Davis, Donald R.; Weinstein, David E.

Article

What Role for Empirics in International Trade?

Aussenwirtschaft

Provided in Cooperation with:
University of St.Gallen, School of Economics and Political Science, Swiss Institute for International Economics and Applied Economics Research

Suggested Citation: Davis, Donald R.; Weinstein, David E. (2002) : What Role for Empirics in International Trade?, Aussenwirtschaft, ISSN 0004-8216, Universität St.Gallen, Schweizerisches Institut für Aussenwirtschaft und Angewandte Wirtschaftsforschung (SIAW-HSG), St.Gallen, Vol. 57, Iss. 4, pp. 441-468

What Role for Empirics in International Trade?

Donald R. Davis and David E. Weinstein*
Columbia University

Das vorrangige Ziel der Aussenwirtschaftstheorie sollte ein profundes Verständnis der Determinanten der in der realen Welt anzutreffenden Handelsströme ein. Die empirische Analyse aktueller Handelsbeziehungen spielt dennoch nur eine untergeordnete Rolle. Obwohl Fortschritte im Bereich der neoklassischen Theorie, der steigenden Skalenerträge, des unvollkommenen Wettbewerbs, der Wachstumstheorie und zuletzt auf dem Gebiet der «Economic Geography» das Verständnis internationaler Handelsbeziehungen enorm vergrössert haben, blieben ähnliche Erfolge auf der empirischen Seite bisher aus. Mit Blick auf die empirischen Arbeiten im Bereich technologischer Unterschiede, unterschiedlicher Faktorausstattungen und v.a. der Heckscher-Ohlin-Vanek (HOV) Theorie gelangt der Artikel zu dem Schluss, dass die wichtigsten empirischen Fragen der Aussenhandelstheorie immer noch offen sind. Allerdings wird die Heckscher-Ohlin Theorie auch in Zukunft ein wichtiger Bestandteil jedes empirischen Erklärungsversuches aktueller Handelsmuster sein.

Keywords: Bertil Ohlin, trade empirics, Heckscher-Ohlin-Vanek model, factor price equalization, intra-industry trade
JEL-Codes: B23, B40, F10, F14

1 Introduction

The centennial of BERTIL OHLIN's birth is an outstanding opportunity to reflect on his work. One must acknowledge that much of what has come to be known as the Heckscher-Ohlin theory of trade has been mediated by the contributions of others, including some contributors to this volume. But none would gainsay that, even for a specialist in the field, there is great delight and inspiration to be found in reading the original texts. Deep economic intuition and breadth of vision grace each page. It would not be far off the mark to observe that a great deal of the theoretical work in positive trade in the last half century – including some of the most recent – has involved elaboration of ideas for which OHLIN already provided interesting treatments.

---

* Prepared for the Conference on the Centennial of BERTIL OHLIN's Birth, Stockholm, October 1999. The article appears by kind permission of MIT Press as a reprint from the book "Bertil Ohlin: A Centennial Celebration (1899–1999)", edited by RONALD FINDLAY, LARS JONUNG and MATS LUNDAHL, chapter 17.

The centennial of OHLIN's birth also provides stimulus to take the long view of our own field of international trade. Others will be discussing theoretical developments. Our charge is to consider empirical developments, addressing the role of theory only as it has helped to shape this enterprise. The existence of outstanding recent surveys of empirical trade, by some of the world's leading empirical and theoretical researchers (cf. LEAMER 1992; HELPMAN 1999), allows us to forego any attempt at a comprehensive survey. Instead, we will ask how our field has approached empirical research, how data findings have interacted with the development of theory, and how we can strengthen the useful interaction of the two.

Our paper draws a few conclusions. Theory has been the heart of international trade research for the past half century. And a glorious half century of theorizing it has been! Yet this research program has been extraordinarily imbalanced. Moreover, we believe that this imbalance is a serious problem for progress in the field as a whole.

We take the primary objective of our field to be an understanding of the determinants of trade patterns in the world we actually inhabit. Yet empirical analysis of actual trade relations plays a diminutive role in the field. Our field shows little of the two-way interplay between theory and data that is the very life of many fields of economics, such as macro, labor, and others. We believe it is possible to maintain what is beautiful and distinctive about our own field while enriching it in this dimension.

One response to this may be: write interesting data analysis and we will read it! LEAMER and LEVINSOHN (1995) took this perspective as an implicit starting point, looking inward to understand why empirical research has failed to materially affect the views of most trade economists. Such introspection for empirical researchers is both important and necessary.

Yet we believe that such inward looking by empirical researchers addresses only one part of our field's problem. We believe that there needs to be a substantial change in the way that theorists think about data analysis. Theoreticians need to move beyond only working with a few stylized facts to a broader encounter with the empirical work. If there is anything that has been learned on the empirical side of trade in the last decade, it should surely be that spectacular failures of our theories, anomalies, and inconvenient facts are our most precious resources. Failure points the way to success – but only if we learn to embrace and understand facts that are inconvenient for our theories.

Our summary judgment is that, at a deep level, the field has a quite limited empirical understanding of international trade patterns. We can say little about the relative importance of distinct fundamental determinants of trade. Some correlates of trade and production patterns have been established. Some work has been done on international patterns of absorption. But such efforts remain in their infancy. Grappling with the deeper problem of how the pieces really fit together within a world general equilibrium has barely begun.

One might read this and view us as relentless pessimists. This would be a mistake. We believe that skepticism about the state of our knowledge is a very healthy stance for researchers. But, lest we be misunderstood, let us add a few caveats. We are great admirers of what the field has achieved in theory over this half-century and more. To those who have written the beautiful and elegant models that constitute the very language or our field – and you know who you are – we send our cheers! Likewise, the empirical side of trade has a number of researchers – you also know who you are – who have pioneered methods that provide the foundation on which others will build.

Moreover, we see signs of hope, both in the interests of younger researchers and in the reception these have encountered among the leaders of our field. To give only one example, a simple survey, such as that by HELPMAN (1999), can be extremely important in focusing the profession's attention both on the achievements of the recent research and on the outstanding questions that remain.

Indeed, our belief is that an acknowledgement that the truly fundamental questions remain to be resolved is, for researchers, itself a hopeful stance. We believe that international trade economists will rise to the occasion to make our field richer and more complete.

2 Interaction of Theory and Empirics

The folklore of international economics holds that there is a simple difference between the sub-fields of international finance and international trade. In international finance, every theory ever proposed is decisively rejected by the data. In international trade, no theory ever proposed has ever been touched by data. This is, of course, a parody. But like many parodies, it contains a grain of truth.

Data analysis has traditionally played a very marginal role in the field of trade. While macro and labor economics, for example, have the interaction of data analysis and theory as the lifeblood of the field, this has not been so in trade. This is what TREFLER (1995) had in mind when he wrote: "In other fields of economics, the poor performance of a major theory leads to more careful consideration of the data and to new theories that can accommodate the anomalies." By contrast, he argued, the work in trade had by and large only produced conjectures, but no alternatives shown to do better. LEAMER and LEVINSOHN (1995) argue that only two empirical results have materially affected the way international economists think about trade.[1]

The marginal role of empirics in trade is easily discerned in other fora. Graduate reading lists typically feature only a minute selection of empirical papers relative to the body of theory to be mastered. Theorists are vastly more likely to say that their work is inspired by other theoretical work or by a few stylized facts than by any more resolute data analysis. Perhaps the ultimate metric of the extent of the marginalization of empirics within the field is the fact that with the last change of editorship, the *Journal of International Economics* felt it necessary to institute an affirmative action plan for empirical articles.

Why has empirical work in trade, in contrast to other fields, had so little influence on the evolution of the field? There is no single answer. One part of the answer is surely that over much of the last half-century, articulation of the theory has proven very fertile ground. Elaboration of the neoclassical theory, the great advances in commercial policy, increasing returns, imperfect competition, trade and growth, and more recently economic geography – these have been tremendous contributions to our understanding of international economic relations. One certainly cannot say that the field has been sterile.

Yet the field has nonetheless been extraordinarily unbalanced. A second reason for this is that the project itself is rather daunting: to provide a parsimonious characterization of the principal determinants of the structure and evolution of production and absorption, hence trade, across countries. Of course, to say that the project is daunting is also to say that

1 The exceptions are the Leontief paradox and the demonstration by GRUBEL and LLOYD (1975) that a great deal of trade is intra-industry trade. The fact that LEAMER (1980) has strongly challenged LEONTIEF's finding and that we believe there is serious reason to question the meaning of the GRUBEL-LLOYD results (see below) indicates the limited reach of the empirical side of trade.

the returns to success should likewise be high. Certainly the limitations of the data, both in availability and quality, have been an issue. But with improved data collection by a variety of international agencies, and their systematization by various researchers, including LEAMER and FEENSTRA, these constraints are declining.

Finally, LEAMER and LEVINSOHN (1995) make the point that the work has failed to be persuasive because the experiments themselves were often not well formulated. This is no doubt true, but it begs the prior question of why the self-correcting mechanisms that lead other fields to concentrate intellectual firepower on relating the theory to the data had relatively little effect in the field of international trade.

3 Solid Empirics, Low Impact

LEAMER has criticized many empirical papers as not having put enough intellectual capital on the line. But this is only part of the story of the limited influence of data analysis in our field. Many excellent empirical papers, including some by LEAMER himself, have put a lot of intellectual capital on the line, found an important tenet of international trade theory wanting, and ended up off the radar screens of most trade economists. In this section we will explore some results that are well established in empirical trade, which should be part of the empirical toolkit of every trade economist, but which have had very limited impact on the way we think about international trade.

3.1 The Failure of Factor Price Equalization and the Role of Comparative Advantage in the OECD

Let's start with an important fact. It is well known that factor price equalization fails. Wages differ strongly across regions within countries and enormously across countries. International economists tend to hold two stylized facts in their heads with regard to this. The first is that wage differences are small across developed countries, and the second is that they are large between developed and developing countries. These facts are true, but one must also consider the magnitudes. It is not uncommon for wage differentials between developed and developing countries to be on the order of thirty or more. However, even in the OECD, wages vary by a factor of five. These are big numbers. *Figure 1* portrays average compen-

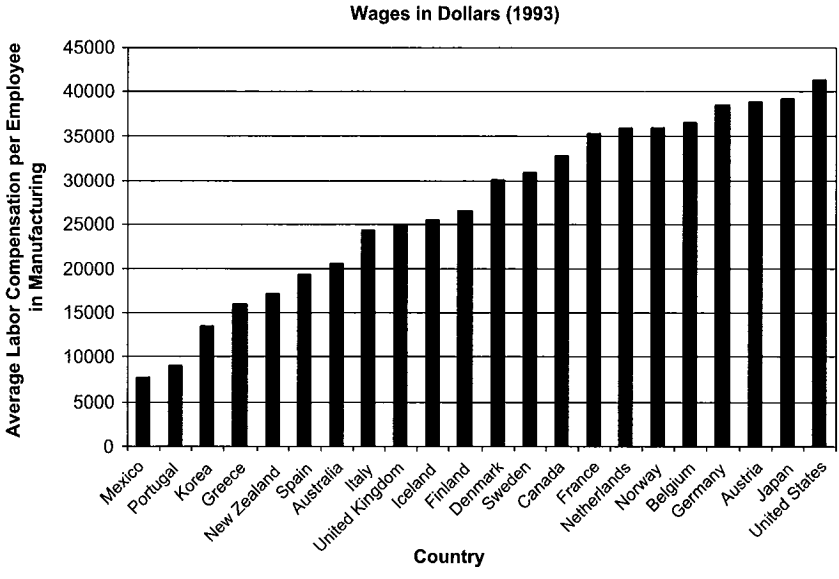
sation within the OECD. Even when considering relatively wealthy countries like Australia, Italy and the US, wages vary by a factor of two or more. The most likely explanations for this wage disparity in the OECD are differences in labor quality, productivity, and differences in endowments. Regardless of which of these stories one finds most plausible, it is hard to escape the conclusion that classical comparative advantage is likely to be quite important in the North.

3.2 One Cone or Many?

LEAMER (1987) was the first to provide solid evidence that one reason for the failure of factor price equalization was the fact that there are multiple cones of diversification. This paper made a clear contribution by putting some important intellectual capital on the line. If factor price equalization were true or if factor price equalization failures were due to factor quality or productivity differences, one would expect to see a linear and not quadratic relationship between country capital labor ratios and output per worker in any given sector. His finding of a quadratic relationship seems to us to be strong evidence against a single cone world. It is interesting to ask why a paper like this is typically not a required reading for graduate students. We teach both the factor price equalization and no-factor price equalization models but spend essentially no time worrying about which world we occupy.

This is especially surprising considering that other studies tend to confirm LEAMER's (1987) basic result. DOLLAR, WOLFF and BAUMOL (1989) make the search for multiple cones a centerpiece of their analysis. Using a somewhat different methodology than Leamer, they find that both industry output per worker and industry capital-to-labor ratios are highly correlated with country endowments. Indeed, the median correlation between industry capital to labor ratios and country ratios is 0.62. Over the course of the next decade, DAVIS and WEINSTEIN (2001a), and SCHOTT (1999) confirmed this basic result using complementary methodologies.

Figure 1 World wages in manufacturing (in 1993 US$)



Source: 1997 OECD Structural Analysis (STAN) Industrial Database for the Years 1970-1995

A generous interpretation of why papers like LEAMER (1987) do not enter the canon is concern with robustness. However, if this is so, one must ask why no one published a critique. A more likely reason for the silence is that LEAMER's result is inconvenient for trade theory. Both "constant returns to scale" and "increasing returns to scale" enthusiasts love the beauty and simplicity of factor price equalization models. Multi-cone models are messy. It is a testament to the power of elegant theory that few seized on the importance of these results. Even though there was strong evidence that a particular cause for the failure of factor price equalization is evident in international data, the general response of trade economists, both empirical and theoretical, has been to continue thinking in terms of factor price equalization models.[2]

Ironically, when OHLIN wrote "complete equality of factor prices is [...] almost unthinkable and certainly highly improbable", he got it half wrong. In spite of being completely improbable, factor price equalization was far too easily thinkable. The fate of LEAMER (1987) illustrates a problem that

2 An interesting contrast is the strong professional interest accorded TREFLER (1993).

empirical researchers face. Studies that put intellectual capital on the line and confirm our preferred view of the world tend to do much better than studies that contradict our priors.

3.3 Industry Level technical Differences in the OECD

Another robust empirical result that tends to get pushed to the side is the role played by Ricardian differences. There have been innumerable studies that have demonstrated that industry-level technical differences in the OECD are large. Within this literature, JORGENSON, KURODA and NISHIMIZU (1987) are notable in finding that even after matching the international data as carefully as is possible, enormous technological differences remain. They found that in 1985 in over two-thirds of the tradable goods sectors they examined, productivity in Japan was either 20% below or 20% above the US level. In perhaps the only trade paper to take both the theory and the data in this area seriously, HARRIGAN (1997) found that these industry technological differences matter for international specialization. Despite the plethora of studies showing industry level technological differences are big even within the OECD, most trade economists abstract from this when thinking about determinants of trade within the OECD.[3] Again the profession seems fairly timid about engaging the data.

3.4 What is Intra-Industry Trade, anyway?

The evolution of our understanding of intra-industry trade illustrates the successes and failures that occur as theorists and empirical economists communicate. KOJIMA (1964) was the first economist to note the large amount of intra-industry trade. GRUBEL and LLOYD (1975) expanded and greatly enhanced this early analysis and laid the foundation for much of our thinking about the empirical importance of intra-industry trade. Ultimately, two popular theories of intra-industry trade arose. The first, based on KRUGMAN (1979) and LANCASTER (1980), held that intra-industry trade is the exchange of horizontally differentiated goods produced with identical factor intensities. The second, based on FALVEY (1981), suggested that intra-industry trade represents vertically differentiated products of different factor intensity. In the end, the KRUGMAN-LANCASTER approach to intra-industry trade became the prevailing view because it could be pre-

3 EATON and KORTUM (1999) are a notable exception.

sented in an elegant, comprehensive, and compelling framework that tied together what were viewed as key stylized facts.

Interestingly, data analysis did not play much of a role in the success of the theory. The debate revolved around two easily observable issues. First, did exports in a sector to a country differ significantly in quality from imports from that country, and second, were they produced using differing factor intensities? A simple test of the first issue is straightforward. Are unit values in bilateral trade similar for exports and imports? Interestingly, the first careful study that put this intellectual capital on the line was not written until over a decade after the original theory. Defining goods to be vertically differentiated if unit values at the 5-digit level differ by more than ±15%, GREENAWAY, HINE and MILNER (1994) find that 70% of UK intra-industry trade is vertical. A similar study by ATURUPANE, DJANKOV and HOEKMAN (1998) found even higher shares for vertical intra-industry trade for other European countries.

Of course, a major worry with this sort of empirical work is whether a mechanical cut off of ±15% is really separating vertical and horizontal specialization. Obviously, as the bands expand more trade will be classified as horizontal, and one is left wondering whether this type of study is really informing us about the world. To get a better sense of the meaning of these results, it makes sense to take a closer look at the data. The most detailed Harmonized Tariff System data is at the ten-digit level. At this level of aggregation, there are 11,297 different agricultural, mining, and manufacturing product categories. Unfortunately, a quick look at these categories suggests that this is a tremendous underestimate of the true level of heterogeneity in the world.

Table 1 presents some sample categories that reflect products that we know something about. What is striking about each category is how much scope for both vertical and horizontal product differentiation there is even at the ten-digit level. Would you feel comfortable entering a good restaurant and asking simply for a bottle of red wine? How much should one pay for a four-cylinder passenger car and what would one expect to receive? Would you order generic Swiss cheese on the internet and feel confident about what would arrive? Clearly, quality differences within these 10-digit categories may support tremendous price differences even when the varieties are sold side-by-side.

Table 1 Sample Ten-Digit Data

HTS Code	Description
0201306000	Meat bovine animals, boneless ex processed, fresh or chilled
0403100000	Yogurt, sweetened, flavored or containing fruit/coco
0406904520	Cheese, swiss or emmenthaler with eye formation
0702002000	Tomatoes, fresh or chilled
0709510000	Mushrooms, fresh or chilled
0711201500	Olives, not pitted
0808100000	Apples, fresh
0901210030	Coffee roasted not decaffeinated for retail under 2 kg
1604142020	Tuna, albacore, no oil airtight container under 7kg
1806900075	Chocolate confectionery put up for retail sale
2007991000	Strawberry jams
2204214005	Red wine grape under 14% alcohol
2208303030	Whiskies, scotch & irish, container not over 4 liters
3004906075	Cough and cold preparations
3004906020	Cardiovascular medicaments
3004400050	Dermatological agents and local anesthetics
3002200000	Vaccines for human medicine
3004400020	Anticonvulsants, hypnotics and sedatives
3004400030	Antidepressants, tranquilizers and other psychiatric agents
3926301000	Handles and knobs
4202219000	Handbags, outer surface of leather, value over $20 each
4901990050	Technical, scientific and professional books
6103110000	Men's or boys' suits of wool, knit
6104531000	W/g skirts of synthetic fibers cont 23% more wool, knit
6104622010	Women's trousers of cotton, knitted
7103910010	Rubies cut but not set for jewelry
8411919080	Parts of turbojet or turbopropeller a/c engines
8703240032	Passenger motor vehicle, 4 cylinder & under
9004100000	Sunglasses
9006530040	Camera, 35mm with built-in electronic flash
9202100000	String musical instruments played with a bow
9306900040	Bombs, grenades, torpedoes, & similar munitions of war
9503411000	Stuffed toys
9506512000	tennis rackets, strung

Consider a typical category: men's and boy's knit wool suit-type jackets and blazers. Even at this level of disaggregation there is still substantial intra-industry trade. In 1994 Japan exported these jackets to 19 different countries and imported them from 31 countries. Overall, Japan's Grubel-Lloyd index of intra-industry trade, even at this tremendously fine division of the data, is 0.20. Interestingly, the two largest suppliers of men's

wool suit jackets to Japan in terms of value are Italy and China. These two countries account for almost one-quarter of total Japanese imports in this sector. This fact alone strongly suggests that even at the ten-digit level, very different types of goods are being aggregated together. Unit values confirm this. The typical unit value for an Italian wool suit jacket is almost seven times higher than that of a Chinese wool suit jacket. The unit values for Japanese exports of suit jackets are triple those of Chinese imports. Clearly there is a lot of vertical differentiation here.

This ad hoc analysis is easy to attack. We only picked a tiny subset of sectors and clearly have an agenda. What if we had picked a sector that we "know" to be homogeneous, like non-durum wheat meant for human consumption?[4] Even here we find Canadian wheat pellets entering Japan with unit values that are 23% above Australian pellets. And this is wheat! If we move slightly downstream to wheat flours, unit values skyrocket to factors of eight or more. This tends to confirm a problem with our tendency to group products we know nothing about as being differentiated only horizontally. Perhaps there is less quality variation in polyacetals, manure spreaders, or bovine semen, but we think there is cause for alarm.

What is worse is that Italian and Chinese suits are likely to be produced with very different factor mixes. This point was made early in the debates over intra-industry trade by FINGER (1975) and CHIPMAN (1992). More recently, DAVIS and WEINSTEIN (2001b) examine the implications of this for our measures of net factor trade. If matched intra-industry trade was the exchange of goods produced with identical technologies, the net factor content of such trade would be zero. In fact, we find that for many OECD countries over half of their net factor trade is accomplished through intra-industry trade. The United States is a particularly striking example. Over two thirds of its net factor trade is accomplished by intra-industry exchange of goods of differing factor intensity. Much of what we call intra-industry trade is simply a data problem that reflects the failure of our industrial classification system to capture the fact that very different goods are being lumped together.

To say that these studies have made little impact on the day-to-day thinking of most trade economists is a gross understatement. A typical graduate student at a top department is likely to believe that intra-industry trade

4 This category is drawn from the Commodity Classification for Japanese Tariff Statistics, which is actually more disaggregated than the HTS system.

being the exchange of goods of similar factor intensity is true simply as a matter of definition. That this bears little relation to measured intra-industry trade does not even present itself as a problem. Such gross errors would be inconsequential if not for the fact that they form the core around which a great deal of theorizing occurs. And our beautiful models hold a tenacious grip on the way we view the world.

3.5 How Similar are Endowments in the OECD?

A final stylized fact often ignored concerns endowment similarity. It is often asserted that OECD countries have endowments of factors that are similar. While it is true that there has been substantial income convergence in the OECD, enormous differences in factor abundance remain. A natural way to measure factor abundance is to divide each country's endowment of a factor by its share of world GDP multiplied by the world endowment of that factor. This produces a unitless measure of abundance that indicates what share of a factor should be exported in a frictionless factor price equalization Heckscher-Ohlin-Vanek (HOV) world (Note the sleight of hand – even we feel compelled to appeal to the frictionless Heckscher-Ohlin-Vanek model as a baseline! But see DAVIS and WEINSTEIN (2001a)).

Table 2 Distribution of Country Measured Factor Abundances in 1990

COUNTRY	Total Labor	Capital	College Educated Labor	High School and Below Educated Labor
World Average	2.23	0.76	1.00	2.90
World Standard Deviation	2.37	0.37	0.87	3.21
World Median	1.16	0.76	0.76	1.43
G10 Average	0.33	1.12	0.84	0.54
G10 Standard Deviation	0.07	0.19	0.32	0.15
G10 Median	0.31	1.12	0.80	0.52
OECD Average	0.45	1.12	0.90	0.75
OECD Standard Deviation	0.25	0.21	0.46	0.41
OECD Median	0.35	1.14	0.79	0.58

Source: Education Data is average for 1985–1990 from BARRO and LEE (1996); Total Labor and Capital Data is from the Penn World Tables Mark 5.6; G10 corresponds to the 10 countries in DAVIS and WEINSTEIN (2001a).

Table 2 reports the results of this exercise for four factors: aggregate labor, capital, college educated labor, and labor with less than a college education. For OECD countries, moving one standard deviation from the median often makes the difference between a country being a predicted exporter or importer of a factor's services. This suggests a *prima facie* case in favor of endowment differences mattering even for trade within the OECD. It is not uncommon to find countries that are in the lowest quartile have abundances that are less than half of those in the upper quartile.

One possible criticism of this is that countries like Mexico, Korea, and Turkey may be driving the results. To see if this were true, we also considered a subset of 10 wealthy and large countries in the OECD (Australia, Canada, Denmark, Germany, France, Japan, Italy, Netherlands, Britain, and the United States). The data reveal substantial differences in endowments even among these countries. The standardized US endowment of college-educated labor is almost five times that of Italy. The United Kingdom's standardized endowment of non-college labor is almost double that of the United States. And Japan's standardized endowment of capital is almost double that of the United Kingdom. Clearly, factor endowment differences are alive and well in the North, although this fact seems largely ignored by the profession.

3.6 How Should We Respond to Uncomfortable Facts?

Ideally, economic theory serves in part as a way to organize key facts about the world. Strategic simplification is essential. A consequence is that our theories are always wrong in some dimensions – this is a necessary fact of life. But we expect them to be right about the key facts around which they are organized. We have presented what we view as important examples in which empirical research has had something substantive to say about the theory, but these facts have had little influence on the way economists think about trade. In certain cases, the romance of the models has had the upper hand on the facts.

We all recognize the fact that anomalies play an important role in the advance of knowledge. But it is always more convenient for the anomalies to grow in someone else's garden. We think that there are important examples – for example, the work of TREFLER (1995), discussed below – where the characterization of an anomaly has played an absolutely crucial role in advancing our understanding of trade patterns. Perhaps the best

we can do is to all take a pledge to work harder to embrace our anomalies as the start of richer theories.

4 Virtually All of the Key Questions Remain Open

One of the great joys of academic economics is to encounter an area in which the most important questions have yet to be resolved. Surely there must have been great excitement as it became evident that a tremendous body of industrial organization theory could usefully be applied to problems of international trade. The same excitement no doubt existed for an earlier generation in consideration of the neoclassical theory of commercial policy, or more recently in work on trade and growth, political economy, or economic geography. Often the simple recognition that an important area and its major problems are open terrain is among the largest steps in finding answers.

Empirical international trade, in our view, is just such a field. Virtually all of the major questions remain quite open. Some are almost untouched. What is the role of increasing returns versus comparative advantage in determining international trade patterns? What role do endowments play in trade patterns beyond North-South trade? How do technological differences at the industry or firm level interact with other determinants of cost in shaping trade patterns? In a world of imperfect integration, how do absorption and production patterns interact? These are absolutely fundamental questions. They are also quite open.

We do not mean in the least to say that the existing empirical literature has taught us nothing about trade patterns. But it is important to understand the limitations either of the questions asked or the answers received. We will consider a few examples.

One of the signal successes of empirical trade is the so-called gravity model. It relates bilateral trade volumes to a parsimonious set of determinants. It fits well whether we look at aggregate trade volumes or instead at industry trade volumes. The fits of the estimating equations really are impressive, with typical values of R^2 in the range of 0.7. The gravity model, once considered a theoretical orphan, now has several sets of parents in waiting, with new ones arriving almost daily.[5] Yet the meaning of the

5 See, for example, DEARDORFF (1998) on gravity in a neoclassical world and FEENSTRA ET AL. (1998) on gravity with oligopolistic competition.

gravity equation's success for our understanding of international trade is worth closer examination.

The core of international trade theory has always focused on the determinants of the pattern of production as the key fact to be explained in understanding trade patterns.[6] Yet the gravity model, e.g. in its industry-level approach, takes the level of production as given, and then seeks to explain the distribution of imports across partner countries. Thus, even if one is willing to be surprised at how well the gravity model fits, the deeper question is what we can infer from these good fits. The fact that the empirical model takes the distribution of production as given should make clear that it would be very hard to use the good fit of the gravity model as evidence for one theory of the determinants of production patterns over any other.

The recent literature focusing on the near-universality of gravity has instead focused on the fact that it might provide evidence of a high degree of specialization, whatever its source. Yet FEENSTRA ET AL. (1998) have shown that gravity can arise even in a homogeneous goods model without a high degree of specialization. For all their good fits, the thousands of gravity models that have been run have done relatively little to inform our understanding of the deep determinants of trade patterns. Papers such as FEENSTRA ET AL. (1998) that actively seek to distinguish alternative models based on their performance in the gravity framework are an important contribution. But this work is still far from complete.

The literature also features important papers establishing robust correlates of international trade and production. Stellar contributions in this genre include LEAMER (1984) and HARRIGAN (1997). Yet, as LEAMER cautions, these represent incomplete tests of the theory. They do not try to get the pieces to fit together. The estimated parameters do not correspond to the structural parameters suggested by the theory.

The interested reader is encouraged to consult the surveys by LEAMER and LEVINSOHN (1995) or HELPMAN (1999). We think our assertion will stand: virtually all of the most important questions in empirical trade remain to be resolved.

---

6 Noteworthy exceptions exist. The LINDER theory is one example, as would be the recent work on economic geography, in which market segmentation leads to a more intimate interaction between demand and production patterns.

5 The Costs of Failing to Distinguish Models Empirically

Our understanding of the determinants of actual trade patterns is not deep. This is a problem in its own right. It becomes a yet larger problem when we turn to normative and policy analysis. This has been quite evident in the very extended discussion in the United States in recent years over the reasons for the rising relative wage of skilled workers and the role that trade may have played in this.

Among the many leading trade economists who contributed to this discussion were JAGDISH BHAGWATI and PAUL KRUGMAN. One of them wrote: "Unusually, serious economists have not by and large argued about theory: with few exceptions they have agreed that a more or less classical Heckscher-Ohlin-Samuelson model is the best framework to use." The other titled a section of a paper "Why FPE [factor price equalization] and [Stolper-Samuelson] Theorems are Inadequate Guides to Reality", with a first sub-section noting the potential gains from exploitation of scale economies as a counterweight to concerns about wage losses. For those who have not followed the debate closely, it might not have been evident that the first quote comes from KRUGMAN, the second from BHAGWATI (co-authored by VIVEK DEHEJIA).[7]

A skeptic could argue that this apparent plasticity of belief about the appropriate underlying framework confirms that policy analysis is just ideology in fancy garb. Or, as a Columbia economics department Christmas skit once averred, it is a case of the assumptions following straight from the conclusions.

Such a skeptic would miss the central point: honest disagreement about which model should be applied in any given context, and even shifting from one to another in different settings, is at present not only respectable but entirely necessary. One reason for this is our reliance on MIT-style theory. This approach asks a model to be crisp and to the point; it does not ask the model to be a picture of reality. The deep beauty and great value of MIT-style theory is unassailable. But, as we have seen, it carries a price when it is not accompanied by a serious effort to distinguish alternative frameworks on empirical grounds. When turning to policy issues, it is a matter of judgment which simple model to apply. Serious

7 The quotes are from KRUGMAN (1996) and BHAGWATI and DEHEJIA (1994).

economists can have honest disagreements. And these disagreements can make all the difference for the conclusions.[8]

6 What Can We Expect of Empirical Work?

We have argued that there would be great value to arriving at a stronger consensus about appropriate models of trade. But at least some prominent voices have expressed skepticism about whether this is a feasible project. This raises a series of questions. What should empirical analysis of positive trade be doing? What interaction should there be with theory? What is the role of estimation? What is the role of testing? What does theory have to learn from empirical work? What is the objective of this entire enterprise? These are among the most basic questions of our field and we spend too little energy grappling with them.

LEAMER (1992) argues that a great deal of empirical analysis fails to be persuasive because it tests propositions that we know to be false. Models in this view are not literally right or wrong; instead they are useful or not. While holding fast to the idea that a persuasive data analysis must be developed in the context of a well-articulated theory, LEAMER issues the injunction: "Estimate, don't test!"[9]. This is, of course, a stricture that LEAMER himself has violated – even in some of his most influential work. This should provide a hint that the injunction is too strong, and for a less accomplished empiricist, could be seriously misleading about the project of empirical trade.

The central object of empirical work in trade is narrowing the range of plausible belief. If all *ex ante* plausible views are untouched as a result of an empirical analysis, then it will strike earth with a resounding nothing. How does one place intellectual capital on the line? Sometimes, we are simply looking for a number. We are willing to take as given, for the exercise, the underpinnings. We want to know a plausible value for an elasticity. We want to know a speed of adjustment. This is LEAMER's "estimation". Such estimation is a thoroughly important part of the enterprise. The accumulation of studies that provide stylized facts about the economy do successively narrow the range of plausible belief.

---

8 It is worth noting, though, that in this issue both BHAGWATI and KRUGMAN arrived at the same substantive conclusion.

9 See, for example, BOWEN, LEAMER and SVEIKAUSKAS (1987).

Some caution, though, is warranted with a subset of such studies. It has become a too-frequent practice to use a framework as the basis for a study, estimate parameters, and if they are in (very gross) accord with the predictions of the theory, to pronounce it as being "consistent" with theory "xyz". Strictly, this is not incorrect, but it is often seriously misleading. This is so particularly when virtually any theory that might be in the least interesting is likely to yield the same or similar predictions. Why not take the extra step and seek to identify predictions that might usefully distinguish the models?

We also believe that the prospects for persuasive testing are more hopeful than are indicated by LEAMER. He is quite right that there is no point in testing and rejecting propositions that we know beforehand to be false. But there is no reason to allow the existence of pointless exercises to define our attitude to testing more generally. We are strongly convinced that researchers can identify hypotheses in which two well-defined theories have contrasting implications, hence in which it is possible to test. The criterion for whether or not this is interesting has to be whether some real intellectual capital is placed on the line via the test. Will we look differently at the world depending on the results of the test? That there are many cases for which the answer is "no" should not discourage us from identifying cases where the answer is "yes". We believe that such well-designed tests can be a crucial part of a research program that successively narrows the range of plausible belief.

In physics, there has long been discussion of a "theory of everything". Its counterpart in international trade is to give a parsimonious account of the world general equilibrium.[10] Is there a way to specify the nature of differences in technology, endowments, tastes, plus the underlying parameters of trade costs that makes sense of world patterns of production, absorption, and trade? That should be our aim. We believe that the field is open for a great deal of progress.

10 Partial equilibrium, of course, is why a dog chases its tail; general equilibrium is why the dog's chase is in vain.

7 Heckscher-Ohlin-Vanek is Dead; Long Live Heckscher-Ohlin-Vanek!

We believe that the project of successively narrowing the range of plausible belief by testing is not only a hypothetical possibility but a process already under way in a number of areas of trade. While a number of areas of inquiry could equally well have served as a model, the focus on OHLIN and our own research proclivities lead us to focus on recent work considering the Heckscher-Ohlin-Vanek (HOV) theory.

The work of BOWEN, LEAMER and SVEIKAUSKAS (1987) is, in our opinion and that of the larger profession, a monumental contribution to the empirics of international trade. Very likely this is the single most widely read empirical paper on trade. We believe that an important reason for the influence of this paper is its substantive conclusion that the Heckscher-Ohlin-Vanek model has little predictive power for the measured factor content of trade. Perhaps oddly, our opinion of the paper's importance has little to do with its substantive conclusions; or rather we think highly of the paper in spite of the fact that its substantive conclusions are unconvincing.

The major contributions of the BOWEN, LEAMER and SVEIKAUSKAS (1987) paper are several. This was the first paper to report results on HOV for a large number of countries, based on a wide array of endowments, trade, and technology. The sign and rank tests employed to measure the model's performance have become standard in the literature. Moreover, the hypothesis testing developed in the later sections of the paper also proved to be very important in later research, such as that of TREFLER (1995).

By far, though, the most important contribution of BOWEN, LEAMER and SVEIKAUSKAS is its conceptual grandeur: it dares to rise to the challenge of assembling all of the empirical pieces to describe a world general equilibrium. In LEAMER's terms, it provides a "complete" test of the HOV theory, employing data on endowments, technology, and trade. That it fails utterly to assemble the pieces in a coherent way is wholly secondary. The attempt itself changed the field.

The results of BOWEN, LEAMER and SVEIKAUSKAS, on their face, were devastating for the HOV theory. In their implementation, factor abundance provides no more information than a coin flip about which country will be measured to export services of a particular factor. What could be more

damning? In spite of their best efforts, they were unable to identify a model that performed better.

Faced with such results, what was the reaction of the profession? On the one hand, the results were likely to be difficult to accept; trade economists receive the HOV theory with mother's milk. It may have seemed very hard to believe that observed differences in endowments really have no influence on net factor trade. On the other hand, the results seemed to lend greater credibility to an emerging consensus that, however important relative factor endowments may have been in the past, they no longer matter much in determining trade patterns. Trade of jute for aircraft may be explained by Heckscher-Ohlin, but the bulk of trade is among countries that hardly differ in endowments, so the dramatic failure of HOV really presents no puzzle.

The next real landmarks in this literature are the papers by TREFLER (1993, 1995). What is most remarkable in TREFLER's papers is that they were written at all. In the wake of the BOWEN, LEAMER and SVEIKAUSKAS paper, it would have been easy to conclude that the HOV theory was a dead end, perhaps something for historians to contemplate, but not a path for new research. TREFLER's sound judgment was that it could not be satisfying to declare the theory dead when we really had no idea why it was failing.

TREFLER asked two key questions. The first follows up directly on the work of BOWEN, LEAMER and SVEIKAUSKAS: are there simple amendments in the spirit of HOV that allow the theory to work? The second is more novel (at least within empirical trade): are the failures systematic? The latter, in particular, proved to be an extraordinarily fruitful question. And the answers TREFLER provides are striking. The most memorable regularity he identifies in the data is what he terms the "mystery of the missing trade". In simple terms, the measured factor content of trade is an order of magnitude smaller than that predicted based on national incomes and endowments. This characterization of the data has been extraordinarily useful in focusing subsequent research on the types of amendments that might be needed to fit the pieces of the puzzle together.[11]

11 For a more complete discussion of TREFLER's methodology and conclusions, see the survey by HELPMAN (1999) and the references therein.

It seemed clear in the wake of BOWEN, LEAMER and SVEIKAUSKAS that the pure factor price equalization version of HOV would be a dismal failure if applied to a broad cross section of countries. This left two paths open. One approach to this is to look for ways to sidestep the problem while continuing to work broadly within the HOV framework. This is pursued in DAVIS ET AL. (1997). The starting point for that paper is to ask what HOV predicts if only a subset of the world shares factor price equalization – an factor price equalization club. This has a definite answer and provides the basis for tests provided a suitable factor price equalization club can be identified. Importantly, the focus on general equilibrium prohibits discarding information on the rest of the world. However, the rest of the world must be incorporated appropriately.

We chose the regions of Japan as our factor price equalization club. This has a number of advantages, including the high quality and comparability of the data, and the heightened plausibility of factor price equalization for regions of a single country. A second important characteristic of DAVIS ET AL. (1997) is that while prior work focused solely on the factor content of trade, we were able to examine separately the HOV theories of absorption and production. This allowed us to see directly where the failures in predicting factor contents might arise, rather than needing to rely on indirect inferences.

The DAVIS ET AL. (1997) paper replicates the failures of the theory identified in prior work for the case in which it assumes that the whole world shares factor price equalization. The mystery of the missing trade is then very evident. However, it also shows that when you drop the assumption of universal factor price equalization, restricting this to the factor price equalization club of Japanese regions, the results improve dramatically. The regions export the services of their abundant factors, and they do so in approximately the right magnitude. The mystery of the missing trade is in large measure eliminated for the regions of Japan. Both the production and consumption theory of HOV fare reasonably well in the Japanese data. This provides a first case of HOV working while considering the problem within a full world general equilibrium.

The problem of getting HOV to work while directly confronting the failure of factor price equalization internationally is addressed in DAVIS and WEINSTEIN (2001a). Prior work on an international sample had focused on two key reasons for the failure of HOV: (1) countries use different techniques of production, possible reasons being efficiency differences or

a breakdown of relative factor price equalization; and (2) the absorption theory based on identical and homothetic preferences may be at fault. Our starting point was to note that while the key hypotheses for the failure of HOV concerned technology and absorption, the prior work employed only a single observation on technology (that of the US) and no data on absorption. An obvious strategy was to assemble more data to explore the nature of these failures directly, which should help in selecting which among the competing hypotheses really matters in trying to get an amended HOV to work.

For details of implementation, consult DAVIS and WEINSTEIN (2001a); we focus here just on the conclusions. In line with the literature on cross-country productivity (see, for example, JORGENSON and KURODA 1990), efficiency differences matter. The failure of factor price equalization matters, even within the OECD: more capital abundant countries use more capital intensive techniques within each industry. Non-traded goods play an unexpectedly important role, both in allowing us to make inferences about the failure of factor price equalization, and also by the fact that when factor price equalization fails they tend to absorb a great deal of the "excess" factor supplies that otherwise might have been available for factor service exports. Finally, trade costs matter, by reducing the opportunities to arbitrage the factor price differences.

Having directly estimated the nature of efficiency differences, the failure of factor price equalization and its implications for production techniques, and the role of trade costs in reducing trade flows, how well does the model predict net factor trade? In considering the answer, it is well to keep in mind that due to the "mystery of the missing trade", the answer in the prior literature is that the model correctly predicts almost nothing. Here, having taken advantage of the new and richer data set, measured factor trade is approximately 60% to 80% of predicted factor trade. The mystery of the missing trade is, in large measure, resolved. Countries export their abundant factors and they do so in approximately the right magnitude. Suitably amended, HOV works.

At this point, it is tempting to append a fairy-tale ending. There was a moment in which all appeared lost for the HOV theory; now the theory has been rescued and provides a beautiful description of the workings of international trade. However, as devoted researchers, we do not believe in endings, fairy-tale or otherwise.

We do, though, believe that the profession's experience with the path of research on HOV holds important lessons. Some of these are substantive. We do believe that HOV, or Heckscher-Ohlin more broadly, will have to be an important component of any empirically based attempt to understand the pattern of trade.

Perhaps, though, the most important lessons have to do with the future approach to research in the field of international trade. There is no reason that this should be a field of very slight empirical content. It can preserve the traditional commitment of the field to elegant general equilibrium modeling and at the same time make progress in terms of matching theory and data in a coherent way. The models that emerge will surely be composites of the various approaches in the literature to trade patterns. However, if we use enough imagination, we can develop these hybrids so that they are both elegant in theory and robust when confronted with data. At least that is how we conceive of the project of future empirical research into trade patterns.

8 Conclusion

The field of international trade is falling short in its central mission. That mission is to understand the causes and consequences of trade in the world we actually inhabit. Trade economists can justly take pride in the theoretical achievements of our field. But these have not been matched with equally illustrious progress on the empirical side. Indeed, data analysis has long played a marginal role in the professional life of our field. Notable individual contributions notwithstanding, virtually all of the most important empirical questions remain open and at times nearly untouched.

The failure of our field to grapple seriously with empirics bears a cost. Our failure to identify a positive model adequate to describe the principal empirical features of trade leaves us in serious straits when we turn to policy analysis. Such analysis requires that we specify a positive model as a foundation. It is easy to appreciate that with empirical analysis having done so little to constrain the model that we select, such policy analysis is likely to be highly sensitive to the analyst's priors of which model is appropriate.

Empirical researchers must shoulder part of the responsibility for this state of affairs. They must insure that their exercises truly place intellectual capital at risk in order for their analyses to be persuasive. But the field more

broadly also needs to accept part of the responsibility. For long stretches it has operated from small collections of stylized facts that at times seem impervious to the intrusion of actual facts. Empirical analysis with substantive insights about the features of the world we inhabit, but which are at times inconvenient for theory, languish in obscurity.

We do believe that there are positive models of what the field can achieve when it is able to concentrate a larger share of its intellectual resources to investigate well-defined empirical projects. While several ongoing research dialogues could usefully serve as exemplars, the focus of our own research interests leads us to focus on verification of the Heckscher-Ohlin-Vanek model. This is truly a case where the contributions of many economists, including failures and successes for the models, played a crucial role in shaping our view of the problem.

The approach we suggest involves a re-conception of the collective project of our field or, at the very least, a strong shift in priorities. Crisp, lucid theory will always play a central role in the field. But this needs to be complemented by a serious encounter with data. Grappling with facts revealed by the data, pressing the limits of what our models can predict, and identifying the contours of the world should be viewed as a central part of the program of our leading empiricists and theorists.

This is a clarion call to a project that we see at least partly in progress. There is a relatively small, but influential, group of well-established empiricists and theorists who have actively undertaken research in this area or considered it at length in their own writings. There is a larger group of younger economists who have made it a key element of their work. It is time for each international economist to accept the challenge to make empirical analysis a central feature of our work and dialogue. We have a world to discover.

References

ATURUPANE, CHONIRA, SIMEON DJANKOV and BERNARD HOEKMAN (1998): Horizontal and Vertical Intra-industry Trade between Eastern Europe and the European Union; in: *Weltwirtschaftliches Archiv* 135 (1): 62–81

BARRO, ROBERT and J.W. LEE, (1996): International Measures of Schooling Years and Schooling Quality; in: *American Economic Review, Papers and Proceedings,* 86 (2): 218–223

BHAGWATI, J. and V. DEHEJIA (1994): *Free Trade and Wages of the Unskilled: Is Marx Striking Again?*; in: J. BHAGWATI and M. KOSTERS (eds.) (1994): *Trade and Wages: Leveling Wages Down?;* Washington, DC: AEI Press

BOWEN, HARRY P., EDWARD E. LEAMER and LEO SVEIKAUSKAS (1987): Multicountry, Multifactor Tests of the Factor Abundance Theory; in: *American Economic Review* 77 (5): 791–809

CHIPMAN, JOHN S. (1992): *Intra-Industry Trade, Factor Proportions and Aggregation;* in: JAMES MELVIN, JAMES MOORE and RAY RIEZMAN (eds.) (1992): *Economic Theory and International Trade: Essays in Memoriam, J. Trout Rader;* New York: Springer

DAVIS, DONALD R. and DAVID E. WEINSTEIN (2001a): An Account of Global Factor Trade; in: *American Economic Review* 91 (5): 1423–1453

DAVIS, DONALD R. and DAVID E. WEINSTEIN (2001b): *Do Factor Endowments Matter for North-North Trade?;* NBER Working Paper #8516; October 2001

DAVIS, DONALD R., DAVID E. WEINSTEIN, SCOTT BRADFORD and KAZUSHIGE SHIMPO (1997): Using International and Japanese Regional Data to Determine when the Factor Abundance of Theory of Trade Works; in: *American Economic Review* 87 (3): 421–446

DEARDORFF, A. (1998): *Determinants of Bilateral Trade: Does Gravity Work in a Neoclassical World?;* in: J. FRANKEL (ed.) (1998): *The Regionalization of the World Economy;* Chicago: University of Chicago Press

DOLLAR, D., E. WOLFF and W. BAUMOL (1989): *The factor-price equalization Model and Industry Labor Productivity: An Empirical Test across Countries;* in: R. FEENSTRA (ed.) (1989): *Empirical Methods for International Trade;* Cambridge: MIT Press: 23–48

EATON, J. and S. KORTUM (1999): International Technology Diffusion: Theory and Measurement; in: *International Economic Review* 40 (3): 537–570

FALVEY, R. (1981): Commercial Policy and Intra-industry Trade; in: *Journal of International Economics* 11: 495–511

FEENSTRA, ROBERT, JAMES MARKUSEN and ANDREW ROSE (1998): *Understanding the Home Market Effect and the Gravity Equation: The Role of Differentiating Goods;* NBER Working Paper #6804; November 1998

FINGER, J. M. (1975): Trade Overlap and Intra-Industry Trade; in: *Economic Inquiry* 13 (4): 581–89

GREENAWAY, DAVID, ROBERT HINE and CHRIS MILNER (1994): Country-Specific Factors and the Pattern of Horizontal and Vertical Intra-industry Trade in the UK; in: *Weltwirtschaftliches Archiv* 130 (1): 77–100

GRUBEL, P. and P. LLOYD (1975): *Intra-Industry Trade: The Theory and Measurement of International Trade in Differentiated Products;* London: Macmillan

HARRIGAN, J. (1997): Technology, Factor Supplies, and International Specialization: Estimating the Neoclassical Model; in: *American Economic Review* 87 (4): 475–494

HELPMAN, ELHANAN (1999): The Structure of Foreign Trade; in: *Journal of Economic Perspectives* 13 (2): 121–144

JORGENSON, D. and M. KURODA (1990): *Productivity and International Competitiveness in Japan and the United States, 1960–1985;* in: C. HULTON (ed.) (1990): *Productivity Growth in Japan and the United States;* Chicago: Chicago University Press

JORGENSON, D., M. KURODA and M. NISHIMIZU (1987): Japan-US Industry-Level Productivity Comparisons, 1960–1979; in: *Journal of the Japanese and International Economies* 1 (1): 1–30

KOJIMA, K. (1964): The pattern of international trade among advanced countries; in: *Hitotsubashi Journal of Economics* 5 (1): 16–34

KRUGMAN, P. (1979): Increasing Returns, Monopolistic Competition, and International Trade; in: *Journal of International Economics* 9(4): 469–479

KRUGMAN, P. (1996): *But For, As If, and So What? Thought Experiments on Trade and Factor Prices;* mimeo; Massachusetts Institute of Technology

LANCASTER, KELVIN (1980): Intra-Industry Trade under Perfect Monopolistic Competition; in: *Journal of International Economics* 10 (2): 151–175

LEAMER, E. (1980): The Leontief Paradox, Reconsidered; in: *Journal of Political Economy* 88 (3): 495–503

LEAMER, E. (1984): *Sources of International Comparative Advantage: Theory and Evidence;* Cambridge: MIT Press

LEAMER E. (1987): Paths of Development in the Three-Factor, n-Good General Equilibrium Model; in: *Journal of Political Economy* 95 (5): 961–999

LEAMER, E. (1992): *Testing Trade Theory;* NBER Working Paper #3957; January 1992

LEAMER, E. and J. LEVINSOHN (1995): *International Trade Theory: The Evidence;* in: G. GROSSMAN and K. ROGOFF (eds.) (1995): *Handbook of International Economics. Volume 3;* Amsterdam: Elsevier, North-Holland: 1339–1394

SCHOTT, P. (1999): *One Size Fits All? Specialization, Trade and Income Inequality;* mimeo; Yale University; October 1999

TREFLER, D. (1993): International Factor Price Differences: Leontief was Right!; in: *Journal of Political Economy* 101 (6): 961–987

TREFLER, D. (1995): The Case of the Missing Trade and Other Mysteries; in: *American Economic Review* 85 (5): 1029–46