NBER WORKING PAPER SERIES

THE STRUCTURE OF FOREIGN TRADE

Elhanan Helpman

Working Paper 6752
http://www.nber.org/papers/w6752

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 1998

This is a greatly extended and elaborated version of the Berhnard-Harms Prize lecture, delivered at the Kiel Institute for World Economics in June 1998. Many thanks go to Don Davis, Gene Grossman, Ed Leamer, Jim Levinsohn, Dan Trefler, and Manuel Trajtenberg for comments on an earlier draft, and to the NSF for financial support. The views expressed here are those of the author and do not reflect those of the National Bureau of Economic Research.

The Structure of Foreign Trade
Elhanan Helpman
NBER Working Paper No. 6752
October 1998
JEL No. F11, F12, F13

ABSTRACT

During the last two decades, new research has greatly advanced our understanding of the structure of world trade. This article reviews the empirical literature that grew out of this effort, emphasizing the ways in which it relied on theoretical developments.

Elhanan Helpman
Department of Economics
229 Littauer Center
Harvard University
Cambridge, MA 02138
and NBER
ehelpman@harvard.edu

The Structure of Foreign Trade*

Elhanan Helpman†

1 Introduction

World merchandise exports amounted to 5.3 trillion US dollars in 1997 and exports of commercial services amounted to 1.3 trillion. These are unprecedented volumes that have expanded much faster than income in the post-war period. More than half of the volume of merchandise trade flows amongst developed countries and less than 15% flows amongst developing countries. The rest, about one third, represents North-South trade – between developed and developing countries.

What explains these large volumes of trade? Why do some countries export computers while others export footwear? Can exports of airplanes be explained in the same way as exports of paper products? Questions of this type have been examined for many years. In attempting to answer them economists developed an elaborate analytical apparatus that has been greatly enriched in the last two decades. They have used the insights from trade theory to examine ever richer data sets in order to discover systematic patterns of trade flows and evaluate how well available theories match these data. Nevertheless, although we do have today better answers to some

---

*This is a greatly extended and elaborated version of the Bernhard-Harms Prize lecture, delivered at the Kiel Institute for World Economics in June 1998. Many thanks to Don Davis, Gene Grossman, Ed Leamer, Jim Levinsohn, Dan Trefler and Manuel Trajtenberg for comments on an earlier draft, and to the NSF for financial support.

†Harvard University, Tel Aviv University and the Canadian Institute for Advanced Research.

of these questions than our predecessors had forty years ago, the evolving structure of world trade defies simple explanations. I will try to describe in this lecture what we know about foreign trade and in what ways our understanding has improved as a result of the last twenty years of research.

2 Early Insights

David Ricardo's theory of comparative advantage, developed at the beginning of the 19th century, has played a major role in modern thinking about trade. Its emphasis on labor productivity renders it useful for analyzing a host of issues, such as the effects of technological progress on patterns of specialization and the distribution of gains from trade. But as far as the structure of foreign trade is concerned, and in particular in relation to data analysis, this theory assumes too much exogeneity to be useful. For this reason empirical studies of trade flows that build only on Ricardo's insights are hardly available. A few of them appeared in the fifties and sixties.[1] As ingenious as they were, it quickly became apparent that they are of limited use. A key reason for this conclusion is, of course, that Ricardian trade models assume that only labor is used to produce goods and services, with fixed labor-output coefficients. Treating these coefficients as given it then proceeds to predict that a country exports products in which it has a comparative advantage; namely, it exports products for which, in comparison to other countries, its labor coefficients are relatively low.[2] Or, put differently, it exports products with relatively high labor productivity.

Although a sensible prediction, the theory is mute on a key ingredient: what causes labor productivity to differ across countries?[3] One major difficulty encountered

---

[1] See McDougall (1951,1952) and Stern (1962).

[2] Let $a^{k}(i)$ be the required labor input per unit output of product i in country k. These are the fixed coefficients. And let w^{k} represent the wage rate in country k, measured in a common currency for all countries (say US dollars). Then if country k exports product i to country l, unit costs must be lower in country k due to competition. Therefore $w^{k}a^{k}(i) \leq w^{l}a^{l}(i)$. Similarly, if country l exports product j to country k unit costs are lower in country l; $w^{k}a^{k}(j) \geq w^{l}a^{l}(j)$. It follows that the exporter's relative labor coefficient is lower than the importer's; $a^{k}(i)/a^{k}(j) \leq a^{l}(i)/a^{l}(j)$.

[3] Eaton and Kortum (1997) provide the first study of a Ricardian model that contains an explana-

by empirical research was the fact that differences in the use of capital provide an important source of variation in labor productivity. Capital-rich countries are able to allocate more capital per worker to all economic activities than capital-poor countries, but they may do so to a different degree in various lines of business. As a result output per worker (or output per hour) may vary across industries to a different degree in capital-rich and capital-poor countries.[4] This raises the question: what determines the allocation of capital to industries and thereby labor productivity? But once the role of capital is taken seriously it may be best to abandon the focus on labor productivity altogether and think about what determines trade flows amongst countries that have more than labor as an input.

Eli Heckscher and Bertil Ohlin provided a framework for thinking about trade in just this type of situation. Both Heckscher (1919) and Ohlin (1924) emphasized the roles of labor, capital and land in agriculture and industry, trying to explain how their availability shapes a country's pattern of specialization and trade. For this purpose technologies are assumed to be the same in all countries. Despite the richness of the approach taken by the founders of the Heckscher-Ohlin theory, Paul Samuelson and his followers elaborated a two-factor two-sector version that became the cornerstone of modern trade theory.[5] Samuelson's version is crisp and elegant. Its focus on labor and capital on the one hand and on exporting and import-competing sectors on the other cuts to the heart of much that matters. For this reason it was widely adopted as the workhorse of the profession.

According to the two-factor, two-sector version of the Heckscher-Ohlin theory, a country should export the product that is relatively intensive in the factor with which

---

tion of labor productivity based on a country's technology level. Using extreme value distributions for labor productivity they derive an equation for bilateral trade flows. They estimate this equation for a sample of OECD countries, using cumulative investment in R&D and the number of scientists and engineers as proxies for technology levels. The results are very good.

[4] This line of reasoning has been formally explored by Ford (1982). It makes the Ricardian and Hekscher-Ohlin theories observationally equivalent in some circumstances.

[5] See Samuelson (1948), Jones (1956-57) and Jones (1965). The argument in the text notwithstanding, Samuelson (1953-54) also provided a very general treatment of the Heckscher-Ohlin trade model.

the country is relatively well-endowed. Thinking about labor and capital as the two inputs, it means that a capital-rich country; i.e., a country that has more capital per worker than its trade partners, should export the capital-intensive product. The argument can be made in two parts. First, if factor prices are not equalized, then the rental rate on capital relative to the wage rate is lower in the capital-rich country. As a result it uses in all product lines more capital per worker than the capital-poor country. But, as shown by Lerner (1952), under these circumstances the capital-rich country has a cost advantage in the capital-intensive product, which it ends up exporting. This implies that its exports are more capital intensive than its imports. That is, if we were to calculate how much capital and labor are embodied in the country's exports and how much capital and labor are embodied in its imports, we should find that for the capital-rich country the ratio of capital to labor is larger in exports than in imports. Second, if factor prices are equalized, then the capital-rich and capital-poor countries use the same ratios of capital to labor to produce identical products. But because the capital-rich country has a disproportionately large amount of capital relative to labor it ends up producing a disproportionately large amount of capital-intensive products. Otherwise it cannot maintain full employment of labor and capital. It then follows that with a similar composition of demand (that results, for example, from identical homothetic preferences in all countries), the capital-rich country exports capital-intensive products in this case too. So we find again that in the capital-rich country the ratio of capital to labor embodied in exports is larger than the ratio of capital to labor embodied in imports.

Leontief (1954) put this prediction to a test. Having worked on US input-output tables, he calculated labor-output and capital-output ratios for a variety of sectors. Matching the sectoral composition of the input-output tables with trade data he then calculated how much labor and capital are embodied in exports and how much in imports. For the latter he used the assumption that the US input-output coefficients also apply to imports, which is the case when foreign suppliers use the same techniques of production as domestic producers. When countries have access to the same technologies and factor prices are equalized this assumption is justified. Otherwise it should be treated as an approximation.

Let a_{Li} be labor input per unit output of product i and let a_{Ki} be capital input per unit output. These coefficients are not constant. With flexible technology they are chosen by manufacturers to minimize unit costs and therefore depend on factor rewards. But taking their values from data it follows that the labor content of a vector of output quantities $\mathbf{x}$ that has a typical element x_i is given by $\sum_i x_i a_{Li}$ while the capital content is given by $\sum_i x_i a_{Ki}$. Using these expressions we find that the capital-labor ratio embodied in $\mathbf{x}$ equals $\sum_i x_i a_{Ki} / \sum_i x_i a_{Li}$. Leontief calculated this ratio twice, once for a vector $\mathbf{x}$ equal to US imports and once for a vector equal to US exports in 1947.[6] Surprisingly, he found that the capital-labor ratio embodied in imports exceeded the ratio embodied in exports. The surprise emanated from the fact that after the war the US was considered to be the most capital-rich country in the world and the Heckscher-Ohlin theory predicts for such a country a higher capital intensity of exports rather than imports. His finding became known as the "Leontief Paradox". Not only was there a paradox in these data, it was also a big one; the capital–labor ratio in imports exceeded the capital-labor ratio in exports by 60%. A large deviation indeed.[7]

Leontief proposed a possible explanation to the paradox: US workers are more productive than foreign workers. If a US worker produces two times as much as a foreign worker the paradox is resolved. But why would US workers be so much more productive, especially after controlling for capital employment? In a sense this explanation resolves one paradox by introducing another.

3 Further Developments

The two-factor two-sector version of the Heckscher-Ohlin theory was extended in the sixties and seventies.[8] Many of these efforts examined the relationship between goods prices and factor rewards on the one hand and the relationship between endowments

---

[6] He did it, in fact, for fractions of these vectors, each one valued at one million dollars.

[7] In later data sets the paradox is less pronounced, but it does not disappear.

[8] See Ethier (1984) for a review of the literature.

and outputs on the other. They have no direct bearing on the main topic of this lecture and will therefore not be discussed. What is central for this lecture, however, is the fact that these extensions provided a surprisingly simple theoretical specification that could be used for empirical research. Despite the fact that the variables used in these specifications had clear empirical counterparts, some of the required data were not readily available. Ed Leamer has done more than anybody else in promoting a research program centered on the construction of such data sets.

Two types of relationships provide the underpinning for the generalized Hecksher-Ohlin theory. First comes production. Let A be a technology matrix with a typical element a_{ij} that represents the quantity of input i used in the manufacturing of one unit of output j. I ignore intermediate inputs. Therefore this matrix describes coefficients for primary inputs only, such as labor, capital and land. It allows distinguishing between various types of capital, such as machines and structures; various types of labor, such as high school dropouts and college graduates; and various types of land, such as pasture and arable land. Other inputs can also be incorporated. A column of this matrix describes the technique of production in a particular sector. Cost-minimizing manufacturers choose these coefficients from the available technology taking factor prices as given. As a result the matrix A in a particular country depends on its technology and factor rewards. In the simplest version the technology is taken to be the same everywhere and factor prices are assumed to be equalized[9]. Under the circumstances the same matrix is used in all countries. Next let $\mathbf{V}^k$ represent the vector of inputs in country k. This is a column vector.[10] Similarly, let $\mathbf{X}^k$ represent the output vector in country k, also a column vector. Then full employment of resources implies

$$A\mathbf{X}^k = \mathbf{V}^k \quad \text{for all } k. \tag{1}$$

The second relationship comes from consumption. Preferences are assumed to be

---

[9] Factor prices are not the same in all countries, not even in the OECD countries. But as O'Rourke and Williamson (1998) demonstrate, trade and migration are powerful forces that lead to convergence of factor prices.

[10] Remember that a row in the matrix A describes the quantities of a particular input used in various industries per unit output.

homothetic and the same in all countries (a strong assumption). In the absence of impediments to trade this relationship implies that the composition of consumption is the same everywhere. Consumption is therefore proportional to aggregate output $\bar{\mathbf{X}} = \sum_k \mathbf{X}^k$. We denote by s^k the share of country k in consumption. It then follows that consumption is given by

$$\mathbf{C}^k = s^k \bar{\mathbf{X}} \quad \text{for all } k. \tag{2}$$

When trade is balanced the share s^k equals country k's share in world income. Otherwise an adjustment is need for trade imbalance.

Equations (1) and (2) provide two fundamental relationships that have been used to formulate empirical specifications. Suppose that inputs and outputs are aggregated into an equal number of categories. In this event the matrix A is square and can be inverted (whenever it is not singular). It then follows from (1) that output is given by $\mathbf{X}^k = A^{-1}\mathbf{V}^k$. Net exports equal the difference between output and consumption: $\mathbf{T}^k = \mathbf{X}^k - \mathbf{C}^k$. Therefore using (2) and $\mathbf{X}^k = A^{-1}\mathbf{V}^k$ we obtain

$$\left(\frac{1}{s^k}\mathbf{T}^k\right) = A^{-1}\left(\frac{1}{s^k}\mathbf{V}^k\right) - \bar{\mathbf{X}} \quad \text{for all } k. \tag{3}$$

This is a linear set of relationships between net exports, normalized by relative expenditure, and factor endowments, also normalized by relative expenditure. Each such relationship, for petroleum products, forest products, machinery or chemical products, can be estimated from cross-country data. We need data on net exports, which are readily available, and factor endowments, which are not readily available in comparable form. Leamer (1984) estimated this linear relationship using a newly constructed data set. This was the first contribution to a new line of empirical research. He did not insist on the equality of the number of inputs and outputs. Nevertheless, the simple linear specification performs very well on a cross section of sixty countries in his data set, for both 1958 and 1975. As expected, availability of oil raises net exports of petroleum products, but it also reduces net exports of machinery in 1975. And abundance of literate, non-professional workers raises exports of labor-intensive manufactures, such as apparel and footwear. These are just examples of the type of effects that were estimated.[11,12]

Estimates of the type provided by Leamer are interesting. But they do not provide a test of the generalized Heckscher-Ohlin theory. The reason is that the theory predicts a relationship between endowments and trade mediated by technology. To test it therefore requires independent information about all three objects: technology, endowments, trade.

Two concepts of trade have surfaced so far: trade in goods and services and the factor content of trade flows. The former is a natural focus of trade theory. But, as explained earlier on, Leontief converted trade in products into factor content. His procedure can be generalized by constructing net exports of factor content, as suggested by Vanek (1968). With identical technology matrixes A in all countries this procedure is quite simple. The factor content of a vector of output $\mathbf{x}$ is $A\mathbf{x}$. Therefore the factor content of net exports of country k, equal to the factor content of its exports minus the factor content of its imports, can be written as $\mathbf{F}^k \equiv A\mathbf{T}^k$. Using the definition $\mathbf{T}^k = \mathbf{X}^k - \mathbf{C}^k$ and the production and consumption relationships (1) and (2) we obtain Vanek's key equation

$$\mathbf{F}^k = \mathbf{V}^k - s^k\bar{\mathbf{V}} \text{ for all } k, \tag{4}$$

where $\bar{\mathbf{V}} = \sum_k \mathbf{V}^k$ is the aggregate endowment vector in the world economy. The left-hand-side represents the factor content of net exports, the right had side provides a measure of factor abundance. When inputs are ordered according to increasing abundance in country k, so that $V_1^k/\bar{V}_1 < V_2^k/\bar{V}_2 < \cdots < V_m^k/\bar{V}_m$, then the vector on the right-hand-side has negative elements in the low-index coordinates, predicting that these inputs are imported on net ($F_i^k < 0$), and positive elements in the high-index coordinates, predicting that they are exported on net ($F_i^k > 0$). So, on net, inputs that are relatively abundant are exported while those that are relatively scarce

---

[11] Linearity is a maintained assumption in Leamer's estimation, and it works well for most sectors. In some sectors, however, such as chemicals, the data favor a non-linear specification.

[12] In addition to estimates of the link between trade and endowments, recent studies have examined the link between production and endowments; see Harrigan (1995) and Reevs (1998). More on this below.

are imported.

Vanek's equation was used by Leamer (1980) to point out a shortcoming of Leontief's procedure: whenever there are more than two inputs a comparison of the ratio of the embodied quantities of two of them in imports and exports does not provide the relevant metric for rejecting the theory.[13]

Vanek's equation can be transformed in a way that makes it useful for empirical research. Bowen, Leamer and Sveikauskas (1987) used the following version of (4), obtained by dividing the equation for input i by $s^k \bar{V}_i$:

$$\frac{\left(F_i^k / s^k\right)}{\bar{V}_i} = \frac{\left(V_i^k / s^k\right)}{\bar{V}_i} - 1 \quad \text{for all } i \text{ and } k.$$

In this version we have on the left-hand-side the factor i content of net exports of country k, adjusted for the spending share of the country, relative to the ***world's*** endowment of this input. This is a natural normalization of the factor content of net exports. On the right-hand-side we have the country's endowment of this input, also adjusted for the spending share, relative to the world's endowment of the input, minus one. The right-hand-side therefore measures how the country's size-adjusted endowment deviates from the world's average. If this measure is above the world's average the input is exported on net. If it is below the world's average it is imported on net.

Two types of tests were performed on this equation, both based on calculations of the right-hand and left-hand-side. These calculations are the first to build on independent information about endowments, technology and trade. One test compares the rank orders of the expressions on the right-hand and left-hand-sides, and examines how well they match. A second test compares the signs of the corresponding expressions on the right-hand and left-hand-sides. Both tests point to difficulties. Bowen, Leamer and Sveikauskas find violations of the sign test in about one third of the cases and violations of the rank order test in about half the cases. This appears

---

[13] This may sound like a neat resolution of the Leontief Paradox. It is not. As pointed out by Brecher and Chaudhri (1982), the fact that the US was a net exporter of labor services has implications for consumption per worker, which are not born out by the data.

to be bad news for the expanded Heckscher-Ohlin theory. But how bad the news is is hard to gauge, because the theory is not tested against a well-specified alternative. For this reason it is also possible to take a more positive attitude and to argue that, since these tests were done on data for 12 inputs and 27 countries (for 1967), the theory explains a reasonably large fraction of the variation, across factors and countries, of the factor content of net trade flows[14].

4 Recent Advances

Although Bowen, Leamer and Sveikauskas (1987) pointed out difficulties with the Vanek equation, they did not investigate whether there are systematic deviations of the data from the theoretical predictions. This important task was undertaken by Trefler (1995). He compiled a new data set, for 33 countries, that disaggregates endowments into nine inputs. The countries in the sample account for three-quarters of world exports and nearly 80% of world income in 1983. In Trefler's data the correlation between F_i^k and $V_i^k - s^k\bar{V}_i$ (taking account of the variation across inputs and countries) equals 0.28, while Vanek's equation predicts a correlation of one.[15] A sign test of the Bowen-Leamer-Sveikauskas type was successful in only about one-half of the cases while Vanek's equation predicts a 100% match. It therefore appears that Trefler's data fits the expanded Heckscher-Ohlin theory as well or as badly as Bowen, Leamer and Sveikauskas' data does.

Trefler shows clearly in what ways these data deviate from the theoretical predictions:

---

[14] Bowen, Leamer and Sveikauskas (1987) also examined a set of alternative hypotheses, including neutral technological differences among countries. See below.

[15] All the reported calculations use normalized data, which is obtained from the original data by dividing every input by the standard deviation across countries of the difference

$$F_i^k - (V_i^k - s^k\bar{V}_i)$$

and by the measure of country size $(s^k)^{1/2}$. Here relative country size is measured by its share in aggregate GDP.

- First, the measures of the factor content of net exports F_i^k are compressed towards zero relative to the factor abundance measures $V_i^k - s^k \bar{V}_i$. That is, even in cases in which both variables have the same sign the former is much smaller in absolute value than the latter. This compression is so striking as to imply hardly any net trade in factor content.[16]

- Second, when sorted by GDP per capita adjusted for purchasing-power-parity (PPP) poor countries have systematically more negative values of $F_i^k - \left(V_i^k - s^k \bar{V}_i\right)$ than rich countries.[17] This means that whenever a poor country exports an input on net, it exports less than predicted by its factor abundance measure. And whenever it imports an input on net, it imports more than predicted by its factor abundance measure. For rich countries the opposite is true.

- Third, poor countries tend to be abundant in more factors than rich countries; they have more inputs for which $V_i^k > s^k \bar{V}_i$.[18]

This characterization is of lasting value. It gives us a better understanding of the ways in which the data do not match the theory. And it provides a clear theoretical challenge: How should the model be modified to better fit the data?

Following Bowen, Leamer and Sveikauskas (1987), Trefler (1995) examined how much improvement in fit (of data to theory) can be obtained from considering simple differences in productivity across inputs and countries. Suppose, as suggested by Leontief, that inputs are not equally productive in all countries. If US labor is, for example, two times as productive as labor in Italy, then a thousand hours of US labor services are equivalent to two thousand hours of labor services in Italy. Taking one country as the benchmark, we can then convert the endowment V_i^k of country k into $\pi_i^k V_i^k$ equivalent units of the benchmark country, where π_i^k is the productivity of input i in country k relative to the same input in the benchmark country. Using the

---

[16] It is not unusual in these data for the absolute value of the factor abundance measure to be 10 to 50 times higher than the absolute value of the factor content of trade.

[17] The correlation between $F_i^k - \left(V_i^k - s^k \bar{V}_i\right)$ and income per capita is 0.87.

[18] The correlation between $V_i^k - s^k \bar{V}_i$ and income per capita is -0.89.

US as the benchmark country then implies that $\pi_i^k = 1/2$ for labor in Italy. Under these circumstances Vanek's equation (4) becomes

$$F_i^k = \pi_i^k V_i^k - s^k \sum_l \pi_i^l V_i^l \quad \text{for all } k \text{ and } i, \tag{5}$$

where the vector of factor content of net exports $\mathbf{F}^k$ is calculated using the benchmark country's technology. Evidently, with no further restrictions the productivity coefficients π_i^k can be calculated exactly from these equations; there are no degrees of freedom left for estimation.

Trefler (1993) used (5) to calculate labor and capital productivity coefficients for a set of countries relative to the US, and he then compared them with the wage rate and the return to capital in these countries relative to the US. According to the theory, if there is factor price equalization, then in a cross-country comparison of a factor reward the relative rewards should equal the relative productivity parameters. Indeed he found the relative factor rewards to be highly correlated with the calculated relative productivity parameters.

Trefler (1995) did not use (5) to calculate the productivity parameters. Instead he used two alternatives to obtain degrees of freedom that allow estimation. First, suppose that the productivity parameters π_i^k are only country-specific. That is, if US labor is two times as productive as Italian labor so is US capital and land. This restriction attributes a single parameter π^k to every country, which then represents its common productivity advantage in all lines of business. This represents a Hicks-neutral variation. Trefler's second observation – namely, that poor countries appear to export too little factor content while rich countries export too much – suggests that such productivity differences can help explain these data. Second, suppose that countries are divided into two sets: one set – consisting of developed countries – has the same productivity as the benchmark country, while countries in the other set – which are less developed – share the same factor-specific productivity parameters π_i.[19] Trefler examined a nested specification of these models. Against the null of the simple Vanek equation he examined the Hicks-neutral specification and a combination

---

[19] Together with the productivity parameters Trefler also estimated the composition of these sets.

of the Hicks-neutral and factor-augmented productivity differences between two sets of countries. Using a model-selection criterion he concluded that the Hicks-neutral specification performs best.[20]

To better account for the "missing trade", i.e., the compression of the F_i^k's towards zero, Trefler also introduced a home bias in demand. He found that this bias together with the Hicks-neutral productivity differences provides the best explanation of the data.

I find the use of a home bias in demand less appealing in explaining the data than the use of technological differences across countries. There is plenty of independent evidence that technologies differ across countries.[21] There is no such evidence for demand patterns, except for biases that are related to income levels. For these reasons a home-bias parameter is too crude a tool to handle this problem.

Davis, Weinstein, Bradford and Shimpo (1997) (DWBS for short) examined separately the production relationship (1) and the consumption relationship (2). They evaluated the production relationship in a cross section of countries and a cross section of Japanese regions, using Japan's input-output table. By computing the left-hand-side and the right-hand-side of (1) they performed rank order tests of the type introduced by Bowen, Leamer and Sveikauskas (1987). They found that the equations do not describe well the international data but are remarkably accurate for the Japanese regions. One possible interpretation of this finding is that techniques of production are very similar across Japanese regions but differ significantly across countries. The latter is consistent with other pieces of evidence. The interesting finding is therefore that very similar techniques are used all over Japan, which is most likely when the technology available to Japanese manufacturers is the same in every region and there is factor price equalization. On the other hand, failures in the international data can emanate either from lack of factor price equalization (which we know to be true from data on wages) or from differences in technological opportunities (which we also know to be the case). At the moment there are no estimates of how much of the

---

[20] Bowen, Leamer and Sveikauskas (1987) also examined productivity differences. Unfortunately, due to a programming error their results are not correct.

[21] See Harrigan (1997) for a good recent example.

failure is attributable to lack of factor price equalization and how much to differences in technology.

DWBS examined equation (2) for Japanese regions. In one version, which is weaker than (2), they tested whether regional absorption (consumption plus investment) vectors are proportional to aggregate Japanese absorption, concluding that they are. Next they tested whether Japanese aggregate absorption is proportional to world production, concluding that it is.[22] It follows that if the Japanese data present difficulties for Vanek's equation it is not due to lack of proportionality in absorption (homotheticity). And in view of the good fit of the production relationship such difficulties also cannot result from cross-regional differences in techniques of production. When testing Vanek's equation (4) directly, however, DWBS found that it does not fit the Japanese data well. As in Trefler's data set, this case too exhibits a "missing trade " phenomenon. Namely, the computed factor content of net exports is much smaller than what is predicted by factor abundance. For example, while net exports of noncollege labor services amount to 3.6 percent of this factor endowment, the factor abundance expression $V_i^k - s^k \bar{V}_i$ predicts imports of 31 percent of the endowment – a large deviation indeed. So how can these different findings be reconciled? Why do DWBS' production and consumption equations fit the Japanese data while Vanek's equation does not?

DWBS propose a partial resolution to this puzzle. Note that (4) is strictly correct only when all countries use the same techniques of production. If, however, Japan uses techniques of production that differ from other countries, then, using its technology matrix, we obtain $\mathbf{F}^J \equiv A^J \mathbf{T}^J = A^J \mathbf{X}^J - A^J \mathbf{C}^J = \mathbf{V}^J - s^J A^J \bar{\mathbf{X}}$, where J stands for Japan. The next step towards (4) requires to replace $A^J \bar{\mathbf{X}}$ with $\bar{\mathbf{V}}$, but it is invalid unless all countries use the Japanese matrix A^J. Avoiding this last step DWBS test the less restrictive equation

$$\mathbf{F}^J = \mathbf{V}^J - s^J A^J \bar{\mathbf{X}}.$$

As expected, this equation fits the Japanese data much better than Vanek's equation,

---

[22] For this test they assumed that the intersectoral flows of intermediate inputs are the same in Japan as in the rest of the world.

substantially reducing the puzzle of the "missing trade".[23]

Evidently, we now know much more about the gaps between theory and data. Trefler uncovered patterns of deviation. Technological differences between countries help to explain them. The work by Davis, Weinstein, Bradford and Shimpo supports this emphasis. But we still do not have a good handle on why techniques of production differ between countries and whether it is possible to explain the data with systematic differences in techniques of production that are driven by a small number of country-specific characteristics. Clearly, every pattern of factor content can be explained with arbitrary differences in techniques of production. This is not useful. The challenge is to find key features of countries that result in differences in techniques of production that fit the data well.

To give an example of what type of country differences matter – which appears to be consistent with the just reported analysis of the Japanese data by DWBS – consider a world of two countries, two inputs and two outputs, in which differences in factor composition are wide and therefore there is no factor price equalization. Let the factors be capital and labor, let the products be food and clothing, and let clothing be capital intensive. The technologies are the same in both countries. Country A has a high capital–labor ratio and therefore a low rental rate on capital and a high wage rate. As a result it employs more capital and less labor than country B would employ in the same line of business. But also suppose that due to the extreme differences in factor endowments country A manufactures only clothing and B produces only food.

Country A's technology matrix is A^A while B's is A^B. And we have $a^A_{Ki} > a^B_{Ki}$ for $i = f, c$, where K stands for capital, f for food and c for clothing. In addition $a^A_{Li} < a^B_{Li}$ for $i = f, c$, where L stands for labor.[24]

---

[23] There is, however, an inconsistency in this argument. If all counties use Japan's technology matrix, then we are justified to replace $A^J\mathbf{X}$ with $\bar{\mathbf{V}}$, and therefore the original Vanek equation should work well. If, on the other hand, other countries do not use Japan's technology matrix, then $\mathbf{F}^J \equiv A^J\mathbf{T}^J$ does not represent the factor content of net trade flows. Namely, in this case the DWBS formulation is correct as a matter of accounting, but $\mathbf{F}^J$ is not a particularly meaningful economic object.

[24] The assumption of complete specialization implies that in a country's technology matrix the

A's net export vector is $\mathbf{T}^A = \begin{pmatrix} -s^A X_f^B \\ s^B X_c^A \end{pmatrix}$, because it imports a fraction s^A of B's output of food and it exports to B a fraction s^B of its output of clothing. Therefore, using A's technology matrix to calculate the factor content of net exports, we obtain

$$\mathbf{F}^A \equiv A^A \mathbf{T}^A = \begin{pmatrix} -s^A a_{Kf}^A X_f^B + s^B a_{Kc}^A X_c^A \\ -s^A a_{Lf}^A X_f^B + s^B a_{Lc}^A X_c^A \end{pmatrix}.$$

From the factor market clearing conditions $a_{Zc}^A X_c^A = Z^A$ and $a_{Zf}^B X^B = Z^B$, where $Z = K, L$, it now follows that

$$\mathbf{F}^A = \begin{pmatrix} -s^A \frac{a_{Kf}^A}{a_{Kf}^B} K^B + s^B K^A \\ -s^A \frac{a_{Lf}^A}{a_{Lf}^B} L^B + s^B L^A \end{pmatrix}.$$

Next note that

$$\mathbf{V}^A - s^A \bar{\mathbf{V}} = \begin{pmatrix} -s^A K^B + s^B K^A \\ -s^A L^B + s^B L^A \end{pmatrix},$$

because $s^A + s^B = 1$. Therefore $F_K^A < V_K^A - s^A \bar{V}_K$ and $F_L^A > V_L^A - s^A \bar{V}_L$, because $a_{Kf}^A / a_{Kf}^B > 1$ and $a_{Lf}^A / a_{Lf}^B < 1$. It follows that the capital-rich country exports on net less capital services than predicted by factor abundance and imports on net less labor services than predicted by factor abundance. A similar conclusion is obtained from an analysis of the labor-rich country.[25] Evidently, lack of factor price equalization induces the use of techniques of production that bias the outcome towards "missing trade". To see why, note that using the matrix of country A introduces no bias in the calculation of the factor content of A's ***exports***. But it biases the calculation of the factor content of A's ***imports***, because foreign manufacturers use different input-output coefficients – they use less capital and more labor per unit output. Therefore the use of A's matrix overstates the amount of capital and understates the amount of labor embodied in its imports. As a result ***net*** exports of capital services are

---

coefficients of factor input per unit output for the product that it does not manufacture represent what it would have used were it to manufacture the product.

[25] These conclusions are independent of which country's technology matrix is used to calculate the factor content of net exports.

understated and so are ***net*** imports of labor services. This is why differences in techniques of production help to explain the phenomenon of "missing trade".[26] But how much "missing trade" can be explained by cross-country variations in factor rewards? At the moment there is no answer to this question.

Gabaix (1997) uses regression analysis to examine a variant of (4). His results are mostly negative. They underscore the extent of the "missing trade".

For a group of countries with proportional consumption vectors $\mathbf{C}^k = s^k\mathbf{C}^R/s^R$, where $\mathbf{C}^R$ is the consumption vector of the entire group and s^R is its share of world consumption. Using this relationship the procedure that leads to (4) can be used to derive instead

$$\frac{1}{s^k}\mathbf{F}^k = \frac{1}{s^k}\mathbf{V}^k - \frac{1}{s^R}A\mathbf{C}^R. \tag{6}$$

To test this relationship for a particular input i Gabaix writes

$$\frac{1}{s^k}F_i^k = \beta_i\frac{1}{s^k}V_i^k - \frac{1}{s^R}A_i\mathbf{C}^R, \tag{7}$$

where A_i is the i'th row of matrix A, and β_i is an auxiliary parameter that can be estimated by regressing the left-hand on the right-hand-side for a cross section of countries.[27] The null hypothesis is that $\beta_i = 1$. Using Trefler's data set Gabaix found that this hypothesis is rejected: β_i is close to zero for the entire sample of 33 countries, for a smaller sample consisting only of the OECD countries, and also for an even smaller sample restricted to the rich European countries. The small size of the estimated β coefficients is just a manifestation of the "missing trade". What is striking, however, is how small they are. For labor, for example, the coefficients are: zero for the entire sample, −0.017 for the OECD countries, and 0.047 for the rich European countries.

Next Gabaix examined whether cross-country differences in factor-augmenting productivity levels help to explain these data. Instead of using Trefler's (1993) procedure of calculating the productivity differences under the assumption that the model is correct and then comparing them to data on factor rewards, however, he tested directly how the model performs when one assumes that factor rewards are proportional

---

[26] A similar point is made by Hakura (1997).

[27] The second term on the right hand side is a constant, the same for all countries.

to factor-augmenting productivity differences. With such productivity differences, that lead to (5), equation (6) becomes

$$\frac{1}{s^k} F_i^k = \pi_i^k \frac{1}{s^k} V_i^k - \frac{1}{s^R} A_i \mathbf{C}^R,$$

where π_i^k is the productivity of input i in country k relative to the benchmark country, which he took to be the US. With factor price equalization π_i^k equals the reward to factor i in country k relative to the US. Therefore one can estimate a modified version of (7)

$$\frac{1}{s^k} F_i^k = \beta_i w_i^k \frac{1}{s^k} V_i^k - \frac{1}{s^R} A_i \mathbf{C}^R,$$

where w_i^k is the reward to factor i in country k relative to the US. The null that $\beta_i = 1$ is again rejected. And these β coefficients remain extraordinarily small. For labor they are: 0.020 for the entire sample, 0.007 for the OECD countries, and -0.027 for the rich European countries. This type of productivity correction does not appear to greatly improve the model's fit.[28]

These results suggest that the difficulties with the basic model may be more severe than previously appreciated, at least as far as the role of factor-augmenting technological differences is concerned. Note, however, that Gabaix's specification does not account for differences in techniques of production that may arise from lack of factor price equalization. It would therefore be interesting to see whether the interactions of factor-augmenting technological differences with variations in techniques of production that result from lack of factor price equalization significantly improve the model's fit.[29]

---

[28] As Gabaix points out, cross-country differences in factor rewards are mostly driven by differences in output levels per unit input. And since the values of F_i^k are very small in these data, calculations (such as Trefler's (1993)) that assume $F_i^k = 0$ for all i and k in (5) give almost identical productivity levels as calculations that use the true values of the F_i^k's. It follows that the factor content of net exports has little value in explaining such productivity differences or differences in factor rewards.

[29] Gabaix's estimation procedure can be improved by exploiting information about $\frac{1}{s^R} A_i \mathbf{C}^R$. When he estimates the coefficient β_i for a particular input he also estimates $\frac{1}{s^R} A_i \mathbf{C}^R$, the constant. But there is nothing to guarantee that this constant equals the true value of $\frac{1}{s^R} A_i \mathbf{C}^R$. An alternative would be to use data to calculate $\frac{1}{s^R} A_i \mathbf{C}^R$ and then regress $\frac{1}{s^k} F_i^k + \frac{1}{s^R} A_i \mathbf{C}^R$ on $w_i^k \frac{1}{s^k} V_i^k$ without

Things do not look so bleak when attention is focused on the pattern of specialization rather than trade. Reeve (1998) estimated a relationship between outputs and inputs for a sample of 20 OECD countries, using data on 15 sectors and five types of inputs: capital, three grades of labor and arable land.[30] He found that cross-country variations in factor endowments explain over 40% of the variation in output levels. Importantly, when allowing for cross-country differences in Hicks-neutral productivity levels (π_i^k is country specific but does not vary across sectors within a country) the fit improves significantly and an additional 7% of the variation in output levels is explained. And decomposing output changes from 1970 to 1985, Reeve finds that shifts in factor endowments and the techniques of production explain over 80% of the changes in the sectoral output levels. But significantly, changing factor endowments contributes about twice as much as changing techniques of production.

All this suggests that as imperfect as the theory is, some of its components fare well empirically. My conclusion from the evidence is that we need to model more carefully the cross-country differences in techniques of production that are driven by both technological differences and differences in factor rewards, in order to close the gap between the theory and the data. This view is strengthened by Hakura's (1997) recent findings. She shows that in data for five of the original EU countries the fraction of observations that pass the Bowen-Leamer-Sveikauskas sign test of Vanek's equation rises markedly when the equation is modified to admit cross-country differences in techniques of production. Like Bowen, Leamer and Sveikauskas (1987) she finds that (4) has the correct sign pattern in about one half of the observations. But this fraction rises to 70-80 percent when each country's techniques of production are used instead of a common technology matrix. And this fraction rises even further when allowance is made for non-traded intermediate inputs.[31] Evidently, allowing for differences in

---

an intercept, and then test whether the slope equals one. Although I believe that this is more appropriate, I also believe that it is unlikely to change the results dramatically.

[30] To justify this procedure with more sectors than inputs, one must assume some extraneous forces, such as transport costs, that limit the possible patterns of production. Otherwise a continuum of production patterns could be consistent with the same factor endowments, as is evident from (1).

[31] See Table 6 in Hakura (1997). In 1980 94% of the observations have the correct sign pattern

techniques of production dramatically improves the fit of factor content equations. Now we need to take one further step forward and identify the forces that induce countries to choose different techniques of production.

5 Economies of Scale and Product Differentiation

Concurrent with the refinement of the Heckscher-Ohlin trade theory and the development of its empirical implications for the factor content of net trade flows, a new theory that emphasizes economies of scale and product differentiation emerged in the 1980s. At the early stages of its development the new theory seemed to threaten the Heckscher-Ohlin orthodoxy.[32] But soon it became clear that explanations of trade provided by economies of scale and product differentiations complement the explanations provided by factor endowments. This called for an integrative view of foreign trade that would allow for an interplay between economies of scale, product differentiation, and factor proportions. Helpman and Krugman (1985) developed such an approach, making allowance for sectors that differ in their sources of scale economies, in production styles, and in market structure.

Much of this research was originally motivated by the observation that large volumes of trade flow between countries with similar factor proportions and that significant trade overlap exists within industries. These facts have not changed. In 1996 the 15 countries of the European Union exported a little over two trillion dollars-worth of merchandise and imported a similar amount. About 65% of this trade was within the EU. At the same time their imports from Japan exceeded their imports from ***all*** of Africa and were more than twice as high as their imports form ***all*** of the Middle East.[33] Evidently, the EU countries trade with countries that are similar to themselves more than they trade with the very different economies of Africa and the Middle East. More broadly, the industrial countries trade with each other much

---

when Italy is excluded from the sample.

[32] See Helpman (1984) for a review of the literature on trade with economies of scale and Krugman (1995) for a review of the literature on trade with monopolistic competition.

[33] See Table A10 in WTO (1997).

more than they trade with less-developed countries.

Measures of trade overlap within industries have remained high.[34] To take a couple of examples, the share of intra-industry trade in the UK was 53.2% in 1970, it increased to 74.4% in 1980 and to 84.6% in 1990. In Germany it increased from 55.8% in 1970 to 56.6% in 1980 and to 72.2% in 1990.[35,36] And these two countries represent a general trend of rising shares of intra-industry trade.

Helpman and Krugman (1995) have shown that economies of scale, product differentiation, and various forms of conduct are compatible with factor price equalization and, as a consequence, with Vanek-type equations for the factor content of trade. For this reason the empirical evidence that I reviewed so far is also relevant for the richer theory developed in the eighties.[37]

Although factor price equalization and the employment of identical techniques of production in all countries can take place with economies of scale, their presence makes this less likely. Scale economies drive countries to specialize in different products, which enhances the motives for foreign trade. For this reason they help to explain large trade volumes between similar countries. At the same time economies of scale make it more likely that countries will employ different techniques of production. This is especially so when there are dynamic economies of scale, be they driven by learning-by-doing or investment in research and development. Both benefit companies by giving them access to technologies that are not available to rivals.[38] Since the empirical evidence does not appear to be consistent with the use of identical

---

[34] These measures consider imports and exports within each industry, evaluate how much overlap there is in every sector, and provide a summary statistic of the share of total trade that consists of such overlaps. According to the Heckscher-Ohlin theory such overlaps should not exist. Davis (1995) has proposed, however, that small differences in individual product-related technologies can produce such overlaps even in a Heckscher-Ohlin framework.

[35] See Table 1.8 in OECD (1996).

[36] Declines in this index are rare, but it happened in Norway.

[37] Grossman and Helpman (1991) show conditions under which they also remain valid in economies that invest in R&D. Such economies enjoy more or improved products over time.

[38] Such advantages exist at least temporarily, as is evident from a variety of technological races in the electronics and pharmaceutical industries. See Grossman and Helpman (1995) for a review of the literature on the links between trade and technology.

techniques of production worldwide, or even within groups of relatively homogeneous countries, the study of country-specific technological developments becomes all the more important for the understanding of international trade.

5.1 Intra-Industry Trade

Grubel and Lloyd (1975) provided the first comprehensive study of the extent of trade overlap. They devised an index to measure this overlap as a fraction of total trade. The Grubel-Lloyd index for intra-industry trade between countries k and l can be expressed as

$$G^{kl} = \frac{2\sum_j \min\left(IM_j^{kl}, IM_j^{lk}\right)}{\sum_j \left(IM_j^{kl} + IM_j^{lk}\right)},$$

where IM_j^{kl} represents the value of imports of sector-j output by country k from country l. By construction this index is between zero and one with higher values representing more trade overlap. Grubel and Lloyd showed that this index was high for many countries. And before the theory of intra-industry trade was properly developed Loertscher and Wolter (1980) established some stylized facts about partial correlations between country and industry characteristics on the one hand and the extent of trade overlap on the other.[39] They found in particular that the share of intra-industry trade is high when:

- the trading partners are highly developed and they are at a similar level of development;
- the trading partners are large and they do not differ too much in size.

To explain large shares of intra-industry trade, sectors with product differentiation were introduced into the theory. There are many products that are differentiated by brand: some are simple, such as breakfast cereal, tooth paste or clothing, while others are sophisticated, such as cars, computers or MRI machines. Some are consumer

---

[39] They did not use the Grubel-Lloyd index to measure the extent of trade overlap. Nevertheless, their index measures the same phenomenon in a somewhat different way.

goods. Many, however, are producer goods, including capital goods such as drilling machines and intermediate inputs such as microprocessors. The theory applies to all of them. It starts by noting that product differentiation typically involves economies of scale. A brand has to be developed, such as a lightweight laptop. Or it has to be designed, such as a dress fashion. In either case, once the nature of the product has been established, manufacturing costs determine the profitability of its production. The company that developed such a product gains some monopoly power, because the market does not provide a perfect substitute for its unique brand. Moreover, companies are driven to differentiate their creations from the brands of rivals.

Economies of scale limit the range of products that are profitably supported by the market; the smaller the economies of scale the more brands become available. With international trade, countries specialize in different brands. When every country demands a wide spectrum of varieties, international trade leads naturally to trade overlap: brand-specific economies of scale lead to intra-industry trade. Although empirical researchers (including Loertscher and Wolter) have sought an association between the degree of economies of scale in a sector and the extent of its intra-industry trade, the theory does not appear to imply such a relationship. What matters is that there are economies of scale, not their size.

Using this theory Helpman (1987) formulated an empirical equation for a cross section of countries that focused on the link between the share of intra-industry trade in bilateral trade flows and a small number of key variables that describe the characteristics of the trade partners. According to the theory the share of intra-industry trade is larger between countries that are similar both in composition of factor endowments and in size. To measure the former he used the absolute difference in GDP per capita. For the latter he used the GDP level of the larger country as one variable and the GDP of the smaller one as a separate variable. For a sample of 14 industrial countries he estimated this relationship and found that the partial correlations had the predicted signs for most of the years between 1970 and 1981. That is, the extent of trade overlap was larger the more similar were the countries' income per capita, the smaller was the larger country and the larger was the smaller country. But these relationships weakened over time.[40]

These findings, however, do not appear to be robust. Hummels and Levinsohn (1995) confirmed them for alternative measures of similarity in factor composition. Replacing income per capita with income per worker, for example, had little effect on the results. And by using the absolute differences in capital-labor and land-labor ratios as measures of factor similarity, they confirmed that countries with more dissimilar endowment compositions engage in less intra-industry trade. But when they employed panel techniques to estimate these relationships (fixed and random effects) the results changed dramatically. Now it appeared that most of the variation in the share of intra-industry trade could be explained by country-pair dummies. That is, unspecified characteristics of country pairs explain more than the variables emphasized by the theory. Why this should be so is not understood at this point. But this finding raises an obvious need to broaden the theory in order to arrive at a better empirical specification.

5.2 Volume of Trade

Specialization encourages international trade and economies of scale and product differentiation strengthen the tendency to specialize. As a result economies of scale and product differentiation lead to large volumes of trade.

To see how trade volumes are determined by specialization, consider an extreme case in which every country is completely specialized in a subset of products. Assuming that all have the same homothetic preferences then implies that country k imports from l the fraction s^k of l's output. The volume of trade (imports plus exports) between these two countries therefore equals

$$Q^{k,l} = s^k Y^l + s^l Y^k,$$

where Y^k is the GDP level of country k. If, in addition, spending levels are propor-

---

[40] Helpman constructed Grubel-Lloyd indexes of intra-industry trade from 4-digit industry data for the purpose of this study. At this level of disaggregation the extent of trade overlap is less likely to be a statistical artifact than at lower levels of aggregation.

tional to GDP levels, then

$$Q^{k,l} = \frac{2}{Y} Y^k Y^l, \tag{8}$$

where Y is the world's GDP.

This is a gravity equation. It implies that in a cross section of countries the logarithm of the bilateral volume of trade rises one for one with the logarithm of each country's GDP. Equations of this form – aggregated in various ways – have been estimated time and again, providing a good fit to many data sets.[41] Helpman (1987) used it to calculate the volume of trade amongst 14 industrial countries as a fraction of their aggregate income. According to the theory, this fraction is larger the more similar are the countries' income levels as measured by the similarity index $1 - \sum_k \left(s^k\right)^2$. He found that between 1956 and 1981 this index increased and the fraction of income traded within the group increased as well. This comovement is consistent with models of product differentiation in which specialization in production is driven by brand proliferation.

Hummels and Levinsohn (1995) reexamined Helpman's evidence using (8) for ***bilateral*** trade flows. They confirmed Helpman's finding for a group of industrial countries. But they also applied the same equation to a group of mixed countries – some developed, others less developed – arguing that if product differentiation is the main reason for the good fit, the equation should not perform well in the mixed sample. Although the equation did not fit the mixed sample as well as it did the homogeneous sample, it was remarkably good nevertheless. From this they concluded that the evidence does not lend support to the view that product differentiation is the key to the gravity equation.

Evenett and Keller (1998) developed an estimation procedure that sheds more direct light on the link between product differentiation and the gravity equation. Recall that product differentiation leads to intra-industry trade in addition to specialization. Therefore, they argued, if the gravity equation is driven largely by brand prolifera-

---

[41] Gravity equations go back to Tinbergen (1962). Linnemann (1966) provided a comprehensive analysis. Typical specifications control for the distance between the trade partners, which has a negative effect on the volume of trade.

tion, it should perform better for pairs of countries with large shares of intra-industry trade. Moreover, for country pairs with no intra-industry trade, where factor proportion differences are the main stimulus to trade, they derived gravity equations that depend on whether the factor proportion differences are large enough to lead to specialization in production or not. The modified equations differ in the predicted coefficients, which depend in this case on the shares of various goods in production. As a result they can be subjected to empirical tests.

Using a mixed sample of countries, similar to Hummels and Levinsohn, Evenett and Keller divided the observations of country pairs into those that have less than five percent of intra-industry trade according to the Grubel-Lloyd index and those that have more. They treated the former as country pairs whose trade is dominated by factor proportions. The latter pairs they further divide into five classes that differed in the degree of intra-industry trade. They then estimated, using a resampling procedure, a version of the gravity equation, conditioning its coefficients – whose values theoretically should be one – on the group to which the countries belong in terms of their share of intra-industry trade. The results were that the coefficient was closer to its theoretical value the larger was the share of intra-industry trade. Furthermore, using a variant of the gravity equation that allows for a mixture of homogenous and differentiated product sectors enabled them to estimate varying coefficients across the five classes. They found again that the estimated coefficient was closer to the theoretical value the larger the share of intra-industry trade. A model-discrimination criterion favored the latter model, which allows for both homogenous and differentiated products, over the pure product-differentiation case.

For the sample of country pairs with less than five percent of intra-industry trade Evenett and Keller examined two alternative specifications, assuming in both that these countries trade homogeneous products. In one specification, differences in factor proportions are large enough to lead to specialization and to the gravity equation (8). The other case exhibits overlap in production and a modified gravity equation applies. To compare these two models they divided the country pairs into five classes according to differences in factor composition. Using again a resampling procedure they estimated the coefficients of the pure gravity equation within each class. Contrary to the theoretical prediction the estimated coefficients were smaller rather than larger the more the countries differed in factor composition and they were all much smaller than the predicted value of one. On the other hand, for the model with production overlap the correlation between the estimated coefficient and the difference in factor composition was as predicted. A model-discrimination criterion favored the specification with production overlap. Apparently the evidence does not support claims that the gravity equation is mostly driven by specialization due to marked differences in factor composition.[42]

This evidence lends support to the view that economies of scale with product differentiation are valuable components in the explanation of trade flows. It still leaves room for factor endowments to play a role. But it adds an important layer to the determinants of trade flows.

5.3 The Broad Patterns of Trade

An additional insight is gained by considering the broad patterns of trade that were mentioned in the introduction and discussed at the beginning of this section, namely, the fact that developed economies trade mostly with each other rather than with less-developed countries, and that trade within the group of less-developed countries is only a small fraction of total world trade (about 15%). To see how product differentiation helps to account for these facts, consider a world that consists of n^N symmetrical developed countries (the North) and n^L symmetrical less-developed countries (the South). This means that all countries in the North have the same factor endowments as do countries in the South. But the composition of factor endowments differs be-

---

[42] Deardorff (1998) made the argument that a gravity equation can be derived from a generalized Heckscher-Ohlin model with large differences in factor composition that lead to specialization in production. Although this is a theoretical possibility, the Evenett-Keller evidence suggests that it has little empirical content. Deardorff's alternative derivation of the gravity equation for economies with homogeneous products, which is based on random divisions of imports across the exporting countries, is not even appealing on theoretical grounds. I agree with Grossman's (1998) position on this point.

tween the North and the South. In some sectors there is product differentiation, in others products are homogeneous.

Now suppose that due to differences in factor composition the North specializes in differentiated products while the South specializes in homogeneous products. Since all the developing countries are the same, they do not trade with each other.[43] On the other hand, each country in the North imports a constant fraction of output from every other Northern country. Therefore the volume of trade within the North is $s^N Y^N (n^N - 1) n^N$, where s^N is the share of spending of a typical Northern country – which I take to equal its share in world income – and Y^N is the GDP level of a typical Northern country. Intra-Northern trade therefore equals

$$Q^{NN} = \left(s^N\right)^2 (n^N - 1) n^N Y,$$

where Y is again world income.

Imports of a Northern country from the South equal a fraction s^N of the South's output while each Southern country imports from the North a fraction s^S of Northern output, where s^S is the share of a Southern country in world income and spending. Therefore the volume of trade between North and South equals

$$Q^{NS} = s^N n^N n^S Y^S + s^S n^S n^N Y^N = 2 s^N s^S n^N n^S Y$$

and $Q^{NN}/Q^{NS} = s^N (n^N - 1)/2 s^S n^S$. Since $s^N n^N + s^S n^S = 1$, it follows that the volume of trade within the North relative to the volume of North-South trade is given by

$$\frac{Q^{NN}}{Q^{NS}} = \frac{s^N (n^N - 1)}{2\left(1 - s^N n^N\right)}.$$

For reasonable values of the number of countries and their relative size this ratio is larger than one, indicating that trade within the North exceeds trade between North and South. For example, suppose that there are 20 countries in the North, each one

---

[43] This is of course an extreme condition. Some internal trade would exist in the South if, for example, factor proportions were not the same in every country. But this extreme formulation helps to make the main point. Generally there will be less intra-group trade in the South the more the South specializes in homogeneous products and the less its countries differ in factor compositions.

accounting for 4% of world output. The North accounts therefore for 80% of the world's output and it is four times as large as the South. Under the circumstances intra-Northern trade is almost twice as large as North-South trade.[44]

Davis (1997) argued that the much larger volume of North-North than North-South trade can also be explained with a Heckscher-Ohlin model. He was able to produce this outcome by incorporating differences in techniques of production across goods and sectors as well as suitable cross-country differences in factor endowments. I find this to be an interesting intellectual exercise, but of little practical value. To demonstrate that something can happen does not mean that it is likely to happen. For this reason frameworks in which the phenomenon happens naturally are more appealing on theoretical grounds. Judged by this criterion the broad structure of world trade is more naturally explained with the aid of product differentiation than without it. To be sure, factor proportions matter. But explaining the observed trade pattern with only homogeneous products requires too much fine tuning of the technology to be convincing.

A broader point should be noted here. Some researches have argued that the modern theory of international trade with economies of scale and product differentiation is in fact not needed in order to explain key features of world trade – such as the existence of intra-industry trade, the good fit of the gravity equation, or the difference between intra-North and North-South trade volumes – and that for each feature we can find a structure of a traditional Heckscher-Ohlin model that also produces it. Researchers in this tradition strive to convince us that there is no need to incorporate product differentiation into trade theory. This is, of course, a legitimate debate. It is also intellectually interesting to see how far one can push a model or a line of argument. What I find puzzling, however, is the occasionally expressed preference for models without product differentiation on empirical grounds. The reason I find this attitude puzzling is that product differentiation is so prevalent that it is hard to see why it is even necessary to justify its presence in economic models. If anything, it should be necessary to justify the use of models with homogeneous products

---

[44] It is somewhat lower in the data, only about 1.6 times as large.

only. Just think about the products we consume: food, clothing, furniture, home appliances, sports equipment, cars – almost anything – they all are differentiated products. The same can be said about intermediate inputs and capital goods, as well as about a host of services, such as banking, insurance, or travel. For this reason it seems only natural to think about differentiated rather than homogeneous products. As far as I can see homogeneous products are used in economic theory mostly as a simplifying assumption. Therefore whenever product differentiation helps to explain certain phenomena it is only natural to embrace this assumption rather than to insist on the construction of models with homogenous products that also explain them.

All this being said, the ultimate test is in the direct and indirect evidence. There is plenty of direct evidence, from observation and otherwise, that product differentiation is prevalent. But this does not imply necessarily that the available models of international trade with product differentiation provide a good explanation of the data. Their empirical evaluation depends very much on the indirect evidence: How well do their implications fit the facts? At the moment none of the available models does a great job explaining the data. But adding product differentiation improves the fit between theory and data. Since the inherent richness of the models with product differentiation has not yet been much explored, they also carry the potential of providing even better explanations when subjected to further analysis.

5.4 Economies of Scale

Modern trade theory places significant weight on economies of scale and product differentiation in explaining the structure of foreign trade. Its usefulness has been gauged by a variety of implications that help to interpret some stylized facts, and by the moderate fit of some of its correlates in various data sets. As useful as the research along these lines has been, however, it provides no direct evidence on the extent of economies of scale. In fact, most of the implications that have been examined do not depend on the ***degree*** of economies of scale, just on their existence. Although this is good enough for many purposes, for a variety of welfare-related questions it is quite important to have a sense of how large the economies of scale are. Some micro production studies find economies of scale. But such studies, which are confined to single-country data sets, cannot provide estimates of economies of scale at the sectoral level and may, therefore, underestimate the size of scale effects.

Recently Antweiler and Trefler (1997) developed a methodology for estimating returns to scale at the sectoral level from international data. They constructed a data set for 71 countries, with 37 industries and 11 factors, spanning a twenty-year period from 1972 to 1992. These data provide rich variations that are most suitable for this purpose.

The key theoretical observation that enables them to estimate the degree of economies of scale is that a variant of the Vanek equation (4) holds even when countries use different techniques of production. In such cases, however, it is necessary to calculate separately the factor content of exports and the factor content of imports for each and every trade partner. Let $\mathbf{A}_i^k$ (a vector) be the i^{th} column of the technology matrix A^k in country k, let E_i^k be exports of product i from country k, and let M_i^{kl} represent country k's imports of product i from country l. Then the factor content of country k's net exports is

$$\mathbf{F}^k = \sum_i E_i^k \mathbf{A}_i^k - \sum_i M_i^{kl} \mathbf{A}_i^l.$$

With this definition (4) holds as long as every country has the same composition of consumption.[45]

Next note that the differences in the employed techniques of production can be specified as emanating from differences in factor productivity levels π_{ji}^k, as in Trefler (1993), or from economies of scale. As a result of scale economies the use of inputs per unit output declines (on average) as output expands. Assuming that the productivity parameters π_{ji}^k are proportional to factor rewards leaves scale economies as the only remaining source of cross-country variations in productivity levels. By parametrizing the effects of scale economies one can therefore estimate the suitable coefficients from such modified Vanek equations.[46]

---

[45] Recall that (4) was originally derived under the assumption that A^k is the same in all countries.

[46] The scale function was parametrized as $\left(X_i^k\right)^{-\alpha_{ji}^k}$ for input j in sector i of country k, with

Evidently, this procedure attributes all cross-country variations in productivity levels to differences in scale (except for the π_{ji}^ks). Scale is indeed a key identifying restriction in this study. And the results are very revealing. Productivity is higher the higher the level of output. Costs fall on average by about 0.15 to 0.20 percent when output rises by one percent. These estimates are larger than those obtained from micro studies. But the scale coefficients are not only significantly larger than zero, they also have low standard errors. Allowing for non-homotheticity in production the estimates imply that output expansion leads to shifts in the techniques of production that raise the demand for skilled relatively to unskilled workers.

Harrigan (1998) uses estimates of production functions for a sample of OECD countries to asses whether economies of scale at the sectoral level or Hicks-neutral differences in technological efficiency of countries better explain cross country variations in sectoral levels of total factor productivity. The results are mixed. Differences in total factor productivity are large, and economies of scale do not explain them adequately. Neutral technological differences provide a somewhat better fit. Since one does not exclude the other, it would be interesting to allow for both, scale economies and technological differences – as in Antweiler and Trefler (1997) – in order to find out what fraction of the variation in total factor productivity is explained by each one. Harrigan's data does not allow, however, to accurately estimate this decomposition.

Another inquiry into economies of scale, based on international data, is provided by Davis and Weinstein (1998). Their point of departure is Krugman's (1980) home-market effect. Krugman introduced transport costs for varieties of a differentiated product that are produced with economies of scale. In a simple two-country model with one input (labor) and two sectors, one differentiated the other homogeneous, he showed that the larger country exports differentiated products on net. In this sort of environment, where production functions are the same in both countries, the pattern of specialization is undetermined in the absence of transport costs. But with

---

$\alpha_{ji}^k \geq 0$. This term multiplies a benchmark input-output ratio a_{ji} that is common to all countries, except for the productivity adjustment π_{ji}^k. Namely, the input–output coefficient for country k is $a_{ji}^k = a_{ji}\left(X_i^k\right)^{-\alpha_{ji}^k}/\pi_{ji}^k$. The larger the αs the stronger the economies of scale. The αs were estimated using various restrictions, such as equality across inputs (homotheticity) and sectors.

transport costs the size of local demand determines the profitability of manufacturing differentiated products. The larger country, with its greater demand, supports the development of a disproportionate number of brands and therefore exports them on net. The important role of the size effect is here in demand, not supply.[47]

From the exposition of Krugman's model in Helpman and Krugman (1985, section 10.4) one arrives at the following equation that relates a country's share of output of differentiated products to its share in total spending on differentiated products:

$$\eta_S^k = -\frac{\rho}{1-\rho} + \frac{1+\rho}{1-\rho}\eta_D^k,$$

where η_S^k is country k's share in the supply of differentiated products, η_D^k is its share in spending on differentiated products, and ρ is a parameter between zero and one that is larger the lower are transport costs.[48] Evidently, the country's share of supply rises more than one for one with its share in spending on differentiated products.[49]

More generally, in economies with such features, shifts in the aggregate demand for a country's products have a disproportionately large effect on its output of products with economies of scale. In pure Heckscher-Ohlin-type economies with no transport costs such effects are nil while in Heckscher-Ohlin-type economies with transport costs they are less than proportional. It is therefore possible to discriminate between these alternative models by estimating the effects of aggregate demand on the output of various products.

To estimate the demand for goods in a particular sector of country k, Davis and Weinstein first estimate industry-specific gravity equations for bilateral trade

---

[47] As long as specialization in production is incomplete a labor supply increase in a given country, without a corresponding increase in demand for its differentiated product, is absorbed in the production of its homogenous product.

[48] Helpman and Krugman (1995) derived this equation under the assumption that the fraction of income spent on differentiated products is the same in both countries. Under these circumstances a country's share in demand equals its share in the world's labor force. But allowing for different spending patterns is easy. When country k spends a fraction μ^k of its income on differentiated products its share in total spending on differentiated products equals $\mu^k L^k / \left(\mu^k L^k + \mu^l L^l\right)$, where L^k is its labor force.

[49] The coefficient $\frac{1+\rho}{1-\rho}$ is larger than one.

flows, accounting for distances between countries. Their sample consists of 22 OECD countries and 25 sectors. As is typical for such equations, the trade volume declines with distance. Using these estimates they then construct relative demand levels for every industry and estimate the effect of this demand for country k products, that emanates from all countries including k, on the local supply. Pooling over industries they find that a one percent increase in demand raises local output by 1.6 percent, which is significantly larger than one. This provides evidence in favor of scale effects. Performing the same estimation for every industry separately they obtain less accurate estimates. Nevertheless, in 11 out of the 25 sectors the coefficients are significantly larger than one, confirming the presence of economies of scale.[50]

6 Conclusions

We now have a rich theory of international trade that emphasizes economies of scale, product differentiation and differences in factor composition as key determinants of the structure of world trade. In combination they explain significant parts of specialization patterns, volumes of trade, factor content of trade, and the broad patterns of trade across regions. But despite the massive research effort in the last twenty years, these explanations are still incomplete. This is partly the result of the fact that the nature of world trade is changing at a fast pace. Technological change has modified the patterns of specialization, has reduced trading costs and encouraged larger trade volumes; new countries have joined the trading system, and multinational corporations have spread their net more than ever before. The new economy is marked by plenty of man-made comparative advantage, which is occasionally short-lived, as are many product cycles. All this means that we need a more technologically oriented trade theory and more emphasis on dynamics in order to understand these developments. Although such theories have been elaborated in the 1990s, they have so far had little effect on empirical research. This is the area in which success will

---

[50] Davis and Weinstein (1997b) estimated a similar model on regional data from Japan. There they also found scale effects in a number of sectors.

pay off the most.

References

[1] Antweiler, W. and D. Trefler, "Increasing Returns and All That: A View from Trade," mimeo, 1997.

[2] Bowen, H.P., E.E. Leamer and L. Sveikauskas , "Multicountry, Multi-factor Tests of the Factor Abundance Theory", *American Economic Review*, 1987.

[3] Brecher, R.A. and E.V. Chaudhri, "The Leontief Paradox Continued ", *Journal of Political Economy*, 1982.

[4] Davis, D.R., "Intra-Industry Trade: A Heckscher-Ohlin-Ricardo Approach", *Journal of International Economics*, 1995.

[5] Davis, D.R., "Critical Evidence on Comparative Advantage? North-North Trade in a Multilateral World," *Journal of Political Economy*, 1997.

[6] Davis, D.R., D.E. Weinstein, "Economic Geography and Regional Production Structure: An Empirical Investigation," mimeo, 1997 (forthcoming in the *European Economic Review*).

[7] Davis, D.R., D.E. Weinstein, "Market Access, Economic Geography and Comparative Advantage: An Empirical Assessment," mime, 1998b.

[8] Davis, D.R., D.E. Weinstein, S.C. Bradford and K. Shimpo, "Using International and Japanese Regional Data to Determine When the Factor Abundance Theory of Trade Works," *American Economic Review*, 1997.

[9] Deardorff, A.V., "Determinants of Bilateral Trade: Does Gravity Work in a Neoclassical World?" in J.A. Frankel (ed.), *The Regionalization of the World Economy* (Chicago: The University of Chicago Press), 1998.

[10] Eaton, J. and S. Kortum, "Technology and Bilateral Trade," NBER Working Paper No. 6253, 1997.

[11] Ethier, W.J., "Higher Dimensional Issues in Trade Theory", in Jones R.W. and P. Kenen (eds.), *Handbook of International Economics* (Amsterdam: North Holland), 1984.

[12] Evenett, S.J. and Keller, W., "On Theories Explaining the Success of the Gravity Equation," mimeo, 1998.

[13] Ford, J.L., "The Ricardian and Heckscher-Ohlin Explanations of Trade: A General Proof of an Equivalence Theorem and its Implications", *Oxford Economic Papers*, 1982.

[14] Gabaix, X., "The Factor Content of Trade: A Rejection of the Heckscher-Ohlin-Leontief Hypothesis," mimeo, 1997.

[15] Grossman, G.M., "Comment," in J.A. Frankel (ed.), *The Regionalization of the World Economy* (Chicago: The University of Chicago Press), 1998.

[16] Grossman, G.M. and E. Helpman, *Innovation and Growth in the Global Economy* (Cambridge, MA: The MIT Press), 1991.

[17] Grossman, G.M. and E. Helpman, "Technology and Trade," in Grossman, G.M. and K. Rogoff (eds.), *Handbook of International Economics*, Vol. 3 (Amsterdam: North Holland), 1995.

[18] Grubel, H. and P. Lloyd, *Intra-Industry Trade: The Theory and Measurement of International Trade in Differentiated Products* (London: Macmillan), 1975.

[19] Hakura, D., "A Test of the General Validity of the Heckscher-Ohlin Theorem for Trade in the European Community," mimeo, 1997.

[20] Harrigan, J., "Factor Endowments and the International Location of Production: Econometric Evidence for the OECD, 1970-1985," *Journal of International Economics*, 1995.

[21] Harrigan, J., "Technology, Factor Supplies, and International Specialization: Estimating the Neoclassical Model," *American Economic Review*, 1997.

[22] Harrigan, J., "Estimation of Cross-Country Differences in Industry Production Functions," *Journal of International Economics*, 1998 (forthcoming).

[23] Heckscher, E.F., "The Effect of Foreign Trade on the Distribution of Income", *Economisk Tidskrift*, 1919, in Heckscher, E.F. and B. Ohlin, *Heckscher-Ohlin Trade Theory* (Cambridge, MA: The MIT Press), 1991 (translated, edited and introduced by Harry Flam and M. June Flanders).

[24] Helpman, E., "Increasing Returns, Imperfect Markets, and Trade Theory," in Jones, R.W. and P. Kenen (eds.), *Handbook of International Economics*, Vol. I (Amsterdam: North Holland), 1994.

[25] Helpman, E., "Imperfect Competition and International Trade: Evidence from Fourteen Industrial Countries", *Journal of the Japanese and International Economies*, 1987.

[26] Helpman, E. and P.R. Krugman, *Market Structure and Foreign Trade* (Cambridge, MA: The MIT Press), 1995.

[27] Hummels, D. and J. Levinsohn, "Monopolistic Competition and International Trade: Reconsidering the Evidence", *Quarterly Journal of Economics*, 1995.

[28] Jones, R.W., "Factor Proportions and the Heckscher-Ohlin Theorem", *Review of Economic Studies*, 1956-57 (reprinted in his *Essays,* 79).

[29] Jones, R.W., "The Structure of Simple General Equilibrium Models", *Journal of Political Economy*, 1965.

[30] Krugman, P.R., "Scale Economies, Product Differentiation, and the Pattern of Trade," *American Economic Review*, 1980.

[31] Krugman, P.R., "Increasing Returns, Imperfect Competition and the Positive Theory of International Trade," in Grossman, G. and K. Rogoff (eds.), *Handbook of International Economics*, Vol. 3 (Amsterdam: North Holland), 1995.

[32] Leamer, E.E., "The Leontief Paradox , Reconsidered", *Journal of Political Economy*, 1980.

[33] Leamer, E.E., *Sources of International Comparative Advantage*, (Cambridge, MA, and London, The MIT Press), 1984.

[34] Leamer, E.E. and J. Levinsohn, "International Trade Theory: The Evidence", in Grossman, G. and K. Rogoff (eds.), *Handbook of International Economics*, Vol. 3 (Amsterdam: North Holland), 1995.

[35] Leontief, W.W., "Domestic Production and Foreign Trade: The American Capital Position Re-examined", *Economica Internazionale*, 1954.

[36] Lerner, A., "Factor Prices and International Trade", *Economica*, 1952.

[37] Linnemann, H., *An Econometric Study of International Trade Flows* (Amsterdam: North Holland), 1966.

[38] Loertscher, R. and F. Wolter, "Determinants of Intra-Industry Trade: Among Countries and Across Countries", *Weltwirtschaftliches Archiv*, 1980.

[39] McDougall, G.D.A., "British and American Exports: A Study Suggested by the Theory of Comparative Costs", Part I, *Economic Journal*, 1951.

[40] McDougall,G.D.A., "British and American Exports: A Study Suggested by the Theory of Comparative Costs", Part II, *Economic Journal*, 1952.

[41] OECD, *Globalization of Industry*, Paris, 1996.

[42] Ohlin, B., "The Theory of Trade", 1924, in Heckscher, E.F. and B. Ohlin, *Heckscher-Ohlin Trade Theory* (Cambridge, MA: The MIT Press), 1991 (translated, edited and introduced by Harry Flam and M. June Flanders).

[43] O'Rourke, K. and J.G. Williamson, *Globalization and History: The Evolution of a 19th Century Atlantic Economy* (Cambridge, MA: The MIT Press), 1998 (forthcoming).

[44] Reevs, T.A., *Essays in International Trade*, Ph.D. dissertation, Harvard University, 1998.

[45] Samuelson, P.A., "International Trade and the Equalization of Factor Prices", *Economic Journal*, 1948.

[46] Samuelson, P.A., "Prices of Factors and Goods in General Equilibrium", *Review of Economic Studies*, 1953-4.

[47] Stern, R.M., "British and American Productivity and Comparative Costs in International Trade", *Oxford Economic Papers*, 1962.

[48] Tinbergen, J., *Shaping the World Economy: Suggestions for an International Economic Policy*, New York, 1962.

[49] Trefler, D., "International Factor Price Differences: Leontief was Right!", *Journal of Political Economy*, 1993.

[50] Trefler, D., "The Case of the Missing Trade and Other Mysteries", *American Economic Review*, 1995.

[51] Vanek, J., "Factor Proportions Theory: The N-factor Case", *Kyklos*, 1968.

[52] WTO, *Annual Report 1997*, Vol. II, France, 1997.