions in other market entry barriers (proxied by the number of administrative procedures required to establish a new, domestic or foreign firm) have a negative effect on mark-ups, but that the effect of trade liberalization is less strong when administrative

barriers are more significant, since these may act as a substitute for lower tariffs. As an example, the authors estimate that Colombia could more than halve its average industry mark-up if it reformed its restrictive market entry regulations to the level found in Canada (least restrictive in the sample) and brought down its manufacturing tariff from the current 11 per cent to zero (like Hong Kong, China).

The study also highlights that the impact of tariffs on mark-ups decreases with country size, whereas the impact of entry regulations increases. In other words, smaller countries (that are naturally more open) will see a relatively larger reduction of industry mark-ups when they liberalize their tariff regime while larger countries obtain comparatively better results from reforming their domestic market entry procedures. For instance, Uruguay and Malaysia have the same average tariff level (around 12 per cent), but Malaysia is twice as large as Uruguay in terms of GDP per capita and a marginal change in tariffs in Uruguay has a 14 per cent larger effect on mark-ups than in Malaysia.

iii) Gains from increased economies of scale

While the importance of variety and pro-competitive gains from trade have been established empirically, there is mixed evidence at best of net increases in scale following trade liberalization. Head and Ries (1999) analyze the impact of the Canada Free Trade Agreement (FTA) with the United States for 230 Canadian industries (at the 4-digit SITC level). Following the conclusion of the FTA, almost all Canadian manufacturing industries exhibited substantial rationalization between 1988 and 1994, i.e. a decline in the number of plants accompanied by increases in output per plant.

The authors find that the scale increases experienced by the average industry over that time period cannot be explained by trade liberalization. Their analysis shows that the average US tariff reductions of 2.8 per cent caused a 4.6 per cent scale increase, which was more than offset by the scale decline of 6.1 per cent owing to Canada's own tariff reductions of 5.4 per cent. These effects are similar but larger in imperfectly competitive industries and smaller in high turnover industries, where free market entry and exit of plants appear to dampen scale adjustments.⁵² Roberts and Tybout (1991) obtain similar results looking at a panel of Chilean and Colombian firms over the mid-1970s to mid 1980s.

They examine to what extent changes in plant size can be explained by increased trade exposure, as measured by higher export and import shares or, alternatively, reductions in effective protection. Exposure to foreign competition in the domestic market reduces average plant sizes, while increasing export shares, at least in the short term, have the opposite effect. Again, size adjustment occurs more in industries with low turnover of firms, i.e. where market entry/exit is more difficult.

From these studies, it becomes evident that factors other than scale appear to explain the overall efficiency gains at the sectoral level following trade opening, notably the observed reallocation of market shares towards more productive firms. Such differences between firms have not been modelled in the theoretical approaches presented above, and empirical results of that nature have certainly given a boost to the development of the “new-new” trade theory (which explicitly takes account of firm “heterogeneity”) discussed in Section B.3.

One study that has opposed the effects of liberalization on scale versus selection of firms and shifts in market share is the one by Tybout and Westbrook (1995) on Mexican manufacturing plants covering the 1984 to 1990 period and the liberalization undertaken in 1985. The authors note significant improvements in productivity and average costs during this period. Improvements were largest in the more open sectors, measuring either import or export rates. A number of manufacturing sectors show modest increases in internal returns to scale, but these are only significant for the smallest plants, while the largest plants appear to have reached a minimum efficient scale. Thus, with large plants carrying more weight in sectoral aggregations, increases in openness are associated with relatively small-scale efficiency gains overall. More importantly, open sectors are characterized by some degree of market share shifting towards the more productive plants. However, for the most part, cost reductions and productivity gains are explained by a “residual” factor, which captures the effects from technological innovation, learning-by-doing and other phenomena that are difficult to quantify.

iv) Observed trade patterns and competing theoretical approaches

In order to compare the “new” trade theory with established approaches, a number of studies have further developed the new models, notably the

monopolistic competition model, to yield some empirically testable hypotheses. The question is whether the predictions by the model are consistent with the trade data, notably the results obtained from the gravity approach (see Box 6) and the Grubel-Lloyd measures of intra-industry trade that traditional theories had difficulties to explain, or whether other approaches, both new and old, can better explain the observed relationships.

Hummels and Levinsohn (1995) test the positive association between trade volumes and similarity in size (if countries also have identical preferences), as hypothesized by Helpman (1987) on the basis of a model of monopolistic competition and confirmed empirically by him for a group of OECD countries using a gravity set-up. Hummels and Levinsohn (1995) use instead a diverse sample of developing economies and find that the relationship between size dispersion and the variation in trade volumes, as predicted by the monopolistic competition model, still holds. Since these countries cannot be described as having identical demand structures and as trading predominantly in differentiated products, i.e. as fulfilling the assumptions highlighted by Helpman (1987), it is not clear that the monopolistic competition model necessarily provides the best rationale for such trade flows.

As an alternative test, the authors regress the Grubel-Lloyd indices on a range of measures of factor endowments in each country, such as income per worker or land-labour ratios. In so doing, they are able to confirm that the bilateral share of intra-industry trade is higher for countries that are more similar in terms of factor composition, as in Helpman (1987) and predicted by the monopolistic competition model. However, when more sophisticated econometric methods are employed, the empirical support for the theory becomes mixed. Rather than being explained by factor similarities, much of intra-industry trade appears to be specific to country-pairs and not explained by a common factor.

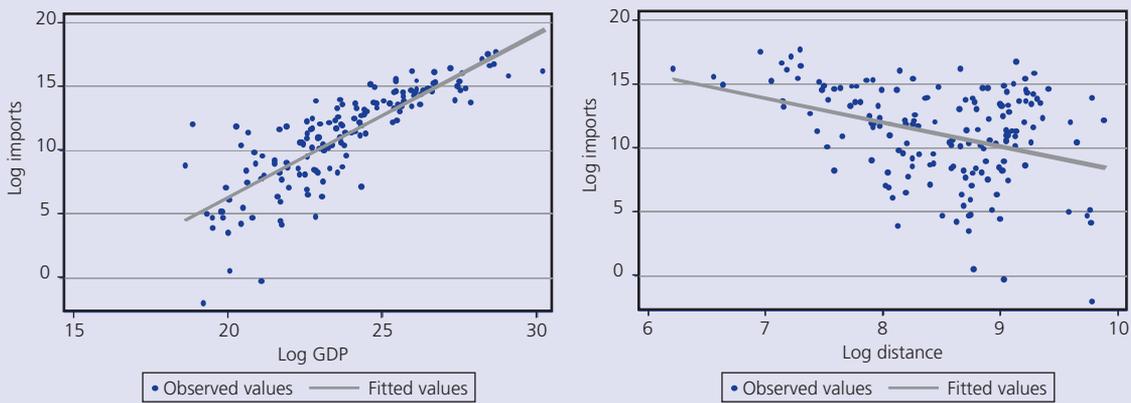
A number of authors have made the attempt to differentiate explicitly between competing models by deriving mutually exclusive, empirically verifiable predictions from each model. Feenstra et al. (2001) hold that the gravity equation is consistent with several theoretical models of trade that, nevertheless, predict certain differences in key parameter values. The authors confirm the predictions of the monopolistic competition model

Box 6
The gravity equation

The gravity equation was developed by Tinbergen (1962) in an attempt to predict the pattern of international trade that would prevail in the absence of distortions. He postulated that the value of bilateral trade between two countries was an increasing function of the gross national product (GNP) of both the exporting country (reflecting the assumption that export supply capacity depended on a country's economic

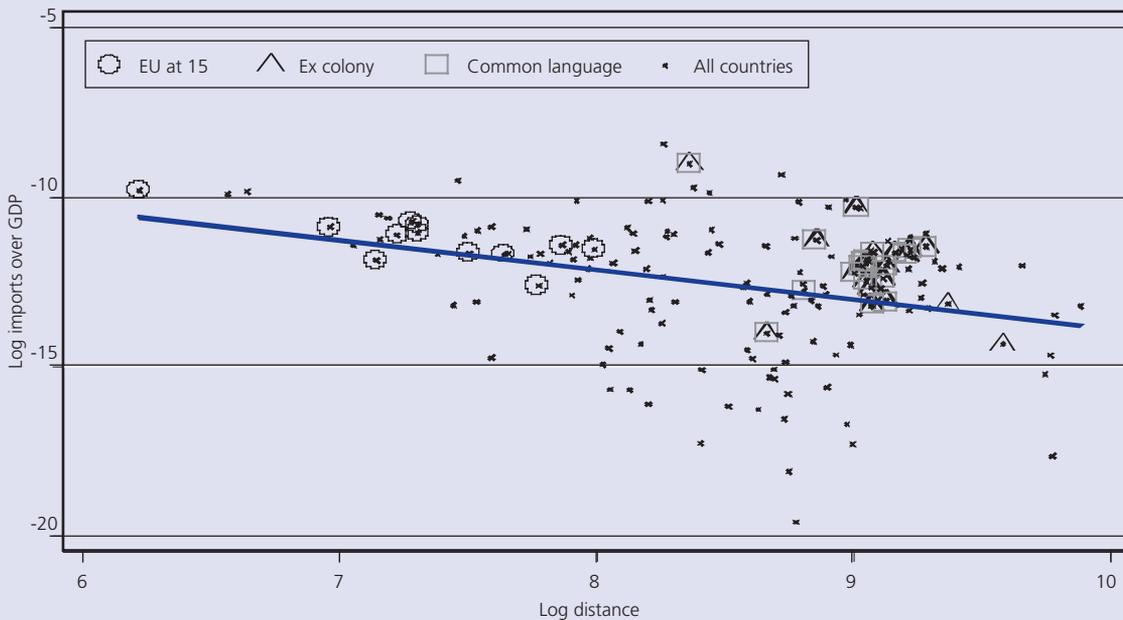
size) and the importing country (assuming that import demand also increased with a country's market size). At the same time, he observed that trade flows were influenced negatively by the "distance" between two countries, as a measure of transportation costs or other obstacles, such as the cost of information on the export market. These relationships are portrayed in Chart A for Spain.

Chart A
Bilateral trade of Spain as a function of GDP of both trading partners and as a function of geographical distance, 2006



Source: Calculation by authors based on UN Comtrade Database (2007).

Chart B
Bilateral trade of Spain as a function of geographical distance as well as other trade barriers, 2006



Source: Calculation by authors based on UN Comtrade Database (2007).

Trade with a range of trading partners increases with both countries' GDP and decreases with the geographical distance. The relative "distance" between trading partners is not confined to geography, but includes other "barriers" that increase trading costs, such as language differences, historical/cultural factors and, not least, trade barriers. In Chart B, the round dots mark Spain's trade with other EU members, the squares denote trade with other Spanish-speaking countries and the triangles refer to trade with former colonies. It can be seen that Spain trades relatively more with countries with which trade "barriers" are lower in one respect or another than with other countries at a similar distance.

This so-called "gravity equation", in reference to Newton's law describing the force of gravity as a function of the product of the masses of two objects and the distance between them,

has been extremely successful in explaining the determinants of bilateral trade flows and the impact of trade policies, such as the creation of free trade areas. Yet, it did not appear to offer any role to comparative advantage. The monopolistic competition model discussed below was the first model that provided a complete theoretical basis for the gravity equation. Previously, Anderson's (1979) Armington model had provided a first, albeit incomplete, theoretical foundation based on differentiation of goods by country of origin. Later, others, e.g. Eaton and Kortum (2002), have been able to derive the gravity equation from the Heckscher-Ohlin framework and Ricardian model respectively. The former is characterized by complete or at least a certain degree of specialization of countries in certain goods, while in the latter modelling approach, countries are not specialized, but owing to transport costs, any particular good is only imported from the cheapest producer.

and reciprocal dumping model with free market entry for differentiated products, while trade in homogeneous goods (i.e. bulk commodities and the like) appears to be better described by alternative approaches.⁵³

Similarly, Evenett and Keller (2002) estimate a gravity equation to test the predictions of the monopolistic competition and Heckscher-Ohlin models.⁵⁴ They split their sample of bilateral import data into two subsets with high and low degrees of intra-industry trade respectively. For the former, they expect trade to be based predominantly on product differentiation and increasing returns to scale and further subdivide the sample according to the level of intra-industry trade. The other subset is sorted according to each observation's differences in factor proportions.

For the first sample, the authors find that a higher share of differentiated goods in GDP is indeed associated with a higher share of intra-industry trade in total bilateral trade. Likewise, when

there is little intra-industry trade (second sample), trade rises with increasing bilateral differences in factor proportions. From these results, it may be concluded that a monopolistic competition framework emphasizing economies of scale and product differentiation is well-suited to explaining trade among industrialized nations ("North-North" trade). By contrast, factor differences appear to play an important role in the trade between developed and developing countries ("North-South" trade), which tends to focus more on the exchange of homogeneous goods.⁵⁵

Despite its obvious empirical relevance, the new trade theory must be seen as a complement to rather than substitute for traditional approaches which continue to play a role in the explanation of trade flows. At the same time, it has triggered further advances in trade theory addressing some unanswered questions, such as which firms will prosper and which ones decline under free trade and where production will take place. These are further discussed in Sections C.3 and D.

3. RECENT DEVELOPMENTS: PRODUCTIVITY GAINS

Until recently, trade literature has not focused much attention on the role of firms in international trade. Mainly for simplification purposes, trade theorists typically used the concept of a representative firm, assuming that all firms in a given industry are identical. In the 1980s, however, firm-level data sets with detailed information on production and trade at the firm level became available.⁵⁶ This information showed considerable differences (“heterogeneity”) between firms and suggested that these differences affected overall outcomes. Trade economists consequently developed a series of new trade models that focus on the role of firms and that explain the empirical findings. These models have identified new sources of gains from trade and new ways in which international trade may lead to resource reallocation (Bernard et al., 2007a).

(a) Differences among firms matter

This sub-section reviews recent firm-level empirical evidence which shows that: (a) most firms, even in traded-goods sectors, do not export at all; (b) of those firms that export, only a few export a large fraction of their production; (c) at the same time, at least some firms export in every industry, with the share of exporting firms being a function of the industry’s comparative advantage; (d) firms that export are different from non-exporters in a number of ways (they are bigger, more productive, pay higher wages and are more capital and skilled

labour intensive than non-exporters); (e) trade liberalization raises industry productivity.

Two important points about the data are worth mentioning. First, the focus is on evidence regarding exporting. This is because until recently, most of the firm-level evidence has been concerned with exporting and the new theories have been developed to account for export-related evidence. Data on US firms’ imports is only briefly reviewed, since this has only recently become available as part of the new transaction-level trade data. This information reveals that data on firms’ imports share many of the features of those on firm exports. Second, while for the time being most of the firm-level evidence is from developed countries, available developing country evidence is also covered in this sub-section. As discussed below, available information suggests that many of the insights drawn from early studies for US exporting firms is confirmed as applicable to other countries when similar firm-level data is made available.

The first point to note is that the share of exporting firms in the total number of firms is relatively small. In 2002, only 20 per cent of all US manufacturing plants and 18 per cent of all US manufacturing firms were exporting (Bernard et al., 2007a; Bernard et al., 2006a). Unfortunately, comparable figures are only available for a small number of other countries. As shown in Table 5, while the fraction for Norway is considerably higher at 40 per cent, figures for France, Japan, Chile and Colombia are in the same 20 per cent range as the US fraction.

Table 5
Share of exporting firms in total number of manufacturing firms

	Year	Share of exporters in total number of manufacturing firms
United States	2002	18
Norway	2003	39.2
France	1986	17.4
Japan	2000	20
Chile	1999	20.9
Colombia	1990	18.2

Note: US: U.S. Census of Manufacturers. Norway: all non-financial joint-stock firms in the manufacturing sector (approximately 90 per cent of the manufacturing industry totals). A firm is an exporter if its exports are over NOK 1000. France: comprehensive data set of French manufacturing firms; 113 countries and 16 industries are included; data fail to account for 20 per cent of total export data. Japan: Survey database of the Ministry of Economy, Trade and Industry which includes all manufacturing and non-manufacturing firms with more than 50 employees and a turnover exceeding 30 million Yen; firms that re-entered and started exporting after 1994 are excluded; unbalanced panel with 22000 observations a year. Chile: Encuesta Nacional de Industria Annual and National Customs Department data; 1991-1999; importers returning goods are recorded as exporters, might lead to overestimation. Colombia: Columbian Manufacturing Census; panel data; plants with 10 or more employees.

Source: United States: (Bernard et al. 2007a); Norway: (Mayer and Ottaviano, 2007); France: (Eaton et al. 2004); Japan: (Kimura and Kiyota, 2006) Chile: (Alvarez, 2004); Colombia: (Brooks, 2006) .

These figures should be interpreted cautiously. The Norwegian sample of firms, for instance, accounts for only 90 per cent of value added. In the Chilean and Colombian cases, only manufacturing firms with more than 10 employees are covered and the Japanese dataset only includes firms with more than 50 employees. If there are proportionally less exporters among the smaller firms than among the larger ones, these figures could be biased upwards.

While exporting is a relatively rare activity, evidence from the United States shows that it occurs in all manufacturing industries (Bernard et al., 2007a). It also shows that exporting is relatively more frequent and export intensity relatively higher in more skill-intensive sectors than in more labour-intensive sectors. Comparing the percentage of firms that export across US manufacturing industries in 2002, they found that 8 per cent of firms were exporting in the apparel sector compared with 38 per cent in the computer and electronic products industry. Similarly, comparing mean exports as a percentage of total shipments across industries, the authors found that the value of exports as a share of total shipments ranged from 7 per cent for firms in the beverages and tobacco products sector to 21 per cent for those in the computer and electronic products sector.

Evidence also shows that exporting is concentrated. A minority of the firms that export make up the bulk of exporting activity measured on a value basis. Based on a ranking of a country's firms in terms of their individual exports, the contributions

to overall exports of the largest exporters can be calculated. The shares of total exports accounted for respectively by the largest 1, 5 and 10 per cent of exporting firms are reported in Table 6 for a number of countries. Figures show that in most of the sample countries, the largest 5 per cent of exporting firms account for more than three-quarters of total exports.

These observations challenge both the old and the new trade theories. The old theory, for instance, can explain why export intensity is higher in skill-intensive sectors but it cannot explain why some firms export but many do not. Similarly, the fact that there are exporters in all industries is consistent with new trade theories but the fact that only some firms export is not.

An examination of firm-level evidence also shows that exporters are different. First, data suggest that US firms that export are more capital-intensive and skill-intensive with respect to their choice of inputs than are firms that do not export. If this suggests that the goods that these exporting firms produce are more capital-intensive and skill-intensive (in line with their input choice), then this would be evidence from the firm level of the United States exporting products in line with its current, underlying comparative advantage.⁵⁷ This result is confirmed by Mayer and Ottaviano (2007), who provide evidence that the export performance of European countries is better in those industries where they have a comparative advantage.

Table 6
Per cent of exports accounted for by largest exporters

	Year	Top 1%	Top 5%	Top 10%
United States	2002	80.9	93	96.3
Belgium	2003	48	73	84
France	2003	44	73	84
Germany	2003	59	81	90
Hungary	2003	77	91	96
Italy	2003	32	59	72
Norway	2003	53	81	91
UK	2003	42	69	80
Chile	1999	49.12	82.25	96.45

Note: Data description (for some countries see Note in Table 3): Belgium: Data taken from Balance Sheet Trade Transactions Dataset; for intra EU trade firms with >250 000 euros trade flows are considered; for extra EU trade firms with >1000 Euro (or one tonne) are considered. Exports reported at the eight-digit level. France: Data are taken from the French Customs. Exports are reported at the eight-digit level. Intra EU trade is reported only for >250 000 Euros; Extra EU trade is reported for >1000 Euros. Germany: Data taken from Federal Statistical Office; Establishment Level Panel Data; manufacturing sector only; covers firms with >20 employees only. Hungary: Data set contains 2043 firms, with exports >100 million HUF; this represents 60-70 percent of total exports. Exports reported at the six-digit level. Italy: Data taken from Capitalia database; survey on Italian manufacturing firms; for firms <500 employees and more than 11 the survey is selective; for firms >500 employees all are included. UK: Data taken from FAME database.

Source: United States: (Bernard et al. 2007a); Belgium, France, Germany, Hungary, Italy, Norway, UK: (Mayer and Ottaviano, 2007); Chile: (Alvarez, 2004).

Second, one of the most robust and most important results found across countries in this literature is that exporting firms are more productive than non-exporting firms. Bernard et al. (2007a) estimate that US exporting plants are more productive than non-exporting plants by 14 per cent for value-added and 3 per cent for total factor productivity. Similarly, Mayer and Ottaviano (2007) estimate that French exporters exhibit a 15 per cent higher total factor productivity than non-exporters and a 31 per cent higher labour productivity.⁵⁸ Given this relationship, a natural question is whether exporting causes firms to be more productive (a “learning by exporting” effect) or whether it is simply the case that more productive firms choose to become exporters, while less productive firms choose not to (a “selection” effect). Here, empirical evidence is mixed. Earlier studies, using data on US firms, followed by a number of studies on exporting firms in countries as diverse as Canada, Colombia, Germany, Mexico, and Morocco, supported the “selection” hypothesis.⁵⁹ Several recent studies, however, find evidence of “learning by exporting”.⁶⁰

Third, another related result is that exporting firms are larger, whether measuring their size by output (domestic shipments) or employment.⁶¹ The fact that these firms are larger also has important implications in relation to the discussion of the relation between trade and productivity. While exporting firms may or may not enjoy any higher productivity growth after they begin exporting (relative to non-exporting firms), evidence suggests that average industry productivity increases following trade liberalization, as a result of the contraction and exit of low-productivity firms and the expansion and entry into export markets of high-productivity firms.⁶²

Finally, there is also evidence that exporting firms pay higher wages than non-exporting firms. Even when examining only intra-industry variation and after taking firm size into account, there is still a wage premium for employees of exporting firms, compared with those of non-exporting firms. Box 7 illustrates how firm-level information can be analyzed using relatively simple techniques.

Box 7
Using firm-level data to analyze export behaviour

Using firm-level information from the Amadeus database, simple techniques are applied to investigate two sets of questions that have been highlighted in trade literature: the exporter premium and the self-selection into export effect. The sample consists of 28,621 medium to large French firms, of which 12,502 (48.9 per cent) recorded non-zero exports in the last available year. The top 5 per cent of exporters account for more than 84 per cent of total exports.⁶³

Exporter productivity premium

The productivity premium of exporters is the average percentage difference in productivity

between exporters and non-exporters, taking firm- and sector-level characteristics into account. It can be assessed using the following simple regression of export status on labour productivity:

$$\ln LP_i = \beta_0 + \beta_1 Export_i + \beta_2 \ln Age + \beta_3 \ln IntAssets + \epsilon_i \quad (1)$$

where *i* indexes firms; *LP* is labour productivity measured as operating revenue per worker; *Export* is an indicator variable for export status; *Age* is the number of years since the firm has been established; *IntAssets* is intangible fixed assets, included as a measure of fixed costs. A full set of sector dummies has been added to equation (1). The regression results are as follows:

OLS regression

Dependent variable: *ln productivity*

dummy for export status	0.095***
<i>ln age</i>	-0.094***
<i>ln intangible assets</i>	-0.023***
R ² 0.4022	Number of observations: 17986

Denotes significance at ***1%
Constant and sector dummies not reported.

The exporter premium, calculated as $100(\exp(\beta_1)-1)$, is equal to 10.1 per cent. That is, after controlling for firm- and sector-level characteristics, exporters are on average 10 per cent more productive than non-exporters.

Self-selection

The hypothesis that high-productivity firms self-select into exporting is also tested. The test uses the estimation results of the following Probit model:

Probit regression

Dependent variable: Probit (export)	
ln productivity	0.057***
ln age	0.005
ln intangible assets	-0.006***
ln number of employees	0.041***
Pseudo R ² 0.2496	Number of observations: 17488

Denotes significance at ***1%
Constant and sector dummies not reported.
Marginal effects are reported.

$$Export_i = \beta_0 + \beta_1 LP_i + \beta_2 \ln Age + \beta_3 \ln Empl + \beta_4 \ln IntAssets + \varepsilon_i \quad (2)$$

where *Empl* is the number of employees, a measure of firm size, and where a full set of sector dummies is added. This produces the following regression results shown below. The estimated elasticity of productivity on the export dummy is positive and significant. Empirical support is found, therefore, for the idea that more productive firms self-select into exporting.

Access to increasingly detailed data that links firms to such characteristics as the number of products they produce and the number of foreign destinations they export to also reveals interesting characteristics of exporting firms (Bernard et al., 2007a). Evidence for the United States suggests that when firms export, they tend to export multiple products. More precisely, while most firms export a relatively small number of products, most exports are done by firms shipping many products. In Europe, data show that top exporters export many products to many locations (Mayer and Ottaviano, 2007). Firms exporting more than ten products to more than ten markets account for more than 75 per cent of total exports. Earlier research also examines the question of “product-switching” and questions related to how firms change their mix of product offerings in response to the pressures of globalization. It also examines how many markets an exporter ships to and to what markets in particular.

This sort of new approach on newly available data suggests many exciting future avenues to improving understanding of how firms that compete in global markets are behaving. Note there is also evidence from US import data on within-product

differentiation (Schott, 2004) that is consistent with a theory of within-product specialization that may reflect differences in product quality, i.e. that rich countries export varieties with high unit values and lower-income countries export varieties with lower unit values.

Finally, using this newly matched data on firm-level transactions, these researchers are able to compare US firms that import with firms that do not import.⁶⁴ The first interesting finding is the high degree of positive correlation between importing firms and exporting firms – firms that import are more likely to export than non-importing firms, and firms that export are more likely to import than non-exporting firms.⁶⁵ Bernard et al. (2007a: Table 8) illustrate a number of other common characteristics between US firms that import and US firms that export – in both cases, the firms are larger, more productive, pay higher wages and are more capital and skill-intensive than firms that do not import.

Missing from this literature, however, is information on firms in industries aside from manufacturing. In particular, because of data limitation, empirical

economic research has had very little to contribute thus far to information on ways in which firms in service industries may be similar or different from those in manufacturing.

Most of the extremely detailed information discussed in this sub-section on what differentiates exporting firms from non-exporting firms or importing firms from non-importing firms relies on detailed data for the United States and European countries. In studies referenced above, a number of these characteristics on productivity and size have been confirmed when examining exporting and non-exporting firms in other countries as well. Nevertheless, there are potential avenues of research, especially in the case of developing countries, that would be worth pursuing. For example, Tybout (2000) describes some extremely interesting characteristics of the size distribution of manufacturing firms in developing countries that may result from an historical environment of protectionism and heavy domestic regulation – oligopolies of extremely large plants with large market shares co-existing with a set of very small plants that are unwilling and/or unable to grow in order to take more advantage of economies of scale.

(b) Models with differences among firms

In the last few years, a new strand of trade models have been developed that incorporate firm-level differences to account for the new firm-level facts discussed above. As summarized by Baldwin (2006b), these so-called “new new” trade models differ from the “new” trade theory models discussed in sub-section 2 by allowing for differences with respect to firms’ marginal costs and fixed market-entry costs that are added to the standard fixed cost of developing a new variety.

This sub-section presents the thinking behind and the basic features of the model introduced by Melitz (2003).⁶⁶ The main reason for focusing on this particular model is that it is proving to be a particularly effective platform for modelling trade with differences among firms (Baldwin, 2006b). In the Melitz model, only a subset of firms exports and there are exporters in most industries. A number of key implications of these features are emphasized, notably the impact of liberalization on average industry productivity through a selection mechanism and its effects on the number of firms

as well as on the number of varieties for consumers. Other models based on differences among firms, embedded in either new trade theory models or in Ricardian models, are subsequently considered and compared with the Melitz model.

The Melitz model is in the tradition of monopolistic competition trade models. More precisely, it introduces differences between firms into the Krugman (1980) model of “new” trade theory. The economy is endowed with a single factor of production: labour (L). There is a single industry that produces horizontally differentiated products. Each firm produces one single variety using a technology with increasing returns to scale. Competition is imperfect but there are many firms. Firms vary in terms of their total factor productivity. Each firm is assumed to draw its productivity level in a “lottery” after paying a one-time fixed sunk cost of entry (or “invention cost” as Baldwin calls it). This can be thought of as a way to model a situation where the firm invests in research and development (R&D) to develop a new variety and there is uncertainty involved in the R&D process regarding the marginal cost of producing the new variety.

In addition to the sunk “invention cost”, if it enters the domestic market, the firm has to pay a fixed entry cost. Similarly, if it wishes to export, it has to pay the fixed cost of entering the export market. Depending on the level of its productivity, the firm will thus either produce or exit, and if it produces, it will either produce only for the domestic market or be an exporter. Only firms with sufficiently low marginal costs will be able to sell enough to cover fixed costs. Firms with the lowest marginal costs will find it profitable to pay the entry cost for both the domestic and the export market, while firms with intermediate productivity levels will find it profitable to pay only the entry cost for the domestic market. In other words only the most productive firms become exporters.

In summary, considering a ranking of all firms according to their productivity level, there are three outcomes for firms and two cut-off conditions – that is, two threshold levels of marginal cost. The least productive firms (i.e. those with a marginal cost above the first threshold level) exit the market, those between the first and the second cut-off points enter but only sell on the domestic market and those with the highest levels of productivity (i.e. with a marginal cost lower than the second threshold level) both export and sell domestically. The threshold

marginal cost for entering the local market depends on the fixed entry cost of entering the domestic market as well as on prices and demand conditions. Similarly, the cut-off marginal cost for entering the export market is a function of the fixed cost of entering the export market, the trade costs, the price and demand conditions.

In this setting, Melitz (2003) shows that increases in the exposure to trade through either a transition from self-sufficiency to trade or a reduction of trade costs will force the least productive firms to exit and reallocate market shares from less productive to more productive firms. He further shows that increased exposure to trade will always deliver welfare gains.

When entry into new export markets is costly, Melitz shows that exposure to trade offers new profit opportunities only to the more productive firms that can afford to cover the entry cost. This also encourages more market entry as prospective firms respond to the higher potential returns associated with good productivity. At the same time, falling trade costs also reduce the minimum level of productivity that firms need to export successfully and thus, the highest-productivity non-exporters enter the export market and existing exporters see their sales grow as they take advantage of new markets. The increased demand for labour by the more productive firms and new market entrants increases overall wages and forces the least productive firms to exit. In other words, the minimum level of productivity needed to survive increases, prompting the lowest-productivity firms to exit and average industry productivity to increase.

Baldwin and Forslid (2004) systematically study the positive and normative aspects of the effects of trade liberalization in the Melitz model. As far as the positive effects of trade liberalization are concerned, they emphasize two main results. First, as mentioned, liberalization has a strong impact on average productivity via a selection effect (the least productive firms drop out of the market) and a reallocation effect (from the less to the more productive firms). Second, trade liberalization leads to an anti-variety effect. Freer trade reduces the number of varieties produced in each country and under reasonable assumptions also reduces the total number of varieties consumed.⁶⁷ Using profits to measure the return to capital, the authors also show that the model displays classic Stolper-Samuelson-like behaviour.

Finally, turning to welfare effects, Baldwin and Forslid break down the total welfare impact into three partial effects: the negative anti-variety effect, the positive productivity effect, and a positive effect related to a substitution and share-shifting effect in favour of imported varieties. They show that considering all the effects together, the positive productivity and substitution/share-shifting effects outweigh the anti-variety effect, so that the overall impact of freer trade is unambiguously positive.

Bernard et al. (2007b) examine how firm, industry and country characteristics interact as trade costs fall in a model that embeds differences in firms into a framework of comparative advantage. The model they use has two countries, two factors and two industries. Each industry is populated by a continuum of firms that each produce a single differentiated variety within their industry. Firms vary in their level of productivity, industries vary in factor intensity and countries differ in terms of factor abundance. Using their model, they demonstrate that increased exposure to trade raises the productivity cut-off necessary for survival, which raises average productivity in both industries. In their model, the strength and importance of firm self-selection varies with the interaction between country and industry characteristics. The rise in productivity is more pronounced in the comparative advantage industry, because firms' export opportunities in this industry are greatest. This outcome magnifies the original differences between countries and thereby boosts the welfare gains from trade. By increasing exporters' profits, falling trade costs also reduce the export productivity cut-off level. Here too, responses vary according to country endowments and industry factor intensity. Another interesting effect is related to aggregate productivity growth. Increases in industry productivity reduce the price of the average variety in each industry and thereby elevate the real income of both factors. This effect may even be strong enough to raise the real wages of both factors. The possibility of such an outcome, which also depends on the model's parameters, contrasts sharply with the predictions of the traditional model.

Yeaple (2005) proposes an alternative explanation of the economic implications of international trade in the presence of differences among firms. In his model, firms are the same when they start out. Differences arise when they choose to employ different technologies and systematically hire different types of workers. In a two-sector economy, firms in one sector

produce a differentiated good. In this sector, firms can choose to employ a medium or high technology. The fixed investment costs for the high technology are higher than for the medium technology.

In the second sector, according to Yeaple, firms produce the same good, employing a standard low technology. Additionally, workers differ in their skills. High-skilled workers have a comparative advantage in using the high technology whereas medium-skilled workers have a comparative advantage using the medium technology. Therefore, in an equilibrium, firms that choose to use the high technology employ highly skilled workers. As the labour market is competitive, high-technology firms pay a higher wage.⁶⁸ Only firms that choose to use the high technology are able to do this and hence attract high-skilled workers. Firms that do not find it profitable to choose the high technology might still find it profitable to use the medium technology and to hire the medium-skilled workers. This might be possible because the medium technology firms pay a lower wage to their workers.

As in Melitz (2003), firms must incur a fixed exporting cost and hence at a sufficiently high level of the latter, only high-productivity firms find it profitable to export. If trading costs are reduced, more firms adopt the better technology, which leads to an improvement in sectoral productivity.⁶⁹ It is interesting to note that a reduction in trade barriers even between identical countries raises the relative demand for skilled workers and the skill premium.⁷⁰

(c) Empirical evidence

The interplay of differences between firms and fixed market entry costs represents the main driver of the Melitz model. As discussed in Section C.3.a, highly detailed data on trading relationships for the United States and France confirm the prevalence of differences at the firm level and also corroborate the existence of destination-specific fixed costs for exporting. In line with these observations, the predictions of the Melitz model both in regard to productivity gains and expected trade patterns have received considerable empirical support.

By comparing pre- and post-liberalization situations, a range of studies have examined the impact of trade reform on average industry productivities and the driving forces behind such developments

– in particular, the rates of firm survival and the relationship between plant productivity and the likelihood of firms to exit the market when confronted with increased competition.

Bernard et al. (2006b) test the predictions of the Melitz model on a panel of approximately 234,000 plants in 337 US manufacturing industries for the 1987 to 1997 period. They find that lower trade costs (resulting from an assumed reciprocal decrease in tariff and freight rates) indeed lead to higher aggregate industry productivity growth. In line with heterogeneous firm theory of intra-industry trade (Melitz, 2003; Bernard et al., 2003; Yeaple, 2005), the probability that a plant will go out of business rises with falling trade costs, and this probability is lower for high-productivity plants.⁷¹ Hence, the exit of lower-productivity firms provides an explanation for the rise in average industry productivity. Baggs et al (2002) are able to obtain similar results in relation to plant survival and sectoral productivity improvements for Canada in the context of the Canada-US Free Trade Agreement (between 1984 and 1998) and Muendler (2004) for Brazil in relation to liberalization efforts between 1989 and 1998.

In an attempt to distinguish the productivity impact of various economic policies, Eslava et al. (2005) find that it was trade opening, and not financial and tax reforms carried out simultaneously between 1982 and 1998, that increased the likelihood of firms exiting the market, in particular for low-productivity plants, and contributed substantially to overall productivity improvements in Colombia. Besides firm selection, Baggs et al. (2002) are also able to confirm a “share-shifting” effect, with Canadian tariff reductions leading to a reallocation of labour resources towards more productive firms. Similarly, Pavcnik (2002) observes aggregate productivity improvements in most sectors in Chile following a range of liberalization measures taken between 1979 and 1986 and is able to attribute sectoral productivity growth to both the exit of less productive firms and the reallocation of resources and market shares from less to more productive firms. As mentioned above in Section C.2.c, Tybout and Westbrook (1995) also find some evidence for shifts in market shares towards more productive firms in Mexican manufacturing sectors that were comparatively more open to trade.

Most of these studies also estimate the impact of falling trade costs on plant level (as opposed to average industry) productivity, although no

such effects can be derived from the basic Melitz model.⁷² According to Bernard et al. (2006b), the data on US firms only give a weak indication that plant productivity goes up when trade costs come down. Pavcnik (2002) observes productivity gains for firms in import-competing sectors and an increasing productivity divergence compared with firms producing non-tradables, but does not find further productivity increases for exporters (which are, however, as expected, more productive initially). For Canada, the evidence is mixed as well, with plant productivity growth also depending on a firm's initial level of productivity (plants at lower productivity levels making further progress) and on the separate impact of US as opposed to Canada's own tariff reductions (leading to productivity increases in high and low-productivity firms respectively) (Baggs et al., 2002; Trefler, 2004).

In order to explain the effect of trade opening on plant productivity, certain assumptions of the Melitz model need to be modified. This is what Bustos (2007) attempts to do by allowing for the possibility that trade liberalization prompts firms to invest and upgrade their technology in order to improve their productivity, an idea first developed by Yeaple (2005), as presented in Section C.3.b above. Looking at a panel of about 1,400 Argentinian firms and at a phase of trade liberalization between Argentina and Brazil from 1992 to 1996, she finds that companies in sectors benefiting from a comparatively higher reduction in Brazil's tariffs were more likely to export and increase their technology spending than firms in industries where trade opening was less ambitious. By demonstrating that both existing and new exporters seek to increase their productivity, she is also able to establish that when tariffs come down, it is the prospect of higher revenues from exporting that causes firms to invest in better technology rather than an exposure to new techniques and know-how from abroad.

Hence, Bustos (2007) shows both formally and empirically that the average productivity gains from trade at the sectoral level are not only explained by the exit of less productive firms (selection effect) and an expansion of market shares of the more productive firms (share-shifting effect) but also by the positive impact of participation in export markets on firm-level performance. A "learning-by-exporting" effect on firm-level performance has also been confirmed for a number of other emerging economies.⁷³ However, her approach only explains productivity improvements of highly productive firms targeting new export

opportunities, although technology upgrading could also play a role in helping firms at lower productivity levels to avoid exit from the market and, therefore, counter the selection effect to some extent.⁷⁴

Besides productivity gains, a second principal line of empirical research has shown that the Melitz model can best explain the hitherto neglected observation that liberalization increases trade not only within existing trading relationships ("intensive margin"), but also via an increasing number of exporters that have not traded before or via exports to destinations not previously targeted ("extensive margin"). Two principal research strategies can be distinguished: studies using firm-level data and studies employing aggregate trade data and an amended gravity approach (see Box 6 in Section C.2.c) that accounts for the many zero trade flows observed in bilateral trading relationships. Some of the studies mentioned above, such as Bernard et al. (2006b) and Muendler (2004), which examine the impact of trade opening on US and Brazilian firms respectively, confirm the prediction of the Melitz model that trade may grow at the extensive margin.

More precisely, Muendler (2004) is able to corroborate Melitz's proposition that among the firms not previously exporting, it is the high-productivity firms that become exporters following a reduction of trade costs. Both studies also find evidence that liberalization increases trade at the intensive margin. Bernard et al. (2006b) find that plants in industries that have undergone a relatively more important decline in trade costs experience higher export growth, while Muendler (2004) shows that existing exporters abandon exporting less frequently than before liberalization.

The Melitz model also had an impact on how gravity estimations should be conducted and, in turn, such studies could be used to probe into the model's explanatory power. According to the Melitz model, the absence of trade may be the consequence of firms' decisions not to enter an export market if their productivity level is not high enough to ensure that expected profits more than compensate for fixed market entry costs. Of course, zero trade may also be due to factors external to the firm, such as insufficient infrastructure. For these reasons, an assessment of the impact of liberalization and other policies on trade flows would be biased if only existing bilateral trade relationships were considered, as has long been the case, and/or if differences among firms were not taken into consideration.

Helpman et al. (2007) demonstrate that once zero trade flows are included in the estimations and the effects of trade barriers and country characteristics on the proportion of exporters are accounted for, higher trade volumes are not just a direct function of lower trade barriers. These higher volumes are also influenced in a more indirect manner by the increased proportion of firms that choose to export to any particular destination. The fact that the authors find these biases of traditional gravity estimations to be substantive provides strong support for Melitz's hypothesis on the importance of differences between firms in explaining international trade flows.

Baldwin and Harrigan (2007) use a gravity approach to determine which one of several theoretical approaches, including Melitz's specification of heterogeneous firms and Krugman's monopolistic competition model, most correctly predicts the effects of trade costs and country characteristics on the volumes and prices of internationally traded goods. The authors determine a higher probability of encountering zero trade flows with increasing distance and smaller market size and conclude that of the models examined, these results are consistent only with the Melitz model.

Baldwin and Harrigan also find that the positive relationship between export unit values and distance is only consistent with a more complex variant of the Melitz model, which is further discussed in the next sub-section (Melitz and Ottaviano, 2008). The latter model, however, predicts a negative relationship between the size of the destination market and the probability of exporting, in contrast to Baldwin and Harrigan's (2007) empirical results. The authors, therefore, propose their own adaptation of the basic Melitz model, which would also fit their finding regarding export unit values and in which firms are different in both quality and price. Firms producing high-quality/high-priced goods are most competitive and can more easily overcome distance-related trade costs, with the average goods prices in remote locations therefore being higher.

As this sub-section has shown, the Melitz model has been an important step forward in explaining observed trade patterns and the underlying gains from further trade opening. Furthermore, its basic framework has proven flexible enough to be extended in various directions and reconciled with other modelling techniques. This has allowed researchers to address a wider set of issues in

international trade, as will be discussed in the next sub-section.

(d) Extensions of the basic framework

Building on Melitz's basic ideas of differences among firms and fixed market entry costs, a number of authors have probed further into the presumed gains from liberalization and possible explanations of observed trade patterns.⁷⁵ Concerning the former issue, the question has been raised as to whether and how the predicted average productivity gains can be squared with the types of gains in the Krugman model discussed in Section C.2.b. This question is particularly relevant in light of the fact that in the Melitz model either no such effects exist (e.g. pro-competitive impacts in terms of reduced firm mark-ups) or, in the case of variety, the effect on the number of available varieties for consumption is ambiguous.

The similarity between trading partners is another important limitation of the Melitz model. The predictions concerning the gains from trade may need to be qualified if countries are dissimilar in terms of size or level of development (as emphasized by the traditional approaches discussed in Section C.1). As far as trade patterns are concerned, the flexibility of Melitz's approach has allowed researchers to address the question of how sectoral characteristics and changes in the broader trading environment determine a firm's decisions to trade rather than set up foreign operations or obtain components from abroad via arm's-length international outsourcing or foreign direct investment (FDI).

Melitz and Ottaviano (2008) have created an encompassing framework that combines the channels for the welfare effects identified by the "new" trade theory with the productivity gains from the selection and reallocation effects among firms that are different. In particular, this framework emphasizes that trade opening leads to a tougher competitive environment also for the surviving firms in the sense that average mark-ups are reduced (and, hence, prices are lower).⁷⁶ This is the case, since, in Melitz and Ottaviano's parameterization of the model, the direct effect of foreign competition on firm-level mark-ups outweighs the selection and share-shifting effects, leading to the survival of only the relatively more productive firms (with higher mark-ups than the less productive firms who exit).⁷⁷

Using data on several hundreds of thousands of firms from 11 European countries and 18 manufacturing sectors over the 1994 to 2003 period, Del Gatto et al. (2006) show that Melitz and Ottaviano's (2008) parameterization of the model is appropriate.⁷⁸ They then apply this model, asking by how much productivity and other performance variables would deteriorate if Europe was still in a state of autarky. Carrying out a simulation with prohibitive tariffs, they find that the costs of "Non-Europe" would be equivalent to a 13 per cent lower level of productivity, 16 per cent higher mark-ups and prices and 23 per cent lower profits on average. Starting from this hypothetical situation, they then examine the effect of a 5 per cent reduction in international trade barriers and obtain an overall productivity gain of 2 per cent along with a decrease in average mark-ups and prices of 2 per cent as well as an increase of average profits by almost 5 per cent.

Falvey et al. (2006b) extend Melitz's model in order to address the important question of whether the gains from trade with firms that are different remain invariably positive when the countries involved are of different sizes and at different levels of development. Regarding the latter, it is assumed that one country has a generally superior technology than the other, i.e. there are not only efficiency differences between firms but also differences at the national level. It turns out that opening up to trade generates the usual productivity effects in both countries, but that the distribution of gains across countries depends on their relative size and, in particular, their levels of technology. As far as size differences are concerned, the "home-market effect" (see Sections B.2 and C.1) is at work with a concentration of industry in the larger market, which also becomes a net exporter of the differentiated product. However, it is interesting to note that better technology can improve the situation of the small country, and, vice versa, i.e. the industry in the larger country may decline if its technology is sufficiently backward.

The most interesting result is that if countries are at different levels of development, the positive productivity effects of trade liberalization may not materialize for the technological "laggard" country. When trade barriers come down, both countries lose part of their domestic market, but exporting firms usually more than make up for the loss. However, if one country is technologically more advanced, it is harder for foreign firms to conquer

that market than vice versa. If sufficiently large technological differences exist, the productivity gains for firms in the technological leader from new exporting opportunities may be so large that they overcompensate the benefit that exporters in the technologically backward country receive from a decline in trade costs and, thus, result in a loss of competitiveness of the latter.⁷⁹ However, for less pronounced technology differences between countries, both sides continue to benefit from reciprocal trade liberalization.

The question of whether the technological gap can be as large as to reverse productivity gains for the technological laggard remains an empirical matter.⁸⁰ Regrettably, with the required firm data being mostly limited to advanced economies (United States and a number of European countries), no case studies appear to exist for the moment that could appropriately assess the impact of liberalization between different countries using information at the firm level.

Helpman et al. (2004) extend the Melitz model by introducing several sectors (instead of only one), some of which are characterized by more firm heterogeneity than others. Their goal is to show that the extent of the differences among firms in productivity plays an important role in explaining the structure of international trade, notably the volume of FDI sales relative to the volume of exports. To that end, they assume the existence of more types of firms. In addition to the less productive firms that are active only in the domestic market and the exporting firms that supply both the domestic and export markets, the authors distinguish a third group of very productive firms that choose horizontal FDI⁸¹ instead of exporting to sell their products in the foreign market. The authors assume that setting up a foreign affiliate (involving costs for establishing a subsidiary, duplication of overhead production costs, etc.) is more expensive than exporting (which entails, for example, only the establishment of a distribution and servicing network).

This set-up allows Helpman et al to predict that the ratio of exports to FDI will be lower in sectors with higher transport costs (since these costs can be saved via FDI)⁸² and in sectors where plant-level returns to scale are relatively weak (since not much is lost if production is not concentrated in one country). The ratio of exports to FDI resulting from this "proximity to market" vs. "concentration"

trade-off is shown to further depend on the degree of differences in each sector, i.e. the variety of firm size (measured by domestic sales), which, among other things, depends on how firm productivity is dispersed.⁸³ Less dispersion implies that the mass of firms is concentrated at the low-productivity/high-cost end of the distribution.⁸⁴ In other words, the more heterogeneous the sector, the more equal is the distribution of firm productivities. This implies that there are relatively more highly productive firms, and these firms will access foreign markets via FDI rather than exports. The authors then test empirically whether these three factors can indeed be seen as key determinants of the observed cross-sector and cross-country variation in export sales relative to FDI flows.

Using export and FDI data of US firms in 38 countries and 52 industries and different measures of size dispersion, Helpman et al are able to confirm the predictions of their model. In particular, it appears that differences between firms constitute as important a dimension in explaining the observed trade-off between exports and FDI sales as tariffs and freight rates (i.e. costs of exporting) and plant-level fixed costs (proxy for the importance of economies of scale) decline.⁸⁵

Finally, the basic idea of differences in firm productivity has been combined with theories of organizational choice and location of production decisions in order to address the observed importance of vertical FDI and arm's-length trade of intermediate goods in certain sectors. The paper by Antras and Helpman (2004), which has already been mentioned in the context of fragmentation in Section C.1.d and the "industrial organization" aspects of which will be more fully explained in Section D.3.b, is built around the core features of the Melitz model.

According to Antras and Helpman, firms vary in their productivities and can either make or buy their intermediate inputs domestically or in a foreign country, with each of these four options being associated with different fixed costs. Outsourcing involves search costs and contractual issues,⁸⁶ while vertical FDI may imply an increased need for monitoring, communication and other management-related costs (so-called "diseconomies of scope"), with the former being assumed to be less costly than the latter, both at home and abroad. Similarly, it is assumed that both outsourcing and integration are cheaper to conduct domestically than in

connection with a foreign provider (i.e. arm's-length international outsourcing or going multinational via vertical FDI). In view of this ordering of fixed costs, firms with relatively higher productivities pursue vertical integration over outsourcing and only the most productive firms do so abroad (where, again, the relatively less productive firms outsource at arm's length and the more productive ones invest in a foreign subsidiary).

This framework allows Antras and Helpman to make predictions about how changes in the sectoral environment, such as trade opening, affect the prevalence of organizational forms and therefore also trade. For instance, a decline in trade costs abroad (or a lower wage rate) leads to more arm's-length international outsourcing and, to a lesser extent, more FDI, and trade volumes are bound to rise. An improvement in the institutional environment abroad may make FDI more attractive relative to arm's-length international outsourcing without affecting overall trade flows; conversely, a better institutional framework at home, while also making integration preferable over outsourcing, including abroad, still leads to a net reduction in imports.

Antras and Helpman's approach can also serve to explain how differences across sectors determine organizational choice and trade patterns. Similar to Helpman et al. (2004) in relation to horizontal FDI, the authors demonstrate that in industries with more productivity dispersion across firms, less outsourcing should be observed both domestically and abroad and more vertical FDI will be undertaken, which increases overall imports. In industries with a relatively higher importance of non-routine activities, such as R&D, that are less readily outsourced than marketing tasks, for example, the authors expect a lower proportion of firms to import components, with the share of FDI still going up relative to arm's-length imports. These issues will be further discussed in Section D.2.b, specifically in relation to offshoring.

This section has discussed yet another source of gains from intra-industry trade, namely productivity improvements at the industry level that come on top of the variety, scale and pro-competitive gains introduced in the preceding section C.2. The analysis has been made possible by the availability of detailed firm-level data confirming the existence of significant differences in firm characteristics and trading patterns. Although these gains are of a static nature (i.e. reflect a comparison of the

situation before and after trade opening in terms of resource allocation, product availability, prices etc.), access to a larger variety of intermediate inputs, increased market size and confrontation with foreign competition may also affect firms' incentives to innovate and invest in research and development.

In fact, some of the empirical studies reviewed in Section C.3.c have found evidence of increases in the productivity of individual firms following trade opening. With technological progress being the principal driver of long-term economic performance, trade has the potential to result in important dynamic gains as well. These issues will be examined in a comprehensive fashion in the following sub-section.

4. DYNAMIC GAINS

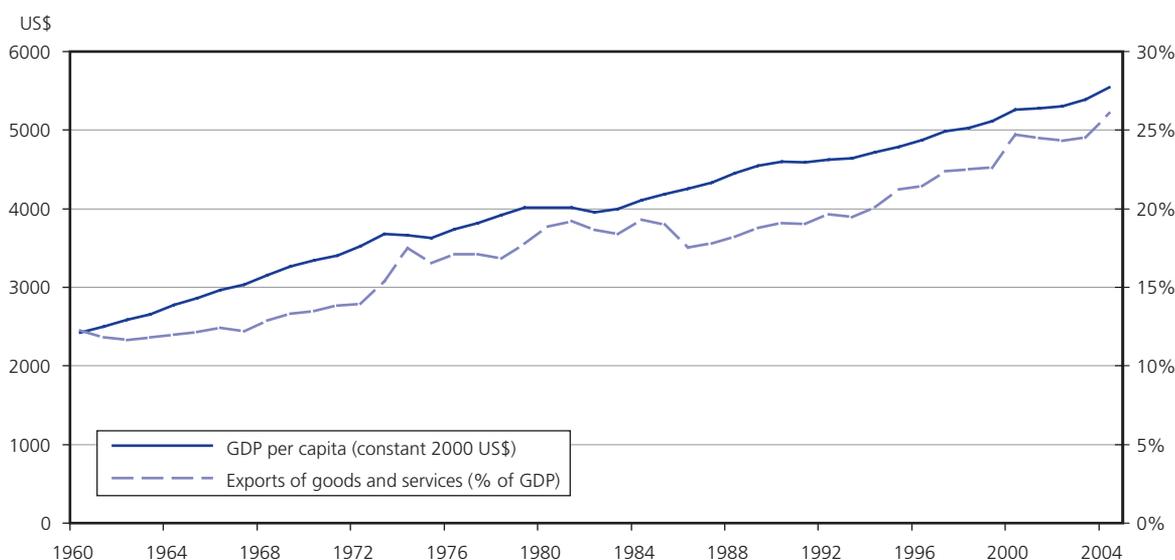
The analysis of the previous sub-sections has highlighted that in static approaches, the effect of trade liberalization is to increase real GDP at world prices. This is the result of an improved allocation of resources through specialization according to comparative advantage, exploitation of economies of scale and the selection of the most productive firms. But what happens when the analysis moves beyond a comparison of two static situations to consider

the more dynamic effects of policy changes? This sub-section will focus on the effects of trade on GDP growth, reviewing both the theoretical and the empirical literature.

(a) Trade and growth: an overview of theoretical predictions

The traditional theory of economic growth does not take international linkages into account. It is generally built on the assumption that countries produce and consume in isolation, so with no trade among them there can be no transfer of knowledge or technology across national borders associated with commercial relations. However, as discussed in Ventura (2005), the growth experiences of different world regions are intimately linked and cannot be analyzed in isolation. Three facts should be highlighted. First, the world economy has experienced positive growth for an extended period of time. Second, in the same period world trade has been growing at an even faster pace (see Chart 7). Third, the data illustrate a strong positive correlation between the growth of GDP and the growth in trade (see Chart 8). This correlation does not imply that one leads to the other,⁸⁷ but it reveals an important relationship between these two variables.

Chart 7
World GDP per capita and world exports, 1960-2004
(Per cent of GDP)



Notes: World GDP per capita is measured in constant 2000 US\$; World Exports contain goods and services exports and are presented as a share of World GDP.

Source: World Development Indicators 2007.

that by continuing to equip workers with more capital goods, a point will eventually be reached where additional capital becomes redundant. In this model, in the absence of technological progress (e.g. new and more efficient uses for capital), the effects of diminishing returns would eventually cause economic growth to cease. In other words, the neoclassical model of economic growth displays a sustained positive rate of growth in per capita output only in the presence of an exogenous improvement in technology (hence, the name “exogenous growth model”).

What shapes technological improvements and, therefore, long-term economic performance? Endogenous growth theory formalizes the determinants of technical progress and its role in the process of economic growth. Arrow (1962) provides a first attempt to explain technical change by assuming that growth in technology is an unintended by-product of the production of new capital (a phenomenon named “learning by doing”). Learning by doing allows growth in technology to become endogenous in the sense that increased capital accumulation (for instance, due to a surge in the propensity of saving in a country) would affect the rate of technological change and, therefore, the rate of GDP growth. One limitation of the learning by doing approach is that technological change does not depend on deliberate economic decisions.

Recent research provides an explanation of intentional investment in innovation based on a well known argument by Schumpeter (1942). New technologies provide market power and investment in innovation is motivated by the prospect of future profits. Romer (1986) employs a model of monopolistic competition (Dixit and Stiglitz, 1977; Ethier, 1982) where there is a continuum of intermediate goods used in the production of the final good. Each intermediate good is produced by a local monopolist. Romer shows that growth in technology in this framework is the result of a continuous increase in the variety of intermediate goods. The underlying assumption is that a

large number of intermediate inputs raises the productivity of capital and labour (for instance, because of the increased specialization of labour across an increasing variety of activities). Romer (1990) extends this model by assuming that inventing new goods (by investing in R&D) is an intentional economic activity motivated by the pursuit of profits. Firms that intend to enter a new intermediate sector must pay a sunk cost of product development, which is compensated by monopoly profits (rents). This model exhibits increasing returns to scale in the intermediate good sector and endogenous growth in output per capita.⁸⁸

In the Romer model, a technological innovation consists of a new good that does not displace existing goods (“horizontal innovations”). However, technical change can also take the form of an improvement in the quality of an existing good (“vertical innovation”) that makes old goods obsolete. This obsolescence (or “creative destruction”, as Schumpeter (1942) first referred to it) was initially formalized by Aghion and Howitt (1992) and Grossman and Helpman (1991). Growth is generated by the R&D activities of new entrants aimed at introducing higher-quality products in the market that will replace the products of incumbent firms. These two approaches (horizontal and vertical innovation) are best seen as complementary, as they describe different contexts in which technological progress takes place.

Recent developments in economic growth literature have emphasized the role of economic institutions (e.g. property rights) and political institutions (e.g. the form of government) in the growth process. The underlying idea is that these institutions affect the organization of production and shape the ability of firms to accumulate and innovate (or adopt more advanced technologies developed elsewhere) and, ultimately, determines the growth rate of a country (Acemoglu, 2008; Helpman, 2004).

i) Trade and factor accumulation

Models that analyze the effects of international trade on the accumulation of capital have two key insights. First, when economies are open, growth in one region cannot be analyzed in isolation from the growth experience of other regions. Second, the interaction between trade and growth depends on the nature of trade in which countries engage.

When international trade is in intermediate goods (i.e. trade in inputs used in production of final goods) and is determined by differences in relative factor abundance across countries (as in the model of Heckscher-Ohlin in Section C.1.b), the prices of factors are determined in world markets. In a world with free trade, factor prices equalize and each country (if small relative to the rest of the world) will take these prices as given when it makes accumulation decisions. This simple result has important implications for economic growth, as factor prices shape incentives to invest and accumulate capital.

In particular, as shown in Ventura (1997), the growth process in the presence of international trade can be quite different from that predicted by the (closed economy) neoclassical growth model. Even if an integrated world economy cannot sustain growth through capital accumulation only (i.e. it is still subject to the law of diminishing returns), periods of exceptional growth may be achieved by small open economies through savings and investment. If a small open economy adopts policies that increase its investment rate, it can accumulate capital without affecting its relative price (which is determined in international markets), thus avoiding diminishing returns due to capital accumulation.

As argued by Ventura, an example of this mechanism at work is the growth miracle of the East Asian tigers in the 1970s and 1980s, where small open economies were able to grow with high investment rates for an extended period of time. According to the predictions of the traditional neoclassical growth theory, rapid capital accumulation in East Asian countries should have been associated with falling rates of return on investment in new capital (the law of diminishing returns). However, international trade made the difference. As capital stock grew, resources were shifted into capital-intensive industries whose output was for a large share exported. This allowed an extended period of (export-led) growth.⁸⁹

The interaction between international trade and economic growth is different when trade is driven by technological comparative advantage (as in the model of Ricardo in Section C.1.a). Acemoglu and Ventura (2002) build a model where each country specializes in the production of a single intermediate good used in the production of a final good (the so-called Armington technology). This implies that countries are small relative to the global market, but have market power in the goods they supply to the rest of the world. In particular, each country affects its terms of trade (the price of imports relative to exports) by varying the amount of exports. If, for instance, an economy increases its supply of a good in the international market, the price of that good declines, implying a worsening of the terms of trade of the country (an increase in the price of imports relative to exports).

Consider the effect of rapid capital accumulation and, therefore, growth in an open economy relative to the rest of the world. Faster growth increases exports and worsens the terms of trade which, in turn, reduces the rate of return on capital and moderates incentives to accumulate. The same (but opposite) effect would be at work were a country to experience a fall in its growth performance relative to the rest of the world. In other words, a model of Ricardian trade and growth implies that open economies will tend to experience similar growth rates in the long term (clearly, similar growth rates do not necessarily entail similar income levels across different regions of the world).

A lesson that can be drawn from these models is that the type of international trade a country engages in shapes its growth performance and the dynamic effects of trade liberalization (Acemoglu, 2008). Countries that specialize in the production of standardized goods, exploiting relative abundance of a factor of production (e.g. unskilled labour) can expect to face sustained growth through the accumulation of capital (i.e. with high savings and investment rates) as predicted by the model of Ventura (1997). On the other hand, countries that trade in highly specialized sectors (e.g. high-tech industries), where they are likely to have some degree of market power, are likely to face terms of trade effects of the sort described in Acemoglu and Ventura (2002).

The models discussed in this sub-section predict important effects of international trade on economic growth. However, such effects are bound

to disappear in the long term. The reason is that these models do not take into account the role of international trade on the major factor determining the growth process (improvements in technology). This will be the focus of the next two sub-sections.

ii) Trade and learning by doing

Does international trade encourage sustained growth? As economic growth in the long term depends on technological advances, models of endogenous growth are needed to address this question. These models highlight the major factors determining technological progress, which in turn lead to improvements in the productive use of factors of production and growth. This sub-section starts by reviewing models that provide a sceptical answer to the question of the dynamic effects of trade liberalization (Young, 1991; Matsuyama, 1991; Galor and Mountford, 2006). The argument is based on the presence of learning by doing externalities in some sectors of the economy – that is, improvements in productivity in the overall economy engendered by the experience of producing in some specific sectors. The next sub-section analyzes the effects of international trade on incentives to innovate (i.e. to invent new or higher-quality goods and/or new or more efficient production processes).

Consider for simplicity a world with two countries (A and B) and two sectors (1 and 2). Sector 1 is characterized by some form of learning by doing externalities so that increasing production in this sector augments the overall productivity of the economy. Such opportunities are not present (or only to a smaller extent) in the other sector. Assume further that Country A has a small initial comparative advantage in the production of Sector 1. International trade in this economy leads each country to specialize fully in the sector in which it has a comparative advantage from the first period, implying static gains from trade (as in the standard Ricardian model). However, in subsequent periods, the productivity of the country specialized in Sector 1 (Country A) increases, while the productivity of the other country stagnates. The reason is that, by specializing in Sector 2, Country B loses the opportunity to exploit the learning by doing externalities of Sector 1.

In short, in these models with learning by doing, trade has different effects. The static benefits of international trade may come at the cost of dynamic losses for some countries and gains for

others. Because of initial conditions, trade prompts a number of countries to specialize in sectors of the economy with low growth potential. As a result, these countries may be negatively affected by opening their economies to international markets. These models are often mentioned as an argument in favour of trade restrictions in countries that fear dynamic losses from trade liberalization.

While it is a clear theoretical possibility, the relevance of the “wrong specialization” argument hinges on two assumptions that do not have strong empirical foundations. First, specialization in the “wrong” (or “right”) sectors is a permanent feature of trade relationships. However, as the example discussed for Chinese Taipei in Box 9 on “Success stories of export-led growth” shows, export structures may well change over time. Second, and more importantly, these models assume that knowledge does not flow across countries. However, as discussed in Section C.4.b, international trade is associated with an increase in knowledge spillovers. This implies that, contrary to what is predicted by these models, it is certainly possible that knowledge developed in Sector 1 in Country A positively affects productivity (and, therefore, growth) in both sectors in Country B.

iii) Trade and innovation

A different avenue through which trade can affect economic growth is by reinforcing or dampening incentives for firms to innovate. There are several mechanisms at work. First, trade liberalization increases the size of the market (scale effect). Second, to the extent that knowledge travels with the exchange of commodities and inputs, trade liberalization enlarges the scope of knowledge spillovers. Third, an increase in the degree of openness of an economy will typically enhance product market competition (competition effect). Fourth, decreasing trade barriers affect the distribution of production in different areas of the world (international product cycle). Last, trade liberalization may have an influence on institutions and government policies that shape economic incentives of firms. All these mechanisms affect economic growth through their effect on technical change.

Trade, market size and knowledge spillovers

The core message of groundbreaking work on endogenous growth in open economies is that trade

liberalization is conducive to an improvement in long-term economic performance (Rivera-Batiz and Romer, 1991; Grossman and Helpman, 1991). The effect of trade on growth in these models can be seen as the dynamic analogue to the static gains from trade of Krugman's model in Section C.2. Consider two economies trading intermediate goods used in the production of a final good. The possibility of freely trading with the other economy creates a larger market, thus increasing profitable opportunities for producers of intermediate goods. This greater profitability increases incentives to invest in R&D and translates into a higher rate of innovation (i.e. the rate at which new varieties are introduced) and a higher growth rate for the global economy.

This basic insight has been extended in a large number of directions. Some of these extensions reinforce the positive effect of trade on the pace of endogenous technical change. This is the case of models where inputs produced in one country increase the productivity of R&D in the other country after trade liberalization (international knowledge spillovers). In particular, if under free trade R&D activity in each country benefits from the discoveries made in other countries, trade liberalization will increase incentives for firms to engage in research, thus boosting economic growth. However, other extensions of this framework point to the presence of potential counteracting effects. These might originate from differences across countries in human capital (Grossman and Helpman, 1991), size (Feenstra, 1990) or the initial stock of knowledge (Devereux and Lapham, 1994).

Moreover, endogenous growth models do not take into account the differences between firms. As shown in Baldwin and Robert-Nicoud (2007), once firm-level differences are taken into account, trade can either increase or reduce the rate of innovation and, ultimately, economic growth. It is important to emphasize that, notwithstanding the presence of such counteracting effects, the ultimate result of trade on innovation may still be positive, provided that there are large international knowledge spillovers. This further motivates the significance of empirical investigations into cross-border knowledge flows (discussed in Section C.4.b).

Trade and competition

There is a general consensus that international trade reduces domestic firms' market power and that static gains are associated with this pro-competitive

effect (Helpman and Krugman, 1985). Whether an increase in competition associated with a wider exposure to international markets has positive effects on firms' incentives to innovate has been the object of a recent academic debate. From a theoretical point of view, this affect is ambiguous. The traditional work relating competition and growth dates back to Schumpeter (1942) and predicts that competition should reduce innovation as it lowers the monopoly rents that a successful innovator can expect to enjoy. However, several recent theoretical works show that competition can have positive effects on incentives to innovate. Aghion et al. (2005) build a model where firms may choose to invest more in R&D in an effort to escape competition. In this framework, more intense competition leads to higher R&D investment and innovation. Peretto (2003) obtains a similar result in a trade model where a reduction in tariffs increases firms' exposure to foreign competition and increases incentives to invest in cost-reducing innovation that allows firms to cut prices and acquire larger market shares.

Ultimately, the relationship between competition and innovation is an empirical question. The literature has pointed to a positive relationship between product market competition and innovation (Nickell, 1996; Blundell et al., 1999) and, more recently, to a hump-shaped relationship (Aghion et al., 2005), where the Schumpeterian (i.e. negative) effect on innovation tends to dominate only for high levels of competition. A possible interpretation of the available empirical evidence is provided by Aghion et al. (2005). They argue that the effect of competition on growth actually depends on the technological characteristics of a sector or an industry – specifically, the distance of the sector from the world technology frontier (i.e. the most advanced technology). In particular, the effects of lowering barriers to international trade on innovation and growth are positive overall, but trade liberalization may hurt some sectors that are further from the technology frontier (Aghion and Griffith, 2007).

Trade, innovation and imitation

The discussion up to now has focused mostly on the effects on innovation of trade between similar countries. However, international trade also links countries that are at different stages of development and this type of “North-South” trade is the fastest growing component of world trade. Accordingly, an important question is whether trade between

countries at different stages of development fosters world growth.

An important empirical fact is that most innovations take place in a small number of advanced economies and are later transferred to the rest of the world.⁹⁰ The presence of international trade enriches the process of technology diffusion in several ways. In particular, trade determines a process of an “international product cycle”, whereby certain innovative products previously produced in technologically advanced economies are imitated and produced in less developed countries. A case has been made both for and against strong intellectual property rights (IPRs). On the one hand, less imitation stimulates new inventions by increasing the expected reward to innovators. But on the other hand, this raises the cost of innovation by increasing wages in more advanced economies (as more production will be carried out there) and by limiting the entry of new competitors (Helpman, 1993; Maskus, 2000).⁹¹ Recent cross-country evidence, however, generally points to a positive effect of IPR protection on growth and on investment in physical capital and R&D (Gould and Gruben, 1996; Ginarte and Park, 1997; Kanwar and Evenson, 2003; Falvey et al., 2006a).

Another important consideration relates to the fact that less advanced economies have different technological needs from more advanced countries and that innovations developed in the latter group may be inappropriate for the former (Acemoglu and Zilibotti, 2001). Weak IPRs in less advanced economies encourage firms in advanced nations to target the needs of their own markets, thus lowering productivity in sectors where developing countries are specialized. This negative growth effect may spill over to advanced economies, thus lowering world growth. Bonfiglioli and Gancia (2007) formalize this mechanism and provide supporting evidence. Overall, this literature suggests that a proper design of IPR rules may well foster beneficial technical change and increase growth (Maskus, 2000).

Trade, institutions and policies

The quality of the institutional framework has long been recognised as an important component of a well-functioning market. A large number of recent studies highlight the importance of a sound institutional framework for economic growth (see Box 8). Countries with “better institutions” tend to invest more in infrastructure, training and education, are more efficient in this respect and innovate more.

A key question is whether trade can have a positive effect on institutions that promote growth (Rodrik, 2000b). Several direct and indirect channels may be at work. First, trade reform may imply adoption of external requirements. For instance, membership of the WTO requires the embracing of certain institutional norms (e.g. transparency in trade policy, and rules on industrial policy or property rights consistent with WTO provisions). Tang and Wei (2006) find that WTO/GATT accession has favourable effects on growth by committing countries to policy reform. Second, institutional reforms (and the preferences that underlie such reforms) may be the indirect consequence of the workings of market forces associated with trade. Acemoglu and Robinson (2005) show that democratization may result from changes in income distribution associated with trade liberalization. The empirical work of Rodrik et al. (2004) confirms that international trade has a positive impact on the institutional framework, thus suggesting that trade liberalization can indirectly enhance growth performance by improving institutional quality.

If trade can improve the institutional framework, its effect on government policies is in principle ambiguous. While economic integration is rapidly advancing, international political integration is advancing at a slower pace or not advancing at all. The increase in international spillovers associated with trade may prompt national governments to choose worse policies as they do not fully internalize the effects of their decision on the rest of the world (Ventura, 2005; Epifani and Gancia, 2008). An example is government subsidies, when at least part of these subsidies are redistributed by governments in response to rent seeking (i.e. activities of firms, such as lobbying or bribing, aimed at obtaining private benefits from policy-makers). When firms engage both in R&D and rent-seeking activities, trade liberalization – by increasing competition in the market – may prompt firms to increase their demands for public subsidies and, therefore, to devote more resources to rent-seeking and less to R&D. This shift from a productive to an unproductive activity may reduce innovation and growth (Brou and Ruta, 2007).

(b) Empirical evidence on trade and growth

The previous sub-section shows that, while the presumption is that free trade entails dynamic gains,

recent theoretical models point to the existence of several offsetting effects. The impact of trade policy on economic growth is thus an important empirical question, which is the focus of the present sub-section. This question is tackled from three points of view: the cross-country evidence on the impact of trade on growth rates, the microeconomic evidence on the effect of trade liberalization on firms' productivity, and the evidence on the relationship between trade and international knowledge spillovers. While the conclusions of the macroeconomic evidence have been questioned recently, the evidence on knowledge spillovers and firm productivity provides a more clear-cut (even if indirect) answer about the positive effect of international trade and growth.

i) Macroeconomic evidence (openness and growth in cross-country analyses)

As highlighted in the introduction, there is a high correlation between the rate of economic growth and the volume of international trade. However, these data do not necessarily imply that trade leads to growth. In the past few years, a large body of literature has investigated this relationship. Although many of these studies find a positive effect of several measures of international trade on economic growth, this macroeconomic literature can hardly be seen as conclusive.

Researchers have focused on two different trade-related indicators, namely trade volumes and trade policies. This distinction appears to be crucial, since changes in trade policy (e.g. changes in tariff rates) are often not accompanied by a change in trade volumes (Dollar and Kraay, 2002). On the other hand, a change in trade volumes can take place without a policy change (e.g. increased trade volumes due to a reduction in transport costs).

Trade volumes and economic growth

A number of studies (Sachs and Warner, 1995) find a positive relation between volumes of international trade and economic growth,⁹² even after taking into account other differences across countries and controlling for their effects. Moreover, the relationship appears to be extremely robust regarding different trade model specifications (Levine and Renelt, 1992) when considering all countries for which reasonable data are available. However, these studies do not provide a convincing answer when it

comes to determining whether trade causes faster growth or whether economies that grow quickly also trade more.

Recent work (Frankel and Romer, 1999; Ades and Glaeser, 1999; Alcalá and Ciccone, 2003; Alcalá and Ciccone, 2004) directly addresses this problem of causality using econometric techniques. In particular, Frankel and Romer (1999) find that higher volumes of trade drive economic growth. However, a consensus on the reliability of the results has not yet been reached. The method used by Frankel and Romer (1999) relies on the assumption that country geographical characteristics influence growth exclusively through the trade channel. However, critics of this approach – most notably Rodríguez and Rodrik (1999) – point out that geographical and other characteristics of countries may affect growth through various other channels (for instance, the presence of disease may affect public health, the quality of human capital and therefore growth performance).

Trade policy and economic growth

A different set of papers focuses on the link between trade policy and growth. This is obviously of interest as policies can be directly changed by governments. However, studies have shown mixed results. The main difficulty with this approach relates to the availability of suitable measures of trade policy. First, detailed data on trade restrictions are scarce before 1985. Second, many types of trade restrictions (tariffs, quotas, embargos, import and export licences etc.) are set up differently across countries. This implies substantial difficulties in systematically measuring these indicators of trade policy. Not surprisingly, results obtained in this literature have turned out to be dependent on the nature of the trade policy.

One example is the measure constructed by Sachs and Warner (1995) named “number of years of open economy”. Using this variable, Sachs and Warner (1995) found that trade openness is associated with higher growth rates. Several later studies confirmed this result (Sala-i-Martin, 1997; Fernandez et al., 2001). However, these findings strongly hinge on the fact that this trade openness indicator includes a black market premium on exchange rates and state export monopolies as trade barriers (Rodríguez and Rodrik, 1999). Without these variables, no conclusion on the relation between trade policy and growth can be drawn. More recently, Wacziarg

and Welch (2003) built an alternative measure of openness and report a robust positive relation between trade liberalization and growth. The main advantage of their measure is that it allows them to exploit not only differences across countries but also changes over time within countries.

Apart from the measurement challenges mentioned above, a debate continues on the existence of causal links and the direction of any causality (as already discussed for trade volumes). Since government policy responds to economic and political objectives, it cannot be viewed as independent and therefore the distinction between cause and effect is again not clear (Rodrik, 2005). Moreover, it is often difficult to disentangle the effects of trade liberalization from other domestic policies that governments may implement simultaneously and which may also have important effects on growth (Rodriguez and Rodrik, 1999).

ii) International knowledge spillovers

As the discussion of theoretical models emphasized, international knowledge spillovers are crucial for the realization of the dynamic gains from trade. An important area of empirical research investigates the relevance of international flows of knowledge, finding support for the proposition that R&D carried out in one country has positive effects on trading partners.⁹³

In order to study the importance of trade for the diffusion of international technology, a general approach of the empirical literature is to study the effect of foreign technology (as measured by foreign R&D, patents, etc.) on total productivity of the factors of production. The literature also stresses the importance of the sectoral composition of imports in determining technology spillovers. The empirical research distinguishes between imports originating from developed countries and those originating from developing countries, and between imports of technology-intensive goods and imports of non-technology-intensive goods. The idea is that imports originating from industrial countries have a higher embodied technology content than imports from developing countries. Similarly, imports of capital goods or imports of machinery and equipment have a higher average technological content than total manufacturing goods.

A benchmark study was conducted by Coe and Helpman (1995) in which they found that technological spillovers are higher when a country imports relatively more from high rather than low-knowledge countries (i.e. there is an import composition effect). Second, they concluded that, for a given composition of imports, technology transferred from abroad is greater the higher the overall level of imports. Coe et al. (1997) extend this analysis to examine the diffusion of technology from highly industrialized countries to 77 developing countries. They show that total factor productivity in developing countries is positively and significantly related to the R&D in their industrial country trading partners. They show that this effect is stronger when employing machinery and equipment import data as opposed to overall manufacturing or total (goods and services) import data. Several other studies confirm that international knowledge spillovers are more important when imports have a larger content of capital goods and machinery (Xu and Wang, 1999; Gera et al., 1999).

In addition to the “direct” (i.e. bilateral) R&D spillovers which are related to the level of R&D produced by the trading partner, “indirect” spillovers may exist even if countries do not trade with each other. A simple example clarifies this point. Assume that there are three countries: A, B and C. It is possible that Country A benefits from the technology produced in Country C, without importing from this source, if Country B trades both with A and C. Lumenga-Neso et al. (2005) find that “indirect” trade-related spillovers are empirically as important as the “direct” ones discussed above. This result points out the importance of an open multilateral trading system for the diffusion of technology, as what matters is not how much knowledge a country’s trading partner produces but how much knowledge the trading partner has access to through trade relations with the rest of the world.

Some empirical studies that examine the evidence of international technology diffusion use patent data (rather than R&D statistics) as a measure of technology transfer. Sjöholm (1996) examines citations in patent applications of Swedish firms to patents owned by inventors in other countries. He finds a positive correlation between Swedish patent citations and bilateral imports, a result consistent with the hypothesis that imports contribute to international knowledge spillovers. Branstetter (2001) extends this technique to consider firms

both in the United States and Japan. This study highlights the possibility that knowledge spillovers may be asymmetric – that is, while there is no evidence that American firms benefit from the R&D activity of Japanese firms, the reverse is generally true. Finally, Bottazzi and Peri (2007) find evidence that R&D in the United States increases patent applications in trading partners. Their data suggest that such effects are stronger after five to ten years (i.e. there is a delay in the diffusion of knowledge spillovers).

iii) Microeconomic evidence (trade and firm productivity)

Recent evidence on the relation between trade and productivity growth focuses on firm-level data. The effects of trade policy reforms are an essential indicator of what is happening to an economy as a whole, as aggregate growth is a reflection of improvements at the firm level. While the debate on the macroeconomic effects of trade on growth is still quite open, the microeconomic evidence provides more clear-cut answers on the dynamic gains from trade.

Firms that engage in exporting are generally more productive than non-exporters. An important question is whether exporting has a direct impact on firms' productivity (and, hence, on economic growth). A related question concerns the effects of trade liberalization episodes on the productivity of domestic firms. These issues are briefly considered below (for more extensive analysis, see Section C.3).

Early studies find no evidence of improved productivity at the plant level as a result of beginning to export (Bernard and Jensen, 1999; Clerides et al., 1998). However, more recent research using firm-level data from several developing countries reveals that exporting is associated with a boost in firms' productivity.⁹⁴ These findings may suggest that the aggregate gains from exporting are larger for emerging market economies than they are for advanced countries due to the presence of some form of "learning by exporting". In addition, several studies conducted for different trade liberalization episodes in both developed and developing countries provide new evidence on the positive effects of trade on firm productivity.⁹⁵

Box 9 Success stories of export-led growth

A series of important studies in the early 1970s demonstrated the high cost of protectionism in developing countries (Little et al., 1970; Balassa and Associates, 1971). They set in motion a major rethinking of the role of trade in development. The idea that trade can become an engine of growth was accentuated by the success of a number of economies in East Asia. Starting in the 1960s, Chinese Taipei, South Korea, Hong Kong and Singapore used exports to promote sustained growth and industrial transformation, as Japan did at the end of World War II. While these five countries represent only 4 per cent of the world population, they have become an important pillar of the modern industrial world and of the international trading system in a short period of time. Firms from this region are leaders in industries such as electronics, shipbuilding and automobiles and overall these five economies continue to grow faster than many other industrialized countries.

Once these economies moved away from import-substitution regimes and made firm commitments to more export-oriented and outward-looking policies, the pressure of global competition pushed them to keep costs low and to achieve higher and higher levels of performance. Coupled with other important ingredients (e.g. efficient administration based on merit, a large labour force eager to improve its skills and living standards, large levels of aid from international organizations and technology transfers from the United States), this policy shift resulted in export growth rates reaching 20 per cent (or more) per year over extended periods of time. According to Vogel (2008), the response to the challenges generated by more outward-oriented policies created a "virtuous circle". Companies and workers increased their effort to learn more, improve standards, boost productivity and absorb new technologies. The ability to compete in international markets increased confidence, which fuelled new rounds of efforts and success. In each round, more

Changes in the product structure of Chinese Taipei's merchandise exports, 1963-2003

(Percentage shares)

	1963	1973	1983	1993	2003
Agricultural products	59.3	15.4	8.0	5.1	2.5
Mining products	2.7	0.9	2.4	1.7	3.5
Manufactures	38.0	83.6	89.1	93.0	93.7
Iron and steel	3.0	1.3	2.5	1.6	3.9
Chemicals	5.1	1.5	2.4	5.1	8.1
Other semi-manufactures	11.7	12.1	11.6	9.6	6.8
Machinery and transport equipment	1.5	23.5	26.2	44.4	55.7
Office and telecom equipment	0.3	16.3	13.9	23.8	35.8
Electrical machinery and apparatus	0.3	2.7	3.6	6.5	8.1
Textiles	11.7	12.8	7.2	9.6	6.2
Clothing	3.0	16.1	11.9	4.4	1.4
Other consumer goods	1.8	16.3	27.4	18.4	11.6

Source: WTO Secretariat.

advanced knowledge, technologies and skills were acquired that allowed these countries to specialize in new sectors.

Chinese Taipei exemplifies this virtuous circle. In the 1960s it turned from traditional import substitution to a strong export-oriented development strategy. This policy shift led to increases in the average ratio of exports to GNP from 8.8 per cent in the 1950s to 18.5 per cent in the 1960s, 42.4 per cent in the 1970s and 50.3 per cent in the 1980s. Average GNP growth rates were 10.2 per cent in the 1960s, 8.9 per cent in the 1970s and 7.6 per cent in the 1980s. The table above shows how economic growth was accompanied by a change in the economy's export structure, moving away from agricultural products and textiles in the 1960s to clothes and "other consumer goods" (including toys and watches) in the 1970s and 1980s, and finally to office and telecom equipment in the 1990s. This indicates that international trade can play an important role in shifting an economy's resources into the most dynamic economic activities.

The most vibrant economy of recent years has been undoubtedly China. From 2000 to 2007, it alone accounted for 13 per cent of the world growth in output. This contribution is likely to increase in the next two decades according to recent estimates (Winters and Yusuf, 2007). From a historical perspective, the growth of China is unprecedented. When growth in the United States reached its peak in the middle of the 19th century, income doubled within a single generation. At current growth rates, income in China would rise a hundredfold within one generation. Similarly, the increase in China's share of world exports has no historical precedent. Analysts debate the extent to which the performance of the Chinese economy is export-led. There is little doubt, however, that its growth process shares important similarities with the economic history of the other East Asian success stories. First, the take-off coincided with a shift towards outward-oriented policies. Second, China has demonstrated over time the ability to upgrade its performance in more technologically advanced sectors.

5. CONCLUSIONS

Two centuries of economic research on trade have substantially improved our understanding of the gains from trade and thus the causes of trade. An overview of the economic literature shows that early contributions are still relevant but that they are insufficient. New theories have been developed

to account for new forms of trade and for new information on trade. The old, the new and the "new new" theories should be seen as primarily complementary. No single cause can explain the complex trading relations that we observe today. In 2017, economists will celebrate the anniversary of the first publication of David Ricardo's "Principles of political economy and taxation". After two

centuries they still consider the law of comparative advantage as one of the most important results in economics and perhaps even the “only proposition in social sciences that is both true and non-trivial”.

As explained, one of the main contributions of the law of comparative advantage has been to point out the fact that there are many more circumstances under which international trade is beneficial than most people appreciate. More recent theories have provided further causes for trade, identifying gains that can explain trade flows left unexplained by the traditional model. Recent empirical work confirms that beyond differences in technology and differences in endowments, other factors play an important role in the explanation of trade patterns. Finally, traditional and new trade theories have been extended to analyze the dynamic effects of trade on the economy. This line of research points out that another cause of international trade is its potential to enhance economic growth.

The overview of trade literature undertaken in this section focused on the causes of trade. It only touched upon a number of very important issues that trade models also address. Patterns of trade were only discussed in relation to the causes of trade. Patterns of trade will be discussed in more detail in Section D, which focuses on two phenomena that deeply affect trade flows: the geographical concentration of production and the fragmentation of the production process. Also, the predictions of the trade models regarding the distribution of income will be addressed in Section E. Finally, policy issues are addressed in Section F. The results reviewed in this section and the models presented will serve as a basis for these discussions. Table 7 below provides a summary of some of the basic characteristics of the trade models that were reviewed. It should facilitate the reading of subsequent sections.

Table 7
Trade theories

	Traditional trade theory Ricardo, Heckscher-Ohlin	“New” trade theory Krugman(1980)	Heterogeneous firms models Melitz (2003)
Gains from trade (causes)			
Specialization	Yes	No	No
Economies of scale	No	Yes	Yes
Pro-competitive	No	Yes	No
Variety	No	Yes	No ¹
Aggregate productivity (through selection/ reallocation)	No	No ²	Yes
Trade patterns			
Inter-industry	Yes	No	No
Intra-industry	No	Yes	Yes
Exporters and non-exporters within industries	No	No	Yes
Distribution			
Trade liberalization affects relative factor rewards	Yes	No	No

¹ Variety effects are ambiguous. See Section C.3.

² In the Krugman model, “productivity” in the integrated market also increases in the sense that the same total amount is produced at lower average cost due to exploitation of scale economies. However, the Krugman model is silent about which firms remain in business, since it does not include differences among firms. Once firms are distinguished according to their productivity level, as in the Melitz model, the exit of less productive firms itself leads to improvements in overall industry productivity.

Note: The table refers to the basic versions of the models (row 2). As discussed in the text, models that combine features from different approaches presented in the table have been developed.

Source: Based on Table 1 of Bernard et al. (2007a).

Endnotes

- 1 Normative trade theory is concerned with making welfare judgments about trade and trade related policies and events. (Corden, 1984)
- 2 For a simple explanation, see Markusen et al. (1995). For a more elaborate discussion, see Wong (1995).
- 3 This formulation is in Deardorff (2005a). See Corden (1984) for more details.
- 4 That is, production costs per unit of output are constant regardless of the level of output.
- 5 The assumptions required for perfect competition include the following: no supplier or consumer can affect prices (the complete absence of any monopolistic power in the market), any adjustment to change is instantaneous, and all consumers and producers are in possession of complete (perfect) information.
- 6 Externalities arise when economic activity produces effects that are not properly priced in the market place. A firm that pollutes the environment without paying for the social costs of pollution is generating a negative externality. A firm that generates new knowledge which others can avail themselves of without paying for the costs of generating the knowledge creates a positive externality.
- 7 Economies of scale arise when the cost of producing a unit of production falls with the level of output.
- 8 Sub-sections 2 and 3 consider other prominent theories that focus on different kinds of gains from trade and provide different explanations of the patterns and causes of trade.
- 9 The opportunity cost is the cost of choosing one option over another (sometimes the next best alternative).
- 10 Haberler (1930) showed that what is important for trade, the gains from trade, and comparative advantage, even in the Ricardian model, is not labor costs *per se*, but rather the opportunity cost, at the margin, of producing one good instead of the other. See Deardorff (2005b)
- 11 Dornbusch et al. (1977) simplified the Ricardian model with many goods. See Krugman and Obstfeld (2006).
- 12 This conclusion is not restricted to comparative advantage as embedded in the Ricardian model. It generalizes to comparative advantage embedded in the Heckscher-Ohlin model.
- 13 See Krugman and Obstfeld (2006) for a simple presentation of the Heckscher-Ohlin model and Jones and Neary (1984) for a formal derivation of the core propositions. Ethier (1984) states eight propositions reflecting the principal variants of these theorems.
- 14 See for instance Feenstra (2004), Bhagwati et al. (1998) and Ethier (1984).
- 15 “Stickiness” of prices (and wages) implies that these are not completely flexible. In other words, prices do not adjust quickly and “clear” markets, i.e. equilibrate supply and demand.
- 16 The effects of the three types of distortions are discussed in Bhagwati et al. (1998).
- 17 They test the Ricardo-Haberler-Deardorff theorem. This fully-specified multi-country, multi-good general equilibrium formulation of comparative advantage was developed by Deardorff (1980) and simply states that $p_{AT} > 0$ where p_A is a country’s N-good autarky price vector and T is its corresponding N-good net import vector. See Bernhofen (2005) for a summary of the test.
- 18 Kiyota (2007) suggest that the test of the law of comparative advantage by Bernhofen and Brown (2004) is not correct when trade is not balanced. They show that the empirical test of the law of comparative advantage sometimes fails when trade imbalance is taken into account.
- 19 As discussed below, empirical work on the Heckscher-Ohlin model highlights that technological differences are a major determinant of trade patterns.
- 20 Other explanations of the paradox included the fact that contrary to what Leontief assumed, US and foreign technologies are not the same, that the data for 1947 may be unusual because of the proximity to WWII or that the US was not engaged in free trade as the Heckscher Ohlin model assumes.
- 21 Trefler (1995) identified anomalies in the data which further research could aim to understand. In particular, he showed that measured net factor trade is approximately zero, a result he described as “the case of the missing trade”. See the overview of this literature in Davis and Weinstein (2003b).
- 22 The integrated equilibrium concept assumes that a world with imperfect mobility of factors of production across regions or countries may replicate the essential equilibrium of a fully integrated economy, provided that goods are perfectly mobile. This paradigm has played a central role in international economics.
- 23 Section C.1.e discusses how international factor movements fit into the theory of comparative advantage.
- 24 The model has two sectors (X, Y), two factors, labour L (immobile across nations) and K, which represents internationally-mobile capital services. K could represent managerial services that could be provided without any physical factor movement at all as in Deardorff (1985). X is labour-intensive, Y is intensive in capital services. .
- 25 As it is the case with models with more goods than factors, when both goods are tradeable there is a production indeterminacy due to the footlose nature of capital services.
- 26 An elasticity measures the responsiveness of one economic variable to another, usually comparing percentage changes. Here, an elasticity of unity implies that a one per cent increase in the trade of goods in related to a one per cent increase in services trade.
- 27 This sub-section draws on Baldwin (2006c).
- 28 The terms “intermediate goods”, “stages of manufacturing” or “production blocks” could be used instead of “tasks”.
- 29 See also the discussion of the contribution of the “new paradigm” to the explanation of the decision to offshore in Section D.
- 30 Consider a simple setting with 2 nations (H and F) and one final good, produced using two tasks. Assume H has a productivity edge over F in both tasks, and this edge is equal for both tasks (this rules out offshoring based on comparative advantage because H does not have a comparative advantage in either task). When it offshores a task to F, H uses its own technology. However, there are task-specific offshoring costs that deteriorate H’s technology when used in F. H firms will offshore a task if the wage gap (equal to the productivity gap in this framework) more than offsets offshoring costs. Offshoring will boost the real wage in H (productivity effect), while keeping the wage in F unchanged.
- 31 The traditional offshoring literature assumes that technology is country-specific, not firm-specific, so that a firm uses the technology of the country where it operates.

- ³² Krugman and Obstfeld (2006) discuss overall gains from trade in the specific factors model. Bhagwati et al. (1998) formally show that factor immobility cannot be an impediment to gains from trade. Free trade will still dominate autarky, because immobility between sectors eliminates the production gain but leaves open the consumption gain.
- ³³ These comparative statics results are formally derived in Bhagwati et al. (1998) for example.
- ³⁴ Markusen (1983) shows that trade and factor movements are complements when trade is caused by differences in demand, by increasing returns to scale, by imperfect competition or by domestic distortions.
- ³⁵ The Grubel-Lloyd index is calculated as $1 - \frac{|X - M|}{X + M}$, with X and M denoting exports and imports respectively. If, in a sector, a country is either only an exporter or only an importer, the second term will be equal to unity and, hence, the index will be zero, indicating the existence of inter-industry trade. Conversely, if in this sector, a country both exports and imports, the index will be closer to one the more similar in value imports and exports are.
- ³⁶ Helpman (1987) demonstrates that two countries of unequal size do not trade as much as two countries of similar size, holding constant the sum of both gross domestic products (GDP). Helpman's similarity index is defined as $1 - \left\{ \frac{GDP_i}{GDP_i + GDP_j} \right\}^2 - \left\{ \frac{GDP_j}{GDP_i + GDP_j} \right\}^2$, where GDP_i refers to real GDP. See Feenstra (2004: 146).
- ³⁷ In this section, only economies of scale internal to the firm are discussed, i.e. cost reductions depend only on the size of the firm itself. By contrast, average costs for any firm may also be lower the larger the industry to which it belongs (independently of the size of the individual firm). Such an industry is then characterized by external economies of scale. The concentration of an industry in one or few locations, like the semiconductor industry in Silicon Valley, will be further discussed below in Section D.2.
- ³⁸ Due to transport costs, the authors identify an optimal number and location of production sites. The trade-off between plant size and transport costs will be reverted to further below.
- ³⁹ To be clear on terminology, "differentiated product" refers to one type of product (say, ice-cream), for which different varieties exist (chocolate, vanilla, etc.). In other words, the term "varieties" refers to different forms of the same type of product.
- ⁴⁰ An oligopolistic market structure is likely to develop, where firms in their pricing decisions not only consider consumer demand, but also the expected responses of competitors to their behaviour. These responses, in turn, depend on competitors' expectations about the firm's pricing decisions and so on, giving rise to a complex web of strategic interactions. These issues do not affect the basic nature of the gains from trade in a setting with increasing returns of scale, while complicating matters tremendously.
- ⁴¹ This is known in the literature as the "zero profit condition".
- ⁴² For the moment, trade is supposed to be costless.
- ⁴³ The home market effect is further discussed in Section C.1.a.
- ⁴⁴ To maximize profits, a monopolist sets the price at a level, where marginal revenue equals marginal cost.
- ⁴⁵ In other words, with a lower market share in the foreign than domestic market, a firm's exports typically face a higher elasticity of demand than its domestic sales.
- ⁴⁶ The firm makes an additional profit as long as its exports sell at a price (net of trade costs) that, even if lower than the domestic price, is still higher than its marginal costs.
- ⁴⁷ Of course, by presuming Cournot competition, i.e. each firm expects that its own output decision will not affect the decisions of its rival, the model excludes the possibility that both former monopolists could collude and prevent mutual entry from happening.
- ⁴⁸ Krugman (1990) shows that these results continue to hold when there are more than just two firms.
- ⁴⁹ The underlying idea is the concept of consumer surplus. When a good is not available and demand is zero, it should be valued at its reservation price, i.e. the maximum price a buyer is willing to pay for it. When it appears in the market, there will be positive demand and the price will be less than the reservation price and the consumer welfare gain can be calculated as the area under the demand curve between the two price levels.
- ⁵⁰ Of course, the possibility that other varieties simultaneously disappear, leading to an increase in the price index, is also taken into account.
- ⁵¹ The median mark-up for the 21 developed countries in the sample is 1.60 and for the 21 developing countries it is around 1.90. This difference could be due to greater market uncertainty in light of which investors are likely to demand higher returns. Also, average tariffs tend to be higher and entry regulations more numerous in developing countries, which also explains the higher mark-ups.
- ⁵² The result that home and foreign tariff reductions have opposite effects on scale is consistent with the monopolistic competition model. Without foreign market opening, the market size for domestic producers stays unchanged, while the reduction of own tariffs increases the number of firms in the domestic market. The authors show that these patterns are also consistent with an oligopoly model of segmented markets and imperfect competition (Cournot) among firms, similar to the reciprocal dumping model discussed above, since home tariffs raise the delivered costs of foreign firms, inducing them to contract which and, hence, cause domestic firms to expand.
- ⁵³ The alternatives are the Armington model, i.e. national product differentiation by country of origin, and reciprocal dumping model with entry barriers respectively.
- ⁵⁴ In the terms of the authors, the monopolistic competition set-up is referred to as the "IRS" model indicating the existence of "increasing returns to scale" and product differentiation as opposed to the Heckscher-Ohlin framework assuming constant returns to scale ("CRS") and homogenous products. Their results reject the presumption of complete specialization for both the Heckscher-Ohlin and monopolistic competition models. Conversely, the authors find support for imperfect specialization of production among countries, whereby the degree of specialization is a function of relative factor abundance. In other words, the main comparison in the paper is between the Heckscher-Ohlin model with incomplete specialization and a framework that combines this approach for CRS goods with an IRS framework such that not all goods are assumed to be differentiated.
- ⁵⁵ This conclusion provides an explanation of the puzzle identified by Hummels and Levinsohn (1995) that was mentioned in the previous paragraph: the correlation between size similarity and trade volumes among OECD countries found by Helpman (1987) is due to the importance of IRS-based trade, while such a correlation among non-OECD countries can be explained with specialization driven by differences in relative factor endowments.
- ⁵⁶ For the time being, access to firm level data is still limited and the potential for improvement in this area

- is considerable. A first problem is mere existence of the data, as statistical agencies within countries must repeatedly perform firm-level surveys which are costly to administer. Second, if statistical agencies within a national government are decentralized, firm-specific information from one agency's survey (e.g., manufacturing activity) may be difficult to link to data collected by another agency (e.g., customs), let alone to data from surveys involving a different economic actor (e.g., matching individual-level surveys on labor market activity to firm-level surveys on manufacturing activity). Third, firm-level data that would be comparable across countries is in particularly short supply. For independent researchers, access to firm-level data is complicated for confidentiality reasons. In relatively concentrated industries, even using codes to mask the firm's name may not be sufficient as merely examining the data may reveal the identity of the firm. Firms would prefer this information not be available to the general public (more specifically, their competitors), as it may reveal details and confidential information about business strategies and trade secrets, for example. Thus, while the quality of firm-level data is quite high in the United States, the number of academic researchers working in this area is quite limited because of research barriers to entry demanded by such confidentiality concerns. For example, to access data on these firms from the US Bureau of the Census, non-government researchers have to go through a process of becoming government employees (nominally) and receive "Special Sworn Status," and even then access to the actual underlying data may be limited. Nevertheless, as the results of this exciting new area of research reveal, if more such data were collected and additional barriers to its access by economic researchers were removed, it would provide a tremendous amount of useful information with insights for policy formulation.
- This approach and results are consistent with Schott (2004), which exploits product-level US import data from high- and low-wage countries, showing that unit values within products vary systematically with exporter relative factor endowments and exporter production techniques – empirical facts that reject the factor proportions theory of specialization across products but which are consistent with such specialization within products. Nevertheless, the evidence on comparative advantage theory at the micro level may break down once we move across countries, as apparently the relationship between exporting and capital as well as skill intensity also holds in some developing country-level studies as well such as the one on Chile by Alvarez and Lopez (2005). This is inconsistent with a comparative-advantaged based theory that firms in such countries that would be taking advantage of relative endowment differences more intensively would be more unskilled labor intensive than non-exporting firms in developing countries.
- The technique used to estimate TFP is the one proposed by Olley and Pakes (1996). It is used to control for simultaneity. Simultaneity might occur, if a firm has private information and hence adjusts the factor demand to the change in its TFP.
- Examples of country-level studies include Colombia (Roberts and Tybout, 1997); Mexico, and Morocco (Clerides et al., 1998); France (Eaton et al., 2005); Germany (Bernard and Wagner, 2001); and Canada (Trefler, 2004).
- See Van Biesebroeck (2005), Aw et al. (2000) and De Loecker (2007).
- Mayer and Ottaviano (2007) provides evidence for several European countries while Bernard et al. (2007a) provide evidence for US firms.
- See the discussion in Sections C.3.c and C.4.b.
- We only have access to the top firms, which are selected according to the fulfilment of at least one of the following criteria: minimum Operating Revenue: €15m; minimum Total Assets: €30m; minimum number of employees: 200. This leaves us with a sample of 28621 firms.
- To clarify, this is not a comparison of import-competing firms, as this research is described in the next section. This section refers to the characteristics of US firms that purchase products from abroad, a segment of trading activity previously neglected in both empirical work and in most theoretical models of trade as well.
- If access to imported products leads to productivity gains to US firms (e.g., new technologies embodied in imported intermediate inputs), this may contribute one explanation to why exporting firms have higher productivity than non-exporting firms.
- The simplified presentation of the Melitz model in this subsection draws heavily on oral presentations of the model by Jeffrey Schott and Richard Baldwin at the WTO Secretariat.
- The number of varieties consumed in each nation falls if the fixed entry cost for imported varieties exceeds the fixed entry cost for local varieties.
- Competitiveness of the market implies, that each worker is paid her marginal product. Therefore the model generates an endogenous wage distribution.
- This follows the discussion of a fall in trade costs in Yeaple (2005).
- The skill premium refers to the extra income carried as a result of the level of skill embodied in jobs.
- The papers feature subtle differences as to how the relationship between the probability of exit and falling trade costs is modelled. However, these differences cannot be properly distinguished in empirical work, as acknowledged in Bernard et al. (2006b).
- To recall, in this model, average industry productivity increases as less productive firms exit the industry and not as a consequence of productivity increases in individual firms via technological change.
- See, for instance, Aw et al. (2000) for firms in Rep. of Korea and Chinese Taipei and De Loecker (2007) in regard to Slovenian firms. Van Biesebroeck (2005) obtains similar results for a panel of firms in Sub-Saharan Africa.
- Such an approach may be one way to explain the productivity improvements of lower productivity and import-competing firms following trade opening observed by Baggs et al. (2002) and Pavcnik (2002).
- To recall, both types of questions focus on the rationale for international trade: (i) If countries trade, what are the specific gains they expect; and (ii), what other factors determine the decision to trade.
- Essentially, trade opening increases market size. The observation that selection is tougher in larger markets has been made by Syverson (2008) who has focused on the effects of US market size (across regions) on the distribution of US establishments. He finds further support for the existence of larger average plant sizes in larger markets along with higher average plant productivity and lower average prices.
- Of course, for surviving firms, average firm size and total profits increase, as does product variety for consumers.
- Combes et al. (2007) show that Melitz and Ottaviano's (2008) results do not depend on the chosen parameterization.
- This "overcompensation" can only happen in the model if the technological laggard is at the same time the larger country.

- ⁸⁰ Unlike welfare effects of reductions in trade costs, which can be positive or negative for one of the countries depending on the technological gap, Demidova (2006) notes that in this model productivity improvements in one country invariably result in welfare losses in the other country. This is the case, as an improvement in a partner's productivity, other things equal, leads not only to a reduction in the production of the differentiated product, but also to a fall in consumed varieties, since the fall in consumption of domestic varieties is not fully compensated for by increases in consumption of foreign varieties. However, she also emphasizes that this result is only valid when both countries produce varieties of the differentiated product. If one country specializes in a different (here: homogenous) product, productivity improvements by its partner in producing the differentiated product increases its own welfare via improved terms of trade.
- ⁸¹ "Horizontal" refers to the same stage of processing, while "vertical" refers to different stages of processing, i.e. the production of intermediate inputs for further use in the production of final products.
- ⁸² The same is true the higher the fixed costs of exporting are in relation to the additional costs of investing in a foreign country.
- ⁸³ For early work on multinational enterprises stressing the proximity vs. concentration trade-off see Brainard (1997) and Markusen (1995). See also the discussion in Section D.1.
- ⁸⁴ Underlying is a Pareto-distribution, which is well-suited to describe phenomena like wealth or performance, where only few individuals achieve the highest values and a large mass is concentrated at the low end.
- ⁸⁵ A range of robustness tests is conducted employing a range of industry-specific control variables and estimation methods that, for instance, allow for the possibility of affiliates re-exporting a portion of their production abroad to third country, i.e. interdependence between a firm's decision to operate an affiliate in one country and its decision to locate affiliates in other countries.
- ⁸⁶ More precisely, a contract is likely to be incomplete since for reasons of asymmetric information and transaction costs, not all possible contingencies can be foreseen and included in the contract. Ex-post, i.e. once the contract is concluded (and "relationship-specific" investments have been made), both sides have an incentive to "defect" on certain promises they made and renegotiate. Such problems can be avoided if a supplier is integrated in a company and management can exercise vertical control over the supply chain.
- ⁸⁷ Refer to Section C.4.b for a discussion of this point.
- ⁸⁸ For an extended discussion of recent contributions based on the Romer model, see Gancia and Zilibotti (2005).
- ⁸⁹ For further discussion on the economic performance of East Asian countries, refer to Box 9 on "Success stories of export-led growth".
- ⁹⁰ For recent evidence see the Global Economic Prospect, World Bank (2008).
- ⁹¹ A number of other papers find that this result depends on the specific channel of technology transfer as FDI and licensing (Glass and Saggi, 1998; Lai, 1998; Yang and Maskus, 2001).
- ⁹² This is often measured as exports plus imports as a ratio of GDP, or simply as the ratio of exports to GDP.
- ⁹³ See Keller (2004) for an extensive review.
- ⁹⁴ E.g. Aw et al. (2000), Van Biesebroeck (2005), etc., discussed in Section C.3.a.
- ⁹⁵ E.g. Pavcnik (2002), Trefler (2004), etc., examined in Section C.3.c.

