

SPECIALIZATION AND THE VOLUME OF TRADE: DO THE DATA OBEY THE LAWS?

James Harrigan

Abstract

The core subjects of trade theory are the pattern and volume of trade: which goods are traded by which countries, and how much of those goods are traded. The first part of this paper discusses evidence on comparative advantage, with an emphasis on carefully connecting theoretical models with data analyses. The second part of the paper considers the theoretical foundations of the gravity model, and reviews the small number of studies that have tried to test, rather than simply use, the implications of gravity. Both parts of the paper yield the same conclusion: we are still in the very early stages of empirically understanding specialization and the volume of trade, but the work that has been done can serve as a starting point for further research.

JEL code: F1

Acknowledgements/Disclaimer

This paper will appear as a chapter in *The Handbook of International Trade*, London: Basil Blackwell, forthcoming 2002, edited by Kwan Choi and James Harrigan. The paper benefited from comments by other *Handbook* authors at the Spring 2001 NBER meeting of the ITI group. The views expressed in this paper are those of the author and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System.

Addresses

James Harrigan
International Research Department
Federal Reserve Bank of New York
33 Liberty Street
New York, NY 10045

james.harrigan@ny.frb.org

Specialization and the Volume of Trade: Do the Data Obey the Laws?

James Harrigan
International Research Department
Federal Reserve Bank of New York

The core subjects of trade theory are the pattern and volume of trade: which goods are traded by which countries, and how much of those goods are traded. Most of the rest of trade theory, such as the analysis of policy and the effects of trade on factor prices and welfare, is grounded in models which explain the pattern and/or volume of trade. As a consequence, it is impossible to assess the relevance of trade theory as a whole unless we understand the empirical performance of the core explanations for trade.

The oldest explanation for the pattern of trade, originally due to Ricardo, is comparative advantage. The law of comparative advantage is an unassailable intellectual cornerstone of economics, but until recently empirical research on it has been scant and only loosely connected to theory. In contrast, the law of gravity as applied to explaining the volume of trade has been the foundation for literally hundreds of applied studies, but the gravity model has had a comparatively shallow (if not shaky) theoretical foundation. This purpose of this chapter is to review and critique the last decade or so of empirical research on comparative advantage and gravity.

The first part of the chapter discusses evidence on comparative advantage, with an emphasis on carefully connecting theory models to data analyses. The second part of the chapter first considers the theoretical foundations of the gravity model, and then reviews the small number of papers that have tried to test, rather than simply use, the implications of gravity. Both parts of the paper yield the

same conclusion: we are still in the very early stages of empirically understanding specialization and the volume of trade, but the work that has been done can serve as a starting point for further research.

1 Testing the General Theory of Comparative Advantage

Economists are proud of the theory of comparative advantage, seeing it as both beautiful and profound: beautiful because of its simplicity and elegance, profound because it is surprising and has deep implications for economic policy and our understanding of real economies.

But is the theory of comparative advantage actually useful for helping understand the world? The most fundamental problem about comparative advantage is that it relates observables (trade flows and specialization patterns) to things which are by their nature almost always unobservable (autarky prices). For example, in Deardorff's definitive modern statement of the theory (1980), the general theorem of comparative advantage for a single country is stated as

The value of net exports evaluated at autarky prices is non-positive (Deardorff 1980, pg. 948).

The assumptions required to prove this result are standard but minimalist: they include convex technology, perfect competition, and the existence of community indifference curves. Tariffs and transport costs are allowed, but not trade subsidies. Like all statements of the theory of comparative advantage, Deardorff's is a comparative general equilibrium result: allowing all goods and factor markets to clear simultaneously, it expresses a relationship that must hold between an endogenous variable in one equilibrium (autarky prices) and an endogenous variable in another equilibrium (net exports)¹.

¹ Stronger statements, such as "a country will export all goods which are cheaper in autarky, and import all other goods", are not possible except in

If it were possible to test this theory there would be a lot at stake. The most interesting reasons why the prediction could fail include non-convex technology and/or imperfect competition, as well as perverse trade policies such as export subsidies. Alternatively, markets simply might not work the way we think they do. In short, failure or confirmation of the law of comparative advantage would be very interesting for theorists as well as policy-makers.

Perhaps surprisingly given the general invisibility of autarky prices, there are two recent papers which offer evidence on the relationship between autarky prices and trade flows. The first of these is "An Experimental Investigation of the Patterns of International Trade" by Noussair, Plott, and Riezman (1995) . Noussair et al run laboratory experiments that are intended to satisfy the assumptions of comparative advantage theory. The experimental economy has two countries and two goods, with given linear production functions that differ across countries and identical preferences for all agents. The experimental subjects are undergraduates at Cal Tech and the University of Iowa. The autarky and free trade equilibria in this economy can be easily computed analytically, so the purpose of the experiment is to see how close the data is to the analytical equilibrium. Before considering the results, it is worth asking what can possibly be learned from this exercise. The authors themselves pose the question

Since the world's international economies are vastly more complicated than the economies created for this study, of what relevance are laboratory data? (NPR, pg. 462)

The author's don't satisfactorily answer this query. Their best attempt at an answer is

restricted models.

The preconditions for the operations of the principles [of comparative advantage] have been introduced by the experimenters. The experiments are able to provide some insights into how models [...] are able to organize the data, given that the situation is one in which the model can meaningfully be applied. The experiment cannot, however, answer the equally important questions about the relative likelihood that nature has created a situation for which the parametric and institutional features of the model are relevant (NPR, pg. 464).

Despite the poor writing, this passage illuminates what the contribution of the paper is. The paper is really about how markets work, about whether they can effectively exploit all the gains from trade that we know are there. It is hard to see how any experimental result could affect our view of the relevance of comparative advantage. The authors' results show that production, trade, and prices converge to the correct values, and they argue that the process of convergence is informative about how markets work. In particular, the paper has an extensive discussion of dynamics and the process of convergence to the full general equilibrium. This may be of interest to theorists but has little relevance for the applicability of the theory of comparative advantage.

In contrast to Nussair et al, Bernhofen and Brown (2000) provide actual historical evidence on the relationship between autarky prices and trade through an examination of Japan's opening to trade in the 1860s. This is a well-executed paper in several ways. First, the authors correctly apply Deardorff's (1980) general statement of the theory of comparative advantage. Second, they argue carefully and (to this non-expert) convincingly that Japan in the mid-19th century met the requirements needed to apply the theory: Japan was completely closed to trade before 1854, and had fairly free trade (in particular, no export subsidies) by the late 1860s. They also recognize the biggest potential problem with applying the comparative advantage prediction to this episode: prices within Japan might

have changed between 1854 and 1870 even in the absence of the opening to trade. The requirements for using 1854 prices in a test of comparative advantage are that economic growth was unrelated to trade, and that growth was not biased in favor of exportables². They make a plausible historical case that these conditions held true. A short table confirms that the Deardorff condition is satisfied: Japan's trade was correctly predicted by her autarky prices³.

Bernhofen and Brown argue that, besides being closed to trade before 1854 and having no export subsidies after opening up, Japan's economy was fairly competitive before and after opening to trade. Does this mean that their results were foreordained? No. As Nussair et al emphasize, just because the competitive and technological conditions of an economy satisfy the assumptions of the theory doesn't mean that the post-trade equilibrium will satisfy the law of comparative advantage - after all, humans and their institutions work in mysterious ways. Put differently, if the data had violated the law of comparative advantage, would it have shaken our faith in the theory? Yes: given the evidence that the authors provide about the structure of the Japanese economy in the mid-nineteenth century, Japan's trade should have been predicted by autarky prices, and if it hadn't we would have had to explain why. The fact that the post-trade general equilibrium behaved as expected is genuine news, and is evidence for the relevance of the theory of comparative advantage. The result is particularly interesting because it involves a large country which became one of the greatest trading nations ever.

In short, the standard view that the theory of comparative advantage has never been tested needs to be modified: with Bernhofen and Brown's contribution,

² A further condition which they do not mention is that tastes must have stayed the same.

³ The paper chooses a normalization for prices such that the magnitude by which the Deardorff condition is satisfied is uninformative. The authors tell me that the next version of the paper will express this magnitude relative to autarky GDP, giving a measure of the size of the gains from trade.

we now have one paper that tests a very general version of comparative advantage, and comparative advantage passes. I think I can speak for many economists who have taught this theory with great fervor when I say "thank goodness".

2 Testing Classical and Neoclassical Models

Bernhofen and Brown is the exception to the rule: for all other contemporary and historical trading economies, we have no evidence on autarky prices. As a consequence, any application of comparative advantage theory must have an intermediate step between autarky prices and trade, one that relates autarky prices to observable features of economies. This means that empirical researchers must usually model production possibilities and preferences. In this section I discuss recent empirical work on these models, but first I will lay out a general model of comparative advantage that can be used to frame the discussion.

2.1 The Neoclassical Theory Of Production and Trade

Comparative advantage is a property of what I will call neoclassical trade models. These models all have at least two goods (so that there is a potential motive for trade), factors which are mobile between alternative uses, convex technology, and perfectly competitive markets for goods and factors. The equilibrium conditions for such economies include zero profit conditions for each sector:

$$p_g^c = a_g^c(w^c) \quad g = 1, \dots, G, \quad w^c \in \mathbb{R}^F \quad (1)$$

where p_g^c is the producer price of good g in country c , and $a_g^c(w^c)$ is the unit cost function for good g given the technology and factor prices w^c that prevail in country c . Constant returns to scale implies that the unit cost functions can be rewritten as

$$a_g^c(w^c) = a_{1g}^c(w^c) \cdot w_1^c + a_{2g}^c(w^c) \cdot w_2^c + \dots + a_{Fg}^c(w^c) \cdot w_F^c \quad (2)$$

where each a_{fg}^c is the cost-minimizing amount of factor f used to produce one unit of good g , which depends on w^c . Collecting all the zero profit conditions together

we can compactly write the system as

$$p^c = (A^c)' w^c \quad p^c \in \mathbb{R}^G, \quad A^c \in \mathbb{R}^F \times \mathbb{R}^G \quad (3)$$

where the $F \times G$ matrix of cost-minimizing input coefficients A^c depends on w^c . The other production side equilibrium conditions are that all factors v , which are in fixed aggregate supply, are fully employed. For a given factor f in country c , full employment is written as

$$v_f^c = a_{f1}^c x_1^c + a_{f2}^c x_2^c + \dots + a_{fG}^c x_G^c \quad f = 1, \dots, F \quad (4)$$

where x_g^c is output of good g , and the dependence of the a_{fg}^c on w is implicit. Collecting all F full-employment conditions together gives

$$v^c = A^c x^c \quad v^c \in \mathbb{R}^F \quad (5)$$

The system given by (3) and (5) is $F+G$ equations in the $F+2G$ unknown factor prices, output levels, and output prices.

Even at this level of generality, and before specifying G extra equations required to close the model, we can say something interesting about these economies. First, if there are at least as many goods as factors, $G \geq F$, then it is possible to solve the zero-profit conditions in (3) for factor prices as a function solely of goods prices⁴:

$$w^c = w^c(p^c) \quad (6)$$

This result, labeled “factor price insensitivity” or FPI by Leamer (1995), is remarkable: factor prices do not depend directly on factor supplies, and if country c is small (so that producer prices are determined in world markets that are unaffected by the output of country c), then factor prices are completely

⁴ If $G > F$, then any F equations from (3) can be used to solve for the F factor prices; the other $G-F$ equations will be consistent by assumption.

independent of domestic factor supplies⁵. Note also that FPI implies that the equilibrium technique matrix A^c will be independent of factor supplies if $G \geq F$, since unit costs depend only on factor prices.

We can also say some interesting things about the relationship between factor supplies and outputs. First, if there are *exactly* as many goods as factors, then A^c is both independent of factor supplies and square, which means it can be inverted. Premultiplying both sides of (5) by this inverse gives

$$x^c = (A^c)^{-1} v^c, \quad (7)$$

which is to say that industry outputs are a locally linear function of economy-wide factor supplies. With more goods than factors, $G > F$, A^c is not invertible: there are many output vectors which satisfy full employment, and which one will obtain in equilibrium will depend on goods market equilibrium conditions. If $G < F$, you might be tempted to think that you could take any G equations from (5) and solve for outputs without reference to the zero-profit conditions (6); the error in that thinking is that with $G < F$ the equilibrium techniques are not independent of factor supplies. In the $G < F$ case, outputs are determinate, but they can't be solved for independently of the zero profit conditions.

Closing the model under autarky requires G goods-market equilibrium conditions, while with trade the G prices are given by global market clearing. With national income Y^c a function of producer prices and factor supplies,

$$Y^c(p^c, v^c) = w^c(p^c, v^c) \cdot v^c, \quad (9)$$

we can define the national indirect utility function $U^c(\tilde{p}^c, Y^c)$, where

⁵ For many decades until Leamer coined the term, the FPI result did not have its own name, and the result was often misleadingly referred to as “factor price equalization” or FPE. The terminology matters in this case, because true FPE - that is, the same factor prices in different countries - requires more assumptions (including frictionless trade and identical technology) than FPI, which is a property of any single economy. Succinctly, FPI is a necessary but not sufficient condition for FPE.

$\tilde{p}^c = \tilde{p}^c(p^c, \tau^c)$ is the vector of consumer prices as a function of producer prices and trade policy instruments τ^c ⁶. With the normalization that the marginal utility of income is unity, Roy's identity gives the Marshallian demands $e_g^c(\tilde{p}^c, Y^c)$ as the negative of the marginal indirect utilities:

$$e^c(\tilde{p}^c, Y^c) = -\nabla_p U^c(\tilde{p}^c, Y^c), \quad e^c \in \mathbb{R}^G \quad (10)$$

With the consumption and production sides of the model specified, it is trivial to write down the net export vector t^c as the difference between the two:

$$t^c(p^c, \tilde{p}^c, v^c) = x^c(p^c, v^c) - e^c(\tilde{p}^c, Y^c(v^c, p^c)) \quad (11)$$

The fact that the determination of net exports can be separated into the determination of production and consumption is very far from being a trivial result, and does *not* generalize to most models with increasing returns and/or imperfect competition: under most such models, output and consumption must be determined jointly (see Helpman and Krugman (1985) and Fujita, Krugman, and Venables (1999)).

The development of the production model of equations (3) and (5) uses primal cost functions, and is useful for understanding the properties of the resulting equilibrium. But if all one is interested in is the equilibrium outputs and factor prices, then the model can be stated much more compactly using duality⁷. Under constant returns and perfect competition, national income is given by

⁶ I am deliberately noncommittal here about how trade policy causes differences between consumer and producer prices. I also ignore the value of trade policy revenue (tariff revenue plus quota rents) in national income, and dismiss the possibility of aggregate trade imbalances, to keep the notation simple.

⁷ Woodland (1982) offers an especially clear and detailed development of the dual approach. For a more compact, if opaque, treatment see Dixit and Norman (1980).

$$Y^c = r^c(p^c, v^c) = \max_{x^c} \{p^c \cdot x^c \mid x^c \in Y^c(v^c)\} \quad (12)$$

where $Y^c(v^c)$ is the compact set of feasible net outputs. Equilibrium outputs and factor prices are then given by

$$x^c \in \nabla_p r^c(p^c, v^c) \quad (13)$$

$$w^c = \nabla_v r^c(p^c, v^c) \quad (14)$$

Note that the gradient in (13) is a set, reflecting the indeterminacy in general of the output vector which will maximize national income. As indicated above, this indeterminacy disappears if $G \leq F$ with no joint production, and we can differentiate the output vector with respect to factor supplies to get

$$x^c = \left[\nabla_{pv}^2 r^c(p^c, v^c) \right] \cdot v^c = [R^c(p^c, v^c)] \cdot v^c \quad (15)$$

If $G=F$ and there is no joint production, then the square matrix $R^c = (A^c)^{-1}$, it is locally independent of factor supplies, and we just have a restatement of equation (7). The notation R^c is chosen to evoke the Rybczynski theorem, since the elements of R^c give the general equilibrium response of outputs to factor supplies.

What can be said in general about R^c ? First, except in the case of $G=F=2$ and no joint production, there is no necessary connection between a sector's factor intensity and its output response to a factor supply increase; for example, it is possible in general that the most capital intensive sector will shrink when capital becomes more abundant. A corollary is that it is impossible to generalize the Heckscher-Ohlin theorem beyond the 2×2 case: a country will not necessarily export goods that intensively use their abundant factors, even if every other condition of the theorem is satisfied⁸.

⁸ Although the Heckscher-Ohlin theorem doesn't generalize, the result that countries will export the services of their abundant factors does generalize to

Second, for $G > F$, we can't say anything about R^c because it doesn't exist due to the indeterminacy of output. This is alarming from an empirical point of view, but many theorists have argued that this case is irrelevant, since even tiny price changes will result in $G-F$ industries shutting down, leaving the economy with $G=F$ ⁹. This is a cute theoretical argument, but does not settle the case empirically.

Third, some important results are available for the general $G \leq F$ case. A natural extension of the 2×2 production structure is to suppose that every sector uses at least two factors, and that there is no joint production. As Jones and Scheinkman (1977) show, with these assumptions and $G=F$, then when a factor supply rises, at least one industry will expand more rapidly than the factor, and at least one industry will contract. To state the result more colorfully, in this "even" case every factor is a friend to at least one industry and an enemy to another. In the uneven case of $G < F$ this result does not hold: an increase in a factor supply may cause all outputs to rise; a factor may be so good-natured that it is a friend to all and an enemy to none¹⁰. But the converse does not hold: for every industry, there is at least one factor whose accumulation will cause it to decline. That is, every industry has an enemy.

This is a rich set of important and testable empirical predictions. If every industry has an enemy, there are clear political economy implications: there will

higher dimensions. See the chapter by Davis and Weinstein for a discussion of the research on this so-called Heckscher-Ohlin-Vanek hypothesis.

⁹ The easiest way to understand this argument is to visualize the textbook two-good, one factor Ricardian model. If relative prices equal relative labor productivities, both goods will be produced in indeterminate quantities, and increases in labor will have unpredictable effects on outputs. But if prices change even slightly, the economy will specialize completely in the good whose relative price has risen, and there will be a unique relationship between labor supply changes and output changes.

¹⁰ The simplest example comes from the textbook specific factors model: accumulation of the mobile factor causes both industries to expand.

be political opponents to policies favoring the accumulation of some factors. The opposite can be said for factors which are friends to all: such factors are likely to face less opposition in the quest for favorable treatment. The effects of factor accumulation are also of interest to policymakers wishing to forecast the future sectoral composition of output. Empirically, there are several points to emphasize. First, the identification of friends and enemies can not be accomplished just by looking at input cost shares. Second, the assumption of no joint production is crucial: with joint production the magnification results in even models do not hold, and the strong friends and enemies results can not be expected. Third, very little can be said if $G > F$: there may be no systematic relationship between relative factor supplies and the composition of output.

Most of the above discussion has concerned equilibrium in a single economy. Without putting further restrictions on how technology and tastes vary across countries, the model can not say much about trade or international differences in outputs and factor prices. The generalized Heckscher-Ohlin approach (what I will call the factor-proportions model) to making cross-country predictions is to assume away *all* international differences in technology and preferences. Further assuming that preferences are homothetic (so that consumption shares don't depend on income), the equation (11) prediction for trade simplifies to

$$t^c(p^c, \tilde{p}^c, v^c) = \nabla_p r(p^c, v^c) - \tilde{e}(\tilde{p}^c) \cdot Y^c \quad (16)$$

where $\tilde{e} \in \mathbb{R}^G$ is a vector of consumption budget shares. In (16), the output and consumption functions are the same across countries (the c superscripts have disappeared from $r(\cdot)$ and $e(\cdot)$), and so could in principal be estimated with cross country data. The prediction can be simplified further by assuming frictionless trade (so that $p^c = \tilde{p}^c = p$), $G=F$, and that endowments are sufficiently close

together so that all countries produce the same goods¹¹. By equation (6), this means that factor prices and hence production techniques will be the same across countries. By equation (7), outputs will vary linearly with factor supplies. Considering (16) as a cross-section prediction at a point in time means that there is no variation in prices, so we can write the prediction for a country's trade as

$$t^c(v^c, Y^c) = R \cdot v^c - \tilde{e} \cdot Y^c \quad (17)$$

In words, trade depends only on relative factor supplies and country size.

This is an elegant prediction, and a version of it was investigated empirically in Edward Leamer's landmark 1984 book (Leamer, 1984). But the elegance of (17) comes at a high price in terms of empirically dubious assumptions, and much of the empirical work on the neoclassical trade model during the past decade has been aimed at relaxing some of these assumptions without giving up the ability to make cross-country inferences.

The assumption of identical homothetic preferences (IHP) used to derive (17) is implausible, and uninteresting in the sense that there is no real theory behind it. Rather, the IHP assumption is just an analytical simplification used to translate the Rybczynski relationship (the mapping from endowments into outputs) into the Heckscher-Ohlin relationship (the mapping from endowments into net exports). An empirical rejection of IHP would rightly have no impact on our view of the underlying production model, although it could be interesting for other reasons¹².

Another reason to be uninterested in IHP is that it treats all demand for

¹¹ Together these assumptions are sufficient for trade to reproduce an “integrated equilibrium” with determinate production. The integrated equilibrium is the allocation which would result in a world with no barriers to the movement of goods or factors. See Dixit and Norman (1980).

¹² In a series of well-done papers, Markusen (1986), Hunter and Markusen (1988), and Hunter (1991) explore the role of non-homothetic preferences in explaining gross trade volumes.

traded goods as coming directly from the demand for final consumption goods. Intermediate products are readily introduced into the production model and have no implications for the production model results mentioned above, as long as we make the distinction between net and gross outputs. But allowing for trade in intermediate goods means that there will be no simple relationship between net exports and national income, even if IHP holds. Since a very large share of the volume of trade is intermediate goods, and essentially all imports require some domestic value added before they enter final consumption (see Rousslang and To (1993)), a more plausible simple model for trade would be that all trade is in intermediate goods, rather than none as assumed by the models that yield equations (11), (16) and (17).

2.2 When worlds collide: data meets the neoclassical model

Whatever assumptions are made about the demand for traded goods, no cross-country predictions can be made without taking a stand on how technology and prices vary across countries. The Leamer (1984) assumptions of identical technology, frictionless trade, and $G=F$ lead to a very simple prediction: for each industry at a point in time, output depends linearly on factor supplies:

$$x_g^c = \sum_{f=1}^F r_{fg} v_f^c \tag{18}$$

This is the model estimated by Harrigan (1995). The empirical model in that paper considers ten large manufacturing sectors and four factor supplies: capital, skilled and unskilled labor, and land. For each industry, Harrigan analyzes a panel of 20 OECD countries and 16 years, using three different strategies for pooling over time: generalized least squares with and without fixed country effects, and a time-varying parameter model. Even the fixed effects model has substantial residual autocorrelation, while the non-fixed-effects models all have first-order autoregressive parameters in excess of 0.9. This means that the parameters of the

model are identified mainly from time series within-country variation - an unfortunate fact since the main interest is in explaining the cross-country distribution of production.

A striking result from Harrigan (1995) is that every industry was found to have at least one enemy, a factor whose abundance and/or accumulation leads to a decline in output (the enemy is usually skilled or unskilled labor, sometimes land). Capital is manufacturing's friend: it is estimated to have a positive effect on output for all ten industries in each specification. Despite the fact that Leamer (1984) used a cross-section of trade data, in contrast to a panel of output data, the inferences about comparative advantage are similar.

Although factor supplies are jointly statistically significant in each regression, the model does poorly in explaining the cross-sectional variation in the data, with large within-sample prediction errors. Harrigan identifies a number of potential explanations for this poor fit (including bad data, scale economies, and government policy) but doesn't mention other possibilities (such as indeterminacy in production, technology differences, different product mixes, or price differences caused by trade policy or transport costs). It seems fair to say that the factor proportions view of the world has mixed support from this paper: a poor overall fit, but fairly solid evidence of a systematic relationship between outputs and relative endowments.

Several papers have been at least partially motivated by the mixed results of Harrigan (1995). Bernstein and Weinstein (1998) focus on the question of output indeterminacy when $G > F$. They correctly note that this is an empirical question which has little to do with counting the numbers of goods and factors in any particular empirical exercise. They begin by noting that, with identical technology, frictionless trade, endowments which are not too far apart, and $G \geq F$, then the full employment conditions (5) will have the same A matrix for all c :

$$v^c = Ax^c \tag{19}$$

They call this the Heckscher-Ohlin-Vanek or HOV model. If, in addition, G is

exactly equal to F , then outputs are the same linear function R of endowments for all c , and as noted above $R = A^{-1}$. They express this implication as

$$AR = I_F \tag{20}$$

where I_F is the identity matrix of dimension F .

Using data from Japanese regions, they confirm that (19) holds, which indicates that techniques and factor prices are the same in all Japanese regions. This is not a trivial finding: it rules out increasing returns at the level of industries, and/or technological differences across regions. They also argue that it rules out $G < F$, but that is wrong: with intra-Japan mobility of factors, factor prices and techniques will be equalized regardless of the relative number of goods and factors. Despite the fact that (19) holds for Japanese regions, (20) fails miserably: outputs are not well-explained statistically by endowments alone, and the linear restrictions embodied in (20) are rejected. From this they conclude that $G > F$ and output indeterminacy is an empirically important fact about general equilibrium production.

They then apply (19) to international data, multiplying the Japanese A matrix by national output vectors x to get predicted national endowments v :

$$\text{predicted } v^c = A^{Japan} x^c$$

These predicted endowments are not at all close to actual measured endowments, which leads them to reject the assumption that all countries produce the same goods using the same techniques. This result, while not new (see the chapter by Davis and Weinstein in this volume for more evidence that techniques vary internationally), is nonetheless worth noting, since it suggests that economists should abandon the simple HOV model of international production.

Abandoning HOV is one thing, but replacing it with something else is another. One of the most appealing aspects of empirical work based on the even model pioneered by Leamer (1984) is that every parameter estimated has a clear structural interpretation. The challenge for researchers wishing to improve on this framework is to develop empirically implementable models which are equally

closely tied to theory, but that relax the stringent assumptions used to derive equations like (17) and (18). In the factor content literature, this has been accomplished using restrictive models of international technology differences (Trefler 1993, 1995) or two-factor models where factor price equalization fails (Davis and Weinstein 1998)¹³. These models may or may not be appropriate for studies of the factor content of trade, but they are too restrictive for studying comparative advantage, as they rule out all but very special types of cross-country technology differences. In searching for amendments to the factor proportions model, it is natural to consider general technology differences as a source of comparative advantage, not least because there is extensive evidence that, even among advanced economies, technology differences are large, ubiquitous, and non-transitory (see, inter alia, Jorgenson, Kuroda, and Nishimizu (1987), Jorgenson and Kuroda (1990), Dollar, Baumol and Wolff (1988), Dollar and Wolff (1993), van Ark (1993), van Ark and Pilat (1993), and Harrigan (1997b, 1999)).

Harrigan (1997a) points out that using the dual, rather than the primal, representation of aggregate technology makes it possible to estimate more general models of specialization. Harrigan assumes that technological differences across countries are Hicks-neutral and industry specific. This can be incorporated into the revenue function approach in a very straightforward way:

$$r^c(p^c, v^c) = r(\theta^c p^c, v^c), \quad \theta^c = \text{diag} \{ \theta_1^c, \dots, \theta_G^c \} \quad (21)$$

where θ_g^c is a scalar productivity parameter which gives the level of technology in industry g of country c relative to productivity in a base country. This is a natural extension of the classical one-factor Ricardian model, and it has the virtue that the technology parameters are, in principle, measurable by applying the theory of total factor productivity (TFP) measurement. The usual derivative property applies to

¹³ See the chapter by Davis and Weinstein in this volume for details on these modeling strategies.

(21), so that outputs are given by the gradient of (21) with respect to prices:

$$x^c = \nabla_p r(\theta^c p^c, v^c) \quad (22)$$

Note that if there are no cross-country differences in relative industry technology levels, that is $\theta_g^c = \theta^c \forall g$, then (by the homogeneity of the revenue function) technology differences become a scalar shift parameter, giving outputs as

$$x^c(p^c, v^c) = \theta^c x(p^c, v^c) \quad (23)$$

Equation (23) illustrates that technology differences which are neutral across sectors affect absolute, but not comparative, advantage¹⁴.

To implement the model given by (22), Harrigan (1997a) follows Kohli (1991) and assumes that the revenue function (21) can be adequately approximated by a translog functional form. This strategy leads to the following estimation equation:

$$s_{gt}^c = \sum_{k=1}^G a_{kg} \ln \theta_{gt}^c + \sum_{i=1}^F r_{if} \ln v_{ft}^c + \varepsilon_{gt}^c \quad (24)$$

where s_{gt}^c is the share of good g value added in country c 's GDP at time t , the a 's and r 's are parameters to be estimated, and ε_{gt}^c has a panel structure with fixed country and time effects to account for other unobservable influences on specialization. There are a number of notable features of this specification. First, it allows Harrigan to simultaneously estimate the impact of Ricardian and Heckscher-Ohlin influences on specialization. Second, since the same technology parameters appear in each equation, it is possible to calculate cross-TFP effects on output shares, which is a key general equilibrium channel. Third, the estimated results do not directly tell us whether each sector has an enemy, in the sense of a

¹⁴ This is the model preferred by Trefler (1995) in his study of the factor content of trade.

factor that causes its output to decline in absolute terms, although the r 's tell us which factors raise or lower a good's share of national income¹⁵. Fourth, because of the use of country fixed effects, all of the model's parameters are identified by within-country time series variation. Fifth, the specification of the model requires no assumption about any form of factor price equalization - this is one of the benefits of using a dual rather than primal approach.

Harrigan's results support the view that non-neutral technology differences are important for specialization. For most sectors, the own-TFP effects are positive, statistically significant, and large. The largest effect is in the biggest sector, Machinery: a 10 percent improvement in relative Machinery TFP raises that sector's share of GDP by around 0.25 percentage points. As an example of the cross-TFP effects, technological progress in Machinery comes at the expense of the Chemicals and Metals sectors, whose share of GDP declines. The inferences about factor supplies are roughly consistent with Harrigan (1995) and Leamer (1984): accumulation of producer durables and High-School educated workers generally lead to expanding manufacturing sectors, while growth in structures and highly educated workers are associated with declining manufacturing. These findings suggest a simple story: the service sector is intensive in non-residential construction (office buildings and retail stores) and college-educated workers (managers, professionals, educators), so that abundance in these factors draws other resources out of manufacturing and into the service sector. By contrast, the manufacturing sectors are intensive in producer durables and medium-educated workers, so that abundance in these factors draws resources out of services and into manufacturing sectors. While plausible, confirmation of this explanation would require data on direct factor shares which are not easily available in internationally comparable form.

Harrigan (1997a) fruitfully extends the literature on comparative advantage

¹⁵It is straightforward to compute the effects on levels, rather than shares, of output, but Harrigan does not do this.

in one direction, by abandoning the Heckscher-Ohlin assumption that countries share the same technology. Another problematic feature of the standard factor proportions approach is the assumption that all countries produce the same goods and have the same factor prices. This "one cone" assumption is explicit in Leamer (1984), Harrigan (1995), and Bernstein and Weinstein (1998), as well as in most of the factor content literature (with the notable exception of Davis and Weinstein 1998). Absolute factor price equalization is easy to reject by direct observation: it would be hard to explain mass migration from the South to the North if wages were the same everywhere. In the factor content literature, the assumption of equal factor prices in levels is sometimes replaced with the weaker assumption that relative factor prices are equalized (see Trefler 1995). Equal relative factor prices implies that, for a given sector across countries, input coefficients are constant and in particular do not vary with aggregate endowments (this is just a corollary of factor price insensitivity). This hypothesis can be tested by a simple non-structural cross-section regression pooled across goods g and countries c :

$$\frac{a_{Kg}^c}{a_{Lg}^c} = \beta_g + \beta \frac{K^c}{L^c} \quad (25)$$

where the left-hand side is the capital-labor ratio in industry g in country c , which is regressed on an industry constant and country c 's aggregate capital-labor ratio. Under the one-cone/FPI hypothesis, $\beta = 0$. There is ample evidence that this is not the case: Dollar, Wolff, and Baumol (1988, Table 2.3) show that capital per worker in individual industries is highly correlated with capital per worker in aggregate manufacturing. More recently, Davis and Weinstein (1998, Table 1) show the same thing, finding that techniques are strongly correlated with aggregate endowments.

A finding that $\beta > 0$ in equation (25) can be explained by a failure of factor price equalization (FPE). Retaining the Heckscher-Ohlin assumptions of

frictionless trade, perfect competition, and identical technology across countries, FPE fails when countries have endowments which are too far apart, and as a result countries produce different goods, the so-called "multi-cone" equilibrium. The fact that we do not observe such specialization in the output statistics may simply be because different goods are lumped together into the same industrial classification. If this were true, then the sign of the Rybczynski effects of endowments on observed output aggregates would differ systematically across countries, and one-cone empirical models like Harrigan (1995) would be misspecified.

By using a dual approach, Harrigan (1997a) skirts this issue, but the general spirit of that paper is a one-cone model, since the translog approximation is assumed to be valid at all points in the sample. Schott (2000) tackles the multi-cone issue directly, and develops an empirical model where the set of produced goods and the associated Rybczynski effects depend on relative factor supplies. Schott's theoretical model is a standard one of two factors (capital and labor), many goods whose techniques of production are independent of factor prices, and many countries¹⁶. The equilibrium of this model has every country producing just two goods, one more and one less capital intensive than the country's aggregate endowment. Define

$$k_g = \frac{a_{Kg}}{a_{Lg}} \quad \text{and} \quad k^c = \frac{K^c}{L^c}$$

to be the fixed capital-labor ratio of good g and the capital-labor endowment of country c respectively, and number goods in order of increasing capital intensity,

$$k_1 < k_2 < \dots < k_G .$$

¹⁶ Schott's model in a more general form dates back at least to Deardorff (1979).

Then there are three possible linear relationships between the output of a good and a country's endowments:

$$\begin{aligned}
 x_g^c &= r_{1Kg} K^c - r_{1Lg} L^c & \text{if } & k^c \in [k_{g-1}, k_g) \\
 x_g^c &= -r_{2Kg} K^c + r_{2Lg} L^c & \text{if } & k^c \in [k_g, k_{g+1}) \\
 x_g^c &= 0 & \text{otherwise,} &
 \end{aligned} \tag{26}$$

where the r 's are positive constants. In words, if a country's capital-labor ratio lies between k_{g-1} and k_g , then capital accumulation leads to an increase in the output of good g , and the opposite if k^c lies between k_g and k_{g+1} . If k^c does not lie between k_{g-1} and k_{g+1} , then country c will not produce any of good g at all; it will be producing the two goods closest to its aggregate capital-labor ratio instead.

Now imagine that a particular industrial classification includes two or more goods with quite different capital intensities (for example, Textiles includes low-quality cotton cloth as well as high-tech synthetic fibers). With such aggregation of goods within a single classification, the Rybczynski effect of capital on output of the aggregate may switch sign more than once: at very low k^c the Rybczynski effect is positive, then it becomes negative as countries move out of the labor intensive good, then positive again as production of the capital intensive good commences, etc. In this case trying to infer the effect of endowments on outputs by pooling across countries with very different endowments is a hopeless muddle: any estimated slopes will be a mix of effects of varying size and sign, and will have no structural interpretation.

Schott pursues two strategies for dealing with the complexities of multiple cones and multiple goods aggregated into a single category. The first is to take the existing output aggregates and estimate a piece-wise linear relationship between outputs and endowments. In this model, Schott simultaneously estimates the Rybczynski effects along with the capital-labor ratios at which the effects change.

Schott uncovers four cones, which is to say that for each output aggregate, four Rybczynski effects are estimated at different point in the sample. This is an intricate empirical model and in the end is not too convincing, as the fitted and actual output levels are not at all close to each other, and the points at which the slopes change sign seem heavily influenced by a very small number of observations (see Schott (2000), Figure 3).

Schott's second empirical strategy is more promising. Rather than work with the usual output aggregates, he constructs three of his own aggregates based on the capital intensity of each country-industry observation. Since (as $\beta > 0$ in equation (25) verifies) sectoral capital intensities are correlated with aggregate capital abundance, Schott's "Heckscher-Ohlin aggregates" group together products in different output categories which are produced by countries with similar capital abundance. For example, Apparel produced in Guatemala is lumped together with Electrical Machinery produced in the Phillipines, while Swedish Apparel is lumped together with Transport Equipment produced in Malaysia.

Using these three HO aggregates and estimating a version of equation (26), Schott finds what the multi-cone reasoning above suggests: the effect of capital accumulation on the least-capital intensive aggregate is at first negative then zero, the effect is first positive and then negative for the middle aggregate, and is first zero and then positive for the third aggregate. In other words, he identifies two cones: in the first cone, countries produce the Low and Middle capital-intensive HO aggregates, and in the second cone countries produce the Middle and High capital intensive HO aggregates. There is some circularity in this procedure, since country capital abundance is in effect used to construct the HO aggregates, making it unsurprising (for example) that capital abundant countries specialize in the capital intensive aggregate. Schott's results are also suspect because he uses total capital within manufacturing, rather than aggregate capital, to measure a country's overall capital-labor ratio; this means he is ignoring capital re-allocation between manufacturing and the rest of the economy. A broader criticism is that Schott's

theoretical model is quite special and is taken perhaps too literally as a framework for data analysis; Leamer (1987), in contrast, also works with a multiple-cone model but regards the piece-wise linearity between outputs and endowments implied by the model to be too special to take seriously. Schott also completely ignores the issue of technology differences, which Harrigan (1997) showed to be important for specialization. Despite these caveats, Schott's study is important for two reasons: it provides some evidence that multiple cones are empirically important, and it forces us to think seriously about the heterogeneity lurking within measured aggregate outputs.

Harrigan and Zakrajšek (2000) consider the multiple-cone issue as well as several other open questions about specialization. As noted above, Harrigan (1997a) measures productivity differences using TFP indices (which, because of data availability, restricts the sample to OECD countries), and estimates a fixed-effects model of specialization that does not use any of the cross-country variation in the sample. Motivated by these limitations of Harrigan (1997a), Harrigan and Zakrajšek develop an empirical model which permits consistent estimation of the effects of factor endowments on specialization while allowing for unobservable technology differences. This allows them to analyze a larger number of countries (including a few from Latin America and East Asia), and to exploit the cross-section variation in the data. Their identifying assumption is that, except for country and time effects, any non-neutrality in technology differences is orthogonal to factor supplies. As in Harrigan (1997a), they adopt a dual-translog approach which allows them to avoid making any assumptions about factor price equalization, and which leads to an estimating equation which is a simplification of (24):

$$s_{gt}^c = \beta_g^c + \sum_{i=1}^F r_{if} \ln v_{ft}^c + \varepsilon_{gt}^c \quad (27)$$

If the country effects β_g^c are also assumed to be orthogonal to factor supplies, then

it is possible to use a random-effects estimator which combines the time-series and cross-country variation in the sample. They also report fixed effects estimates (which use only the time-series variation) and between estimates (which use only cross-country variation) of (27).

Unlike most other papers in this literature, Harrigan and Zakrajšek also consider alternative hypotheses. The statistical alternative they consider is simple, that specialization depends on aggregate productivity rather than on relative endowments:

$$s_{gt}^c = \beta_g^c + \beta_1 \ln \theta_t^c + \beta_2 \left[\ln \theta_t^c \right]^2 + \varepsilon_{gt}^c \quad (28)$$

where θ_t^c is measured as real GDP per worker. This reduced form relationship can be loosely derived from a product cycle model, where new goods are first produced in rich countries and are later produced in poorer countries as technology is transferred. It can also capture multiple cone effects in a flexible way: models such as Schott's would predict an important role for the second-order term as countries move into and out of goods based on their overall per-capita income.

Harrigan and Zakrajšek find that estimating (27) gives a noisy but fairly consistent story about industrial specialization: human and physical capital abundance raise output in the heavy industrial sectors, while physical capital lowers output in food and apparel-textiles. The model has little success in explaining variation in output in the smaller, more resource-based sectors, probably because they have no measurements of resource abundance. Turning to the alternative model (28), results are roughly in line with what would have been expected from the factor proportions results: higher aggregate productivity is associated with lower output of food and higher output in the heavy industrial sectors (fabricated metals and the three machinery categories).

What about multiple cones? Following Schott's line of reasoning, this should show up in parameter instability across different regions of relative factor

supply space. Harrigan and Zakrajšek used a number of formal and informal strategies to find evidence of such instability and found nothing. They did find some weak evidence of quadratic effects in equation (28), but the nonlinearity was only economically important for a single sector, Food.

The bottom line from Harrigan and Zakrajšek (2000) is consistent with the message which has been developed in all of the papers reviewed in this section, as well as the factor content literature reviewed in this volume by Davis and Weinstein:

Relative factor endowments have a large influence on specialization, in ways that are consistent with theory and stylized facts about the international economy. However, factor endowments leave much that is unexplained: there is a great degree of country-specific idiosyncrasy in specialization patterns, and there is also a great deal of noise (Harrigan and Zakrajšek (2000), page 23.)

2.3 What about Ricardo?

The papers discussed in the previous section all work with variants of multi-factor models that have roots in the Heckscher-Ohlin tradition. This might seem odd to a reader familiar only with a much earlier literature on testing trade models, which concluded that Heckscher-Ohlin did very poorly (Leontief 1954) while Ricardo did quite well (MacDougall 1951, 1952). Empirical research on the static Ricardian model was quiescent for nearly three decades after Balassa's last word on the subject (Balassa 1963). In the last few years, however, there have been a few papers on the Ricardian model, including the innovative work by Eaton and Kortum (2001) which is discussed in the context of the gravity equation below. Here I will discuss two recent papers that are very much in the spirit of MacDougall and Balassa.

Golub and Hsieh (2000) argue that a focus on labor productivity variation as the source of comparative advantage is appropriate because other factors of production (such as capital and raw materials) are internationally mobile. This is

an oft-heard argument which has at least two problems with it. First, it risks confusing free trade in financial assets with easy mobility of physical capital goods. What is relevant for comparative advantage is how easily productive factors are reallocated across alternative uses; the ownership of factors is relevant to the level of national income but not to the composition of national product. Despite the vast and rapid international flows of financial capital, I know of no evidence which suggests that physical capital is as easily reallocated internationally as it is intranationally. To make my point transparent, observe that structures are an important component of the capital stock which are immobile even within countries, never mind across borders, while ownership of structures can easily be transferred internationally. The same point can be made with reference to natural resource stocks: land can not move, but its ownership can. The point is not to insist that physical capital and land *are* sources of comparative advantage, but to insist that it is an empirical question. Furthermore, the evidence reviewed in the preceding section suggests that non-labor endowments are relevant to specialization.

These objections notwithstanding, a case can be made that the Ricardian concept of output per worker may be what matters for comparative advantage. First, labor's share of manufacturing value added is quite high, meaning that labor productivity is closely linked to total factor productivity. Second, if differences in labor productivity reflect technological differences, then labor productivity will be a good predictor of specialization. This is the hypothesis that Golub and Hsieh investigate.

Golub and Hsieh do not derive their specification directly from a well-specified model, appealing instead to the earlier literature. This makes it impossible to interpret their results in a structural fashion, but the general idea is intuitive: if relative productivity in sector g in country c is higher than it is in the average sector in country c , then c specializes in g . An illustrative equation is

$$\log \frac{X_{gb}}{X_{gc}} = \alpha_{bc} + \beta \log \frac{a_{gb}}{a_{gc}} + \varepsilon_{gbc} \quad g = 1, \dots, G \quad (29)$$

where X_{gc} is a measure of export success in good g for country c , a_{gc} is output per worker in sector g in country c , ε_{gbc} is a residual, and α and β are parameters. The equation is pooled across goods g (and possibly across country pairs), and a positive value of β is taken as confirmation of the hypothesis that labor productivity determines comparative advantage.

The problem with specifications such as (29) is not just the usual one of a missing alternative hypothesis, which plagues most work on comparative advantage. Rather, the problem is that there is nothing inherently general equilibrium about the specification. Equation (29) simply says that productivity advantage in a sector is associated with export success in that sector, a prediction that arises from any number of partial-equilibrium supply and demand models (for example, a simple Cournot reciprocal dumping model predicts that export success will be negatively related to marginal cost). The critique is not that (29) is inconsistent with a Ricardian model, but that verification of $\beta > 0$ is not evidence in favor of Ricardo over any other explanation. A truly general equilibrium prediction of Ricardian models is that a productivity advantage in one sector can actually *hurt* export success in another sector, but Golub and Hsieh do not investigate this prediction.

The same critique applies to Choudhri and Schembri (2000), which looks at US-Canada trade. Choudhri and Schembri integrate product differentiation into the Ricardian model and derive their estimating equation carefully from theory, but the end result is something similar to (29) above, which relates export success in sector g to relative productivity in sector g . As with Golub and Hsieh (2000), they are silent on cross-productivity effects.

2.4 Conclusions and Unfinished Business - Comparative Advantage

A decade of research on empirical models of comparative advantage has made some progress:

1. We now have our first confirmation of the theory of comparative advantage in its general, autarky price form (Bernhofen and Brown 2000).
2. Many papers have demonstrated that, at least for manufactured goods, relative factor supplies are an important influence on specialization (Harrigan 1995, 1997a, Bernstein-and Weinstein 1998, Schott 2000, and Harrigan and Zakrajšek 2000).
3. Technological differences have been shown to be an important influence on specialization (Harrigan 1997a).
4. The simple even factor-proportions model pioneered by Leamer (1984) is too simple: output indeterminacy (Bernstein-Weinstein 1988), Ricardian effects (Harrigan 1997), and multiple cones (Schott 2000) are all empirically important.

All of the papers reviewed in this chapter have been guided by the view that a careful application of theory is important when investigating the theory of comparative advantage (perhaps this is partly in reaction to the prolonged, confused response of the profession to Leontief's alleged paradox).

It goes without saying that whatever progress has been made, we are a long way from fully understanding the determinants and empirical significance of comparative advantage. Some of the open questions are

1. What is the role of transport costs, or distance more generally, in determining specialization?
2. How do non-comparative advantage influences on specialization, such as increasing returns, interact with technology and factor endowment differences to determine specialization?
3. How are trade flows determined? Is a simple model of preferences enough, or do we need to model income effects and/or the demand for intermediate

goods?

4. What is the appropriate unit of analysis in thinking about comparative advantage? Should we be studying broad industries, or concentrating on firm or plant level models?
5. How can we measure and characterize the cross-sectional and time-series distribution of factor prices and goods prices?
6. Is there evidence for general equilibrium effects of sectoral productivity differences of the sort predicted by simple Ricardian models?

Some of these questions are addressed elsewhere in this volume, but empirical research on comparative advantage is still a young and underdeveloped field. There is plenty of opportunity for good empirical work to continue to sharpen our understanding of the relevance of our basic trade models.

3 The Gravity Equation

James Anderson began his 1979 article "A Theoretical Foundation for the Gravity Equation" by saying that

Probably the most successful empirical trade device of the last twenty-five years is the gravity equation (Anderson, 1979, page 106).

One could say the same thing today, as the gravity equation remains at the center of a great deal of applied research on international trade. Another thing that has not changed since 1979 is that there is great uncertainty about the foundations of the gravity model: what do we mean when we say that it "works", and why does it work? Recently there have been a few papers that try to empirically understand gravity, and a review of this recent research on the foundations of gravity is the focus of this section.

The gravity equation is so-named because it is a theory of trade volumes which is analogous to the physical theory of gravity: trade between a pair of countries depends positively on the product of economic size and negatively on distance, just as the force of gravity between two bodies increases with the product

of their mass and decreases with distance. In its simplest form the gravity equation is

$$M_{cd} = k \frac{Y_c Y_d}{D_{cd}} \quad (30)$$

where M_{cd} denotes the value of imports by country c from country d , D_{cd} is the distance between the two countries, Y is a measure of economic size such as GDP, and k is a constant. Introducing parameters which allow the elasticities of trade volumes with respect to size and distance to differ from one does not change the basic message. As it stands equation (29) is not an economic model, but it is nevertheless plausible. Transport and other trade costs are certainly correlated with distance, so distance will surely reduce trade. Equally obviously, trade between the United States and Japan will probably be larger than trade between Estonia and Portugal. Nevertheless, what is striking about equation (29) is that there is apparently no role for comparative advantage: neither relative endowments nor relative technology levels enter the equation. It is this apparent lack of connection to neoclassical trade theory that led to the widespread conclusion that the gravity equation had no theoretical foundation. A related observation is that neoclassical trade theory is generally not concerned with bilateral trade: in comparative advantage models, a country's trade is determined by its differences from the rest of the world, with no prediction about the pattern of bilateral trade.

Foundation or no, equation (29) fits the data remarkably well. Regressions (in logarithms) of bilateral aggregate trade volumes on the GDP of trading partners and the distance between them typically yield R^2 's in the range of 0.65 to 0.95. What is important about these high R^2 's is that they have led many researchers to use variants of the gravity equation as a benchmark for the volume of trade. This gravity-based benchmark is then used to evaluate economic policy issues such as the effects of protection (Harrigan 1993), openness (Lawrence 1987, Saxonhouse 1989, Harrigan 1996), the merits of proposed regional trade agreements (Frankel,

Stein, and Wei 1997), and the effects of national borders (McCallum 1995, Evans 2000, Anderson and van Wincoop 2001).

3.1 The Theory of Gravity

In fact, there are several theoretical foundations for the gravity model. One of the earliest is due to Anderson (1979), with other contributions from Bergstrand (1985, 1989). These models all have the feature that consumers regard goods as being differentiated by location of production, a modeling trick known as the "Armington assumption" (Armington, 1969). The standard specification for Armington preferences is a variant of the CES functional form:

$$U^c = \left(\sum_{d=1}^C \beta_d c_{cd}^{(\sigma-1)/\sigma} \right)^{\frac{\sigma}{\sigma-1}}, \quad \sigma \in (1, \infty) \quad (31)$$

where c_{cd} is consumption by country c residents of goods produced in country d , and β_d and σ are parameters which are common across all countries. The key feature of these preferences is that goods are differentiated by country of origin only. Another feature of this function is that the marginal utility of consumption of goods from all countries d is always strictly positive, and infinite in the limit as consumption goes to zero. This implies that, whatever the price, country c will consume at least some of every good from every country.

Anderson's and Bergstrand's models, and most other explanations for the negative effect of distance on trade, assume that transport costs are of the "iceberg" form, where for every $t > 1$ units shipped from the exporter, only 1 unit arrives at the importer's location, the other $t-1$ units having "melted" in transit. As long as exporters do not price discriminate across export markets, there will be a single f.o.b. price p_d for country d 's exports, and the c.i.f. price in country c of imports from country d will be $p_d t_{cd}$. In empirical applications, t is usually assumed to be a monotonically increasing function of distance. This way of handling

transport costs is ubiquitous not because it is realistic but because it is very handy: as Grossman (1998) notes,

...few would consider the “iceberg” formulation of shipping costs as anything more than a useful trick for models with constant demand elasticities, and possibly a good approximation to the technology for shipping tomatoes (Grossman, 1988, pg. 30-31).

Hummels (1999) provides some evidence on the actual form of the relationship between transport costs and distance, while a number of authors (including Roberts and Tybout (1997) and Bernard and Jensen (2001)) have shown that there are important fixed costs to trade which are independent of distance.

These reservations about the iceberg assumption notwithstanding, it is crucial to deriving a closed-form gravity equation based on the preferences given by (31)¹⁷. The basis of such derivations is the assumption that all goods are traded, so that national income is the sum of traded goods output, which in equilibrium is the sum of home and foreign demand for the unique good that the country produces. The demand function that arises from (31) is

$$c_{cd} = \beta_d Y_c \frac{(p_d t_{cd})^{-\sigma}}{P_c^{1-\sigma}} \quad (32)$$

where the CES price index P_c is defined as

$$P_c = \left[\sum_{b=1}^C \beta_b (p_b t_{cb})^{1-\sigma} \right]^{1/1-\sigma} \quad (33)$$

The goods market clearing conditions are

¹⁷ The derivation here follows Deardorff (1998) and Anderson and van Wincoop (2001).

$$Y_c = p_c \sum_{b=1}^C c_{bc} \quad (34)$$

These equations can be solved in a way that will generate the gravity equation. Choosing units so that all f.o.b. prices are equal to unity and letting $s_c = Y_c/Y_w$ be c 's share of world income, it can be shown that

$$s_c = \beta_c \sum_{b=1}^C s_b \left(\frac{t_{bc}}{P_b} \right)^{1-\sigma} \quad (35)$$

This equation states that country c 's national income depends on two things: first, the popularity of the goods that it produces (the taste parameter β_c) and second, a GDP-weighted average of its distance from trading partners (assuming that distance and transport costs are positively related). The dependence of income on β_c is unattractive on economic grounds: is it really plausible that the United States has a high GDP because consumers around the world have a taste for US goods? The negative effect of distance on income is much more believable, and is a common implication of economic geography models (see, for instance, the chapter in this volume by Overman, Redding and Venables).

Solving the model (see Anderson and van Wincoop (2001)) for imports as a function of income and trade costs gives the gravity equation for imports by c from d :

$$M_{cd} = \frac{Y_c Y_d}{Y_w} \times t_{cd}^{1-\sigma} \times P_c^{\sigma-1} \times P_d^{\sigma-1} \quad (36)$$

where M_{cd} is the c.i.f. value of imports by c from d . The first and second terms are easy to understand: big countries import and export more than small countries, and trade costs reduce trade volumes with an elasticity of $(1-\sigma)$. The third term is

a substitution effect: if transport costs facing c are high on average so that P_c is large, then c will import more from d . The fourth term varies across exporting sources d and is increasing in a weighted average of d 's transport costs: if d is on average a long way from its trading partners, it will have a low f.o.b. price, so c will import more. More succinctly, controlling for country size and bilateral distance, trade will be higher between country pairs that are far from the rest of the world than between country pairs that are close to the rest of the world.

The relative distance effect explains why the theoretically-derived gravity model of (36) differs from the simple gravity equation (30). To better understand this relative distance effect, consider a world of five equal-sized countries evenly spaced along a line:



Country B will import more from country A than it will from country C, despite the fact that B is an equal distance from both. This is because A is so far from everyone else that aggregate demand for its output will be low, resulting in an f.o.b. export price lower than C's f.o.b. export price. Therefore, B's c.i.f. import prices will be lower from A than from C.

The development of (36) makes heavy use of the CES and iceberg assumptions, but the point that relative as well as absolute distance matters for bilateral trade seems much more general. As a consequence, gravity equations which pool across bilateral pairs without controlling for relative distance are misspecified in a potentially important way. Structural estimation of (36) is difficult because of the non-linear functional form of the price index terms and the presence of the unknown parameters β (for an application which imposes all the structure of the model, see Anderson and Van Wincoop (2001)). An intuitive if ad-hoc proxy for the inverse of the price term defined by (33) can be developed by taking $\sigma = 2$, replacing the unknown β 's with income shares, and choosing units so

that fob prices are unity to get a "centrality index":

$$\Psi_c = \sum_{b=1}^C \frac{s_b}{t_{cb}} \quad (37)$$

According to (37), centrality is a GDP-weighted average of the inverse of trade costs. By the logic of the CES gravity model developed above, bilateral trade should be decreasing in the centrality of the two trading partners, since a central location means there are many alternative nearby sources of supply for the importer, and high demand and hence high f.o.b. prices for the exporter.

Different proxies have been used by several authors in the literature, including Helliwell (1997) and Wei (1996). Wei defines the "remoteness index" as a GDP-weighted average of distance,

$$R_c^{Wei} = \sum_{b=1}^C s_b t_{cb} \quad (38a)$$

which is related in spirit to (33) but can not be derived from any simplification of the CES price index. Helliwell (1997) defines remoteness as

$$R_c^{Hw} = \sum_{b=1}^C \frac{t_{cb}}{s_b} \quad (38b)$$

which makes little sense, since distance from small countries matters more than distance from large countries in determining a country's remoteness.

This discussion makes the point that controlling for relative distance is crucial to estimating a well-specified gravity model, and that there are a number of reasonable ways to measure relative distance (although Helliwell's index is not one of them). However, if consistent estimation of the distance effect ($\sigma - 1$) is what is of interest, then researchers can impose the theory-required unit elasticities on

income and run the following simple regression with country fixed effects:

$$\ln \frac{M_{cd}}{Y_c Y_d} = \delta_c + \delta_d - (\sigma - 1) \ln t_{cd} \quad (39)$$

where the δ 's are dummy variables which sweep out the influence of importer and exporter relative distance.

As a foundation for the gravity equation the Armington model just described relies on some convenient functional form assumptions (icebergs and CES) which can be criticized, but its biggest weakness is that there is no microfoundation for the production side of the model, if one can even say that there is a production side at all. It was left to the "new trade theory" of the 1980s to provide a solid theoretical grounding for the production side of the gravity equation. The monopolistic competition model (summarized elegantly in Helpman and Krugman (1985)) provides just such a model for the zero transport costs case. In the monopolistic competition model, a taste for variety interacts with firms who face increasing returns to scale in the production of varieties. With identical, homothetic CES preferences on the demand side and strong symmetry assumptions on the supply side, the equilibrium of the Helpman-Krugman model provides a rationalization for the Armington utility function (31), with the parameters β_c reinterpreted as proportional to the number of varieties produced in equilibrium in country c . One result is a strikingly simple model for bilateral trade:

$$M_g^{bc} = s_b x_g^c \quad (40)$$

where M_g^{bc} is country b 's imports of good g from country c , s_b is b 's share of world expenditure, and x_g^c is b 's output of good g . Summing over all goods g gives the aggregate gravity model of equation (29) for $k = 1/Y^w$ and $D_{bc} = 1$. One appealing aspect of equation (40) is that it gives predictions on a sectoral basis, and so can be tested using sectoral data on production and trade.

Extending the monopolistic competition model to the case of positive trade costs is straightforward. If all goods are produced in monopolistically competitive sectors, then with CES preferences the equilibrium number of varieties per country is invariant to trade costs (see Krugman 1980), and the model is isomorphic to the Armington model above. As a result, the gravity equation (36) follows immediately.

For the purposes of deriving a gravity equation, the key feature of the Armington and monopolistic competition models is that goods are differentiated by location of production, whether by assumption (Armington) or endogenously (monopolistic competition). Equivalently, countries are completely specialized in disjoint sets of goods. In contrast, Feenstra, Markusen, and Rose (1998, 2001) consider whether it is possible to get a gravity-type relationship in a model of trade in homogeneous goods. Feenstra et al consider a general equilibrium model of "reciprocal dumping", where Cournot-Nash oligopolists sell a homogeneous good in each others' markets à la Brander and Krugman (1983). The model is very simple, with two countries sharing the same technology and a single factor of production. To get analytical results they assume Cobb-Douglas preferences (implying unit elasticity of market demand) and no transport costs in the numeraire sector (guaranteeing factor price equalization). In footnote 2 of Feenstra et al (2001) they assert that Feenstra et al (1998) "derive and illustrate the gravity equation for the reciprocal dumping model", but that is not quite right. The 1998 paper shows, in a two-country world, that the sum of world exports is maximized when country size is equalized, a result that they call "the most important implication of the 'gravity' equation" (Feenstra et al 1998, page 9), but they do *not* show that bilateral imports or exports depend on the product of trading partner GDPs, which is the usual statement of the gravity equation (equation (30) above). In fact, their reciprocal dumping model has some very un-gravity like implications: bilateral trade is increasing in country size only over a limited range, and two-way trade only occurs when countries are of similar size.

A model of trade in homogeneous goods which *does* generate a gravity-type relationship is developed in Eaton and Kortum (2001). Their framework is a multi-country perfectly competitive Ricardian model with a continuum of goods and iceberg transport costs, a complex set of assumptions that nonetheless yields intuitive and elegant implications. The foundation of their modeling strategy is the assumption that country c 's productivity in g , $1/a_c(g)$, is a random variable drawn from the Frechet distribution function

$$F_c(a) = \exp(-T_c a^{-\theta}) \quad (41)$$

where T_c has the interpretation of the absolute technology level in country c and $1/\theta$ is related to the dispersion of productivities across goods, and hence measures the potential for comparative advantage. Country c has wages w_c , so c 's unit cost of producing good g is $w_c/a_c(g)$. With these costs, c supplies g to country d at a cost of

$$p_{cdg} = \frac{w_c}{a_c(g)} \tau_{cd} \quad (42)$$

where $\tau_{cd} > 1$ is the iceberg transport factor between c and d . But d will not necessarily buy from c : it will only do so if c has the lowest c.i.f. price available in d . This will be more likely if c and d are close to each other, and if c is cost-competitive in a wide range of goods.

Prices in country d depend on technology and input costs in the rest of the world, and transport costs. The price index can be shown to be

$$P_d = \sum_{c=1}^C T_c (w_c \tau_{cd})^{-\theta} \quad (43)$$

This price index is increasing in weighted distance: it is higher if you are a long way from countries with good technology (that is, high T_i 's). For $\theta = 1$ it bears a

family resemblance to the centrality index of equation (37) above, which confirms in a very different model the general point that relative distance matters for trade flows. A few more steps gives d 's imports from c :

$$M_{dc} = Y_d \frac{T_c (w_c \tau_{cd})^{-\theta}}{\Phi_d} \quad (44)$$

This is starting to look a lot like a gravity equation, and in fact Eaton and Kortum show that it reduces to the frictionless gravity equation $M_{dc} = kY_c Y_d$ when there are no transport costs. More generally, national income in c will depend on absolute advantage T_c and on c 's location in the world, summarized by P_c , as well as the national endowment of labor. Therefore, just as in the Armington model of (36) above, bilateral trade in this Ricardian model depends on country size, distance, and relative distance. But the effect of distance here is very different than in the Armington model: rather than reduce the volume of a given set of country c goods that are consumed, distance shrinks the set of goods that d chooses to buy from c . In the equilibrium of the model, many countries produce and export the same goods, but they do not sell in the same markets: if a country imports a good it will generally do so from just one source. As such, Eaton and Kortum challenge the view that complete specialization is a necessary condition for the gravity equation.

3.2 Why does Gravity work? Discriminating among alternative explanations

We now have plenty of evidence for the aggregate gravity equation and plenty of theoretical foundations for it. Is Deardorff (1998) correct that, with so many potential fathers, we can not determine the gravity equation's paternity? Deardorff's view is supported by the results of Hummels and Levinsohn (1995), who found that gravity worked just as well for poor countries as it did for rich

countries¹⁸. They argued that this was a surprise, since the production side of the monopolistic competition model (which they took to be the foundation of gravity) is likely to be appropriate for rich, but not poor, countries. A number of recent papers argue that Deardorff was too pessimistic, but empirically selecting among the potential explanations for gravity is still at a relatively early stage.

Most of the evidence that "gravity works" comes from aggregate data, where total bilateral trade is regressed on GDP. This is despite the fact that the models developed to explain gravity often apply also at the sectoral level (see equation (40)). Given this, it is surprising how little work has been done on examining disaggregated gravity equations, or on looking for instances where gravity fails. One recent attempt to do so is Haveman and Hummels (2001), who examine a large data set of bilateral trade flows at the 4-digit SITC level. The most striking result in this paper is the number of bilateral zeros: most potential bilateral trades in a given SITC code do not occur. In particular, when a country imports a good it usually imports it from only one source, and when a country exports a good it usually exports it to a limited number of countries. This pattern could be rationalized by a model of product differentiation with non-CES preferences and/or fixed costs to transporting goods (although no-one has solved such a model), but it is certainly at odds with the standard gravity model specification which assumes symmetric CES preferences and iceberg transportation costs. A large number of zeros is explicitly predicted by the Ricardian model of Eaton and Kortum (2001), which is one of the few predictions about gravity from that model which differ from the gravity predictions of complete specialization models. The Eaton-Kortum model can also be expected to work just as well for poor countries that produce homogeneous goods as it does for rich countries, so the Hummels-Levinsohn critique of the excessively good performance of gravity does not apply

¹⁸ See Debaere (2000) for a critique of Hummels and Levinsohn. Debaere argues that the Hummels and Levinsohn data are friendlier to the monopolistic competition model than they thought.

to the Eaton-Kortum model.

Another paper which looks at disaggregated gravity predictions is Feenstra et al (2001). The authors present a series of simple models that generate gravity-like equations, that is, where both importer and exporter GDP help to explain bilateral trade. The theory models predict that where there is free entry there will be home-market effects, that is, exports will be more than proportional to GDP. This implies that the effect of exporter GDP will be larger than the effect of importer GDP in a gravity equation. Conversely, when there are barriers to entry, there is a reverse home market effect, so the effect of exporter GDP will be smaller than the effect of importer GDP in a gravity equation. The empirical implication is that the GDP elasticities in a gravity equation should be different depending on whether or not there are entry barriers, and the empirical problem is that there are no internationally comparable barriers on sectoral entry barriers. They proceed under the hypothesis that different types of goods might have different types of entry barriers, which would imply different gravity equation coefficients for different types of goods. This is exactly what they find: using a classification scheme due to Rauch (1999), they find large and precisely estimated differences in the GDP coefficients in the gravity equation across differentiated, 'reference priced', and homogeneous goods. Some illustrative results, from 1990, are given in the following table:

Gravity Equation Estimates for Different Types of Goods

	Differentiated	Reference Priced	Homogeneous
Exporter GDP	1.12	0.91	0.54
Importer GDP	0.72	0.74	0.81

Source: Feenstra et al (2001), Table 2.

The interpretation of these results is not straightforward, because of the limitations of the theory and because there is no direct evidence on entry barriers.

Nonetheless, the results are striking and thought provoking, and suggest that further research on how gravity works for different types of trade flows will be fruitful.

Feenstra et al disaggregate one side of the gravity equation, but not the other: their disaggregated imports are always explained by importer and exporter GDP. A different approach is based on equation (40) above, which states that sectoral trade flows depend on importer GDP (demand) and exporter sectoral output (supply). If (40) held true for all sectors, then the ratio of trade to output in sector g among a group of countries C is

$$\text{predicted } \frac{M_g}{x_g} = \sum_{c=1}^C s_c \left(1 - \frac{x_{cg}}{x_g}\right) \quad (45)$$

where M_g and x_g are the intra-group totals of trade and output of good g . Harrigan (1996) calculates that this predicted ratio is about 0.5 among the OECD countries in 1985, and shows that the actual ratio of trade to output is much less than 0.5 and varies by a factor of ten across manufacturing industries:

The volume of trade relative to output within the OECD, 1985

Industry	trade/output
Leather shoes	0.30
Transport equipment	0.26
Basic chemicals	0.24
Electrical machinery	0.18
Textiles	0.14
Basic iron and steel	0.12
Fabricated metals	0.07
Food	0.07
Cement	0.05
Printing and publishing	0.03

Source: from Harrigan (1996), Table 1.

This table suggests some sector-specific explanations for trade volumes: the high volume of trade in leather shoes and transport equipment seems to fit the product differentiation story, while the low volume of trade in cement and publishing are probably due to transport costs (cement is heavy) and home-biased tastes (Frenchmen read few books published in English) respectively. Whatever the sector-specific explanations, the large cross-sector variation in trade relative to output suggests that empirical work on understanding the volume of trade should work with disaggregated data.

Harrigan (1994) was the first to look at equations like (40) in the context of trying to understand the performance of the gravity model, arguing that monopolistic competition predicts that the volume of trade will be higher in sectors characterized by scale economies. The specification in that paper is flawed since it fails to control for bilateral distance, but it is notable that Harrigan finds a fairly robust result that the volume of bilateral trade is higher in sectors with larger scale economy proxies.

Like Harrigan (1996), Lai and Trefler (1999) estimate sectoral gravity equations, but they are much more careful than Harrigan was to use all the structure given by the CES functional form assumption¹⁹. In addition to using the model for policy analysis, they focus on how well the model fits at a sectoral level. To control for distance and other time-invariant influences on bilateral trade, they use panel data with country-pair fixed effects. The fixed effects strategy has the usual advantages and disadvantages: it gives consistent estimates of the parameters of interest but discards the overwhelming majority of the variation in the data, which is in the cross-section of country pairs. They find that the correlation between fitted and actual trade volumes is highest for industries where (they claim) the monopolistic competition model is more appropriate, but they make no attempt to formally identify which industries "should" fit the model's predictions. In their Figure 3, they show that all the hard work in dealing with the CES price term makes no difference to model fit: dropping the price term gives the same correlation as including it. They also confirm in Figure 3 that the fit is largely driven by the output terms on the right hand side, that is, the gravity effect on the supply side. They refer to this as a data identity, but that is not correct: the presence of sectoral output in a disaggregated gravity equation reflects the assumption that products are differentiated. Their final conclusion is that the gravity model doesn't work nearly as well when it is scrutinized at a sectoral level: among other anomalies, the volume of trade is less than predicted, the elasticity of trade with respect to partner production is not one, and the CES structure adds little to a more naive specification.

A paper which argues that we can use aggregate data to see why gravity works is Evenett and Keller (2001). Their approach is to derive the aggregate gravity model using several simple textbook trade models, and then see whether

¹⁹ While Lai and Trefler claim that their paper casts light on the monopolistic competition model, in fact their model takes sectoral production as given, and should therefore be regarded as a general gravity equation in which the Armington and monopolistic competition models are isomorphic.

gravity works better in sub-samples of country pairs that are thought to better fit the presumptions of the different models. Unfortunately, their results are not informative about why gravity works for several reasons. Most importantly, they work with two-country frictionless models, which have no predictions for bilateral trade in a many-country world where trade costs matter. Secondly, they use intraindustry trade indices to stratify their sample, despite the demonstrations by Davis (1995, 1997) that the proportion of intraindustry trade has nothing to do with the causes of gross trade volumes.

3.3 Conclusions and Unfinished Business - Gravity

Despite being a staple of applied analysis because it "works well", the gravity model has been subject to surprisingly little empirical testing. Some of the facts that the papers reviewed in the previous section have uncovered can be summarized as follows:

1. There are many zero observations in disaggregated bilateral trade.
2. The volume of trade is much smaller than predicted by the frictionless gravity model.
3. Relative, as well as absolute, distance and trade costs matter for understanding bilateral trade in a multi-country world.
4. The elasticity of trade with respect to output differs from one and is not uniform across goods. These differences may be related to the type of good and/or market structure.
5. The CES model of preferences doesn't fit the data.
6. The ratio of trade to output varies by an order of magnitude across industries.
7. There is some evidence that the volume of trade is higher in sectors characterized by monopolistic competition and/or scale economies.

This list raises more questions than it answers. What explains the zeros? Why is the volume of trade to output so small, and why is there so much variation in it? What model of consumption might improve on CES? Is it really the case that the

volume of trade is higher in industries with scale economies? No doubt the astute reader can think of other questions left unanswered, and perhaps ambitious readers will try to answer them.

4 General Conclusions

This chapter has surveyed a decade's worth of empirical research on how well the data obey the laws of comparative advantage and gravity. Detailed conclusions from the survey are summarized in sections 2.4 and 3.3.

Given the centrality of comparative advantage and gravity to applied international economics, it is surprising that there has not been more empirical research, and sobering if not frustrating that progress has been so slow. A clear message from this chapter is that, while we have learned something about how specialization and the volume of trade are determined, there are large gaps in our knowledge. The opportunities for future researchers to help fill these gaps are equally large.

References

- Anderson, James E., 1979, "A Theoretical Foundation for the Gravity Equation", *American Economic Review* 69(1): 106-116.
- Anderson, James E., and Eric van Wincoop, 2001, "Gravity with Gravitas: A Solution to the Border Puzzle", NBER Working Paper no. 8079.
- Ark, B. van, 1993, "Comparative Levels of Manufacturing Productivity in Postwar Europe: Measurement and Comparisons", *Oxford Bulletin of Economics and Statistics* 52, 343-374.
- Ark, B. van, and D. Pilat, 1993, "Productivity Levels in Germany, Japan, and the United States: Differences and Causes", *Brookings Papers Microeconomics* 2, 1-69.
- Armington, P.S., 1969, "A Theory of Demand for Products Distinguished by Place of Production", *IMF Staff Papers*, v. 16: 159-176.
- Balassa, Bela, 1963, "An Empirical Demonstration of Classical Comparative Cost Theory", *Review of Economics and Statistics* 4: 231-238.
- Bergstrand, Jeffrey, 1985, "The Gravity Equation in International Trade: Some Microeconomic Foundations and Empirical Evidence", *The Review of Economics and Statistics* 67 (August): 474-481.
- Bergstrand, Jeffrey, 1989, "The Generalized Gravity Equation, Monopolistic Competition, and the Factor-Proportions Theory in International Trade", *The Review of Economics and Statistics* 71 (February), 143-153.

Bernard, Andrew B., and J. Bradford Jensen (2001), "Why Some Firms Export", NBER Working Paper No. 8349.

Bernhofen, Daniel M., and John C. Brown, 2000, "A Direct Test of the Theory of Comparative Advantage: The Case of Japan", manuscript, Clark University.

Bernstein, Jeffrey R., and David E. Weinstein, 1998, "Do endowments predict the location of production? Evidence from national and international data", NBER Working Paper no. 6815.

Brander, James, and Paul Krugman, 1983, "A 'Reciprocal Dumping' Model of International Trade", *Journal of International Economics* 15: 313-321.

Choudhri, Ehsan U., and Lawrence L. Schembri, 2000, "Productivity Performance and International Competitiveness: A New Test of an Old Theory", manuscript, Carleton University.

Davis, Donald R., 1995, "Intraindustry trade: a Heckscher-Ohlin-Ricardo approach", *Journal of International Economics*

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

Davis, Donald R., and David E. Weinstein, 1998, "An Account of Global Factor Trade", NBER Working Paper no. 6785 (November). Forthcoming, *American Economic Review*.

Deardorff, Alan V., 1979, "Weak Links in the Chain of Comparative Advantage", *Journal of International Economics* 9(2): 197-209.

Deardorff, Alan V., 1980, "The General Validity of the Law of Comparative Advantage", *Journal of Political Economy* 88: 941-957.

Deardorff, Alan V., 1998, "Determinants of Bilateral Trade: Does Gravity Work in a Neoclassical World?", in Jeffrey A. Frankel, Ed., *The Regionalization of the World Economy*, Chicago: University of Chicago Press for the NBER, 7-22.

Debaere, Peter, 2000, "Testing 'New' Trade Theory without Testing for Gravity: Reinterpreting the Evidence", manuscript, University of Texas.

Dixit, Avinash, and Victor Norman, 1980, *The Theory of International Trade*, Cambridge, England: Cambridge University Press.

Dollar, D., W. Baumol, and E. N. Wolff, 1988, "The Factor Price Equalization Model and Industry Labor Productivity: an Empirical Test Across Countries", in: R. C. Feenstra, ed., *Empirical Methods for International Trade* (MIT Press, Cambridge, MA).

Dollar, D., and E. N. Wolff, 1993, *Competitiveness, Convergence, and International Specialization* (MIT Press, Cambridge, MA).

Eaton, Jonathan, and Samuel Kortum, 2001, "Technology, Geography, and Trade", manuscript, Boston University. Forthcoming, *Econometrica*.

Evans, Carolyn, 2000, "The Economic Significance of National Border Effects", manuscript, International Research Department, Federal Reserve Bank of New York.

Evenett, Simon J., and Wolfgang Keller, 2001, "On Theories Explaining the

Success of the Gravity Equation", manuscript, University of Texas. Forthcoming, *Journal of Political Economy*.

Feenstra, Robert C., James A. Markusen, and Andrew K. Rose, 1998, "Understanding the Home Market Effect and the Gravity Equation: The Role of Differentiating Goods", NBER Working Paper no. 6804 (November).

Feenstra, Robert C., James A. Markusen, and Andrew K. Rose, 2001, "Using the Gravity Equation to Differentiate Among Alternative Theories of Trade" *Canadian Journal of Economics* 34 (2): 430-447.

Frankel, Jeffrey A., with Ernesto Stein and Shang-jin Wei, 1997, *Regional Trading Blocs in the World Economic System*, Washington DC: Institute for International Economics.

Fujita, Masahisa, Paul Krugman, and Anthony J. Venables, 1999, *The Spatial Economy: Cities, Regions, and International Trade*, Cambridge, MA: MIT Press.

Golub, Stephen S., and Chang-Tai Hsieh, 2000, "Classical Ricardian Theory of Comparative Advantage Revisited", *Review of International Economics* 8(2): 221-34.

Grossman, Gene, 1998, "Comment on Deardorff" in Jeffrey A. Frankel, Ed., *The Regionalization of the World Economy*, Chicago: University of Chicago Press for the NBER, 29-31.

Harrigan, James, 1993, "OECD Imports and Trade Barriers in 1983", *Journal of International Economics* 35: 95-111.

Harrigan, James, 1994, "Scale Economies and the Volume of Trade", *The Review of Economics and Statistics* 76 (2): 321-328.

Harrigan, James, 1995, "Factor Endowments & The International Location of Production: Econometric Evidence for the OECD, 1970-1985", *Journal of International Economics*, v. 39 nos. 1/2 (August): 123-141.

Harrigan, James, 1996, "Openness to Trade in Manufactures in the OECD", *The Journal of International Economics* 40 (February): 23-39.

Harrigan, James, 1997a, "Technology, Factor Supplies and International Specialization: Estimating the Neoclassical Model", *American Economic Review*, v. 87 no. 4 (September): 475-494.

Harrigan, James, 1997b, "Cross-Country Comparisons of Industry Total Factor Productivity: Theory and Evidence", Federal Reserve Bank of New York Research Paper no. 9734.

Harrigan, James, 1999, "Estimation of Cross-Country Differences in Industry Production Functions", *Journal of International Economics* v. 47 no. 2 (April): 267-293.

Harrigan, James, and Egon Zakrajšek, 2000, "Factor Supplies and Specialization in the World Economy", 2000, NBER Working Paper no. 7848 (August) and *Federal Reserve Bank of New York Staff Report* no. 107 (August).

Haveman, Jon, and David Hummels, 2001, "Alternative Hypotheses and the Volume of Trade: Evidence on the Extent of Specialization", manuscript, Purdue University.

Head, Keith, and John Ries, 2000, "Increasing Returns Versus National Product Differentiation as an Explanation for the Pattern of US-Canada Trade", manuscript, University of British Columbia, and forthcoming, *American Economic Review*.

Helliwell, John F., 1997, "National Borders, Trade, and Migration", NBER Working Paper no. 6027.

Helpman, Elhanan, and Krugman, Paul, *Market Structure and Foreign Trade: Increasing Returns, Imperfect Competition, and the International Economy*, Cambridge, MA: MIT Press, 1985.

Hummels, David, 1999, "Toward a Geography of Trade Costs" , manuscript, Purdue University.

Hummels, David, and James Levinsohn, 1995, "Monopolistic Competition and International Trade: Reconsidering the Evidence", *Quarterly Journal of Economics* 110 (3): 799-836.

Hunter, Linda, 1991, "The contribution of nonhomothetic preferences to trade", *Journal of International Economics* 30: 345-358.

Hunter, Linda, and James Markusen, 1988, "Per capita income as a basis for trade", in Robert C. Feenstra, Ed., *Empirical Methods for International Trade*, Cambridge, MA: MIT Press.

Jones, Ronald W., and José A. Scheinkman, 1977, "The Relevance of the Two-Sector Production Model in Trade Theory", *Journal of Political Economy* 85 (5): 909-935.

Jorgenson, D. W., M. Kuroda, and M. Nishimizu, 1987, "Japan-U.S. Industry-Level Productivity Comparisons, 1960-1979", *Journal of the Japanese and International Economies* 1, 1-30.

Jorgenson, D. W., and M. Kuroda, 1990, "Productivity and International Competitiveness in Japan and the United States, 1960-1985", in: C. R. Hulten, ed., *Productivity Growth in Japan and the United States* (University of Chicago Press, Chicago).

Kohli, Ulrich, 1991, *Technology, Duality, and Foreign Trade*, Ann Arbor: University of Michigan Press.

Krugman, Paul, 1980, "Scale Economies, Product Differentiation, and the Pattern of Trade", *American Economic Review* 70: 950-959.

Lai, Huiwen, and Daniel Trefler, 1999, "The Gains From Trade: Standard Errors with the CES Monopolistic Competition Model", manuscript, University of Toronto.

Lawrence, Robert Z., 1987, "Does Japan Import Too Little: Closed Minds or Markets?", *Brookings Papers on Economic Activity* 2: 517-554.

Leamer, Edward E., 1987, "Paths of Development in the Three-Factor, n-Good General Equilibrium Model", *Journal of Political Economy* 95 (October): 961-999.

Leamer, Edward E., 1995, "The Heckscher-Ohlin Model in Theory and Practice", *Princeton Studies in International Finance* no. 77.

Leontief, Wassily, 1954, "Domestic Production and Foreign Trade: The American Capital Position Re-examined", Chapter 30 in Richard E. Caves and Harry G. Johnson, eds., *Readings in International Economics*, 1968, London: Allen and Unwin.

MacDougall, G.D.A., 1951, "British and American Exports: A Study Suggested by the Theory of Comparative Costs, Part I", *The Economic Journal* 61: 697-724.

MacDougall, G.D.A., 1952, "British and American Exports: A Study Suggested by the Theory of Comparative Costs, Part II", *The Economic Journal* 62: 487-521.

Markusen, James, 1986, "Explaining the Volume of Trade: An Eclectic Approach", *American Economic Review* 76: 1002-1011.

McCallum, John, 1995, "National Borders Matter: Canada-US Regional Trade Patterns", *American Economic Review* 85 (June): 615-623.

Noussair, Charles N., Charles R. Plott, and Raymond G. Riezman, 1995, "An Experimental Investigation of the Patterns of International Trade", *American Economic Review* 85 (3): 462-491.

Rauch, James, 1999, "Networks versus markets in international trade", *Journal of International Economics* 48, 7-37.

Roberts, Mark J. and Tybout, James R. (1997) "The Decision to Export in Colombia: An Empirical Model of Entry with Sunk Costs" *American Economic Review* 87 (4): 545-564.

Rousslang, Donald J., and Theodore To, 1993, "Domestic trade and transportation

costs as barriers to international trade", *Canadian Journal of Economics* 26(1), 208-221.

Schott, Peter, 2000, "One Size Fits All? Heckscher-Ohlin Specialization in Global Production", manuscript, Yale University.

Trefler, Daniel, 1993, "International Factor Price Differences: Leontief was Right!", *Journal of Political Economy*, v. 101 (December): 961-987.

Trefler, Daniel, 1995, "The Case of the Missing Trade and Other Mysteries", *American Economic Review*, v. 85 (December): 1029-1046.

Wei, Shang-Jin, 1996, "Intra-national versus international trade: How stubborn are nations in global integration?", NBER Working Paper no. 5531.

Woodland, Alan D., 1982, *International Trade and Resource Allocation*, Amsterdam: North Holland.