On Theories Explaining the Success of the Gravity Equation

Simon J. Evenett

World Trade Institute and Centre for Economic Policy Research

Wolfgang Keller

University of Texas, National Bureau of Economic Research, and Centre for Economic Policy Research

We examine whether two important theories of trade, the Heckscher-Ohlin theory and the increasing returns theory, can account for the empirical success of the so-called gravity equation. Since versions of both theories can predict this equation, we tackle the model identification problem by conditioning bilateral trade relations on factor endowment differences and on the share of intraindustry trade. Only for large differences in factor endowments does the Heckscher-Ohlin model predict perfect production specialization in different countries as well as the gravity equation, and trade is purely in goods produced with different factor intensities. Our empirical analysis yields three findings. First, the predictions of the perfect specialization versions of both theories are rejected by the data and so are unlikely explanations for the empirical success of the gravity equation. Second, a model of imperfect specialization that includes both increasing re-

We have benefited from comments at seminar presentations at Berlin (Humboldt), Boston University, Brown, Dartmouth, Columbia, Geneva (Graduate Institute), Harvard, New York University, Penn, Toronto, University of California (San Diego and Santa Cruz), Wisconsin, and Yale, as well as from participants at the 1998 Midwest International Economics Conference in Ann Arbor, the 1997 NBER Summer Institute, the August 1997 European Economic Association meeting at Toulouse, the September 1997 conference of the European Research Workshop in International Trade/Centre for Economic Policy Research in Helsinki, and the Empirical Trade Conference 1996 at Purdue University. Special thanks go to Don Davis, Peter Debaere, Mahmoud El-Gamal, Gordon Hanson, Elhanan Helpman, Dan Trefler, Lars Peter Hansen as the editor, and especially to an anonymous referee. We also thank Jim Rauch and Karl Aiginger for providing data used in this paper.

[*Journal of Political Economy,* 2002, vol. 110, no. 2]

turns and factor endowments as sources of trade has a mixed performance: it correctly predicts production of more differentiated goods when the level of intraindustry trade is greater; however, the predicted link to factor proportions is tenuous. Third, the predictions of a model with imperfect specialization that relies solely on differences in factor endowments find support in the data. These results suggest that factor endowments and increasing returns explain different components of the international variation of production patterns and trade volumes.

The so-called gravity equation of trade predicts that the volume of trade between two countries is proportional to their gross domestic products and inversely related to trade barriers between them. Empirical research has found that various versions of the gravity equation can account for the variation in the volume of trade across country pairs and over time (see Leamer and Levinsohn 1995).

Since Anderson (1979), it has been increasingly recognized that the gravity equation can be derived from very different models, including Ricardian, Heckscher-Ohlin, and increasing returns to scale (IRS) models (Helpman and Krugman [1985], Bergstrand [1990], Markusen and Wigle [1990], Eaton and Kortum [1997], and Deardorff [1998]; see also the recent survey by Helpman [1998]). Many of these models have one characteristic in common: perfect specialization, where each commodity is produced in only one country. This multitude of models has led some to argue that the gravity prediction cannot be used to test any one of these trade theories (Deardorff 1998). Yet the gravity equation constitutes, perhaps along with the Heckscher-Ohlin-Vanek factor service trade prediction, one of the most important results about trade flows. Therefore, major insights into the causes of international trade might be gained if it could be determined which theory actually accounted for the success of the gravity equation in a given sample of data—an empirical challenge we refer to as a model identification problem.

We address this identification problem by noting that, in a constant returns to scale (CRS) Heckscher-Ohlin world, perfect specialization results only when bilateral differences in factor proportions are large enough. In contrast, when product specialization is the result of IRS, the gravity prediction can be obtained even if there are no differences in factor proportions. Furthermore, these models predict different types of trade: in the Heckscher-Ohlin model, trade occurs exclusively in goods produced with different factor intensities (interindustry trade). In contrast, in the IRS model, some, and potentially all, trade is intraindustry. Consequently, we shall take samples with high shares of intraindustry trade in total trade as those in which IRS-based specialization might drive the gravity equation. Furthermore, those samples with low shares of intraindustry trade in total trade and large differences in factor proportions are those in which a model of Heckscher-Ohlin-based specialization might be behind the gravity equation.

In principle this could lead to a sample-by-sample reconciliation of the perfect specialization models of the Heckscher-Ohlin and IRS trade theories. However, no such reconciliation emerges from our analysis of a large and heterogeneous set of bilateral trade relations in 1985. There is strong evidence that perfect specialization due to differences in factor proportions cannot explain the success of the gravity equation. The IRS model of perfect specialization performs only slightly better in our analysis. Our results suggest that perfect specialization models, which predict a strict proportionality of trade with GDP, overpredict the volume of trade compared to what is actually observed by a large amount. This reinforces the case for considering trade models in which production is imperfectly specialized.

The primary contribution of this paper is that we evaluate gravity equations based on imperfect specialization of production. The imperfect specialization models presented here predict a factor of proportionality of trade with GDP of less than one. By predicting a lower level of trade, all else equal, these models are, a priori, better candidates to empirically match the trade volumes that are observed. Moreover, in these models the degree of specialization is a function of relative factor abundance, a key exogenous variable. With estimates of the former and data on the latter, we have the means to evaluate these models. We find that trade among the industrialized countries ("North-North" trade) can be partially captured by a model that combines trade in perfectly specialized differentiated goods and homogeneous goods. Moreover, trade between less developed and industrialized countries ("North-South" trade) is quite well explained by an imperfect specialization Heckscher-Ohlin model of trade in homogeneous goods.

Our analysis also sheds new light on the empirical relevance of the IRS trade models in general. The influential finding of Helpman (1987), that key implications of the IRS model are consistent with trade data for Organization for Economic Co-operation and Development (OECD) countries, has been called into question. Hummels and Levinsohn (1995) repeated Helpman's analysis with a set of non-OECD countries, whose bilateral trade was not expected to contain much IRS-based trade. They showed that many correlations found by Helpman continue to hold. Rather than providing broad evidence against IRS trade models and, in particular, that these models cannot be responsible for the empirical success of the gravity equation (see Hummels and Levinsohn 1995, p. 828), their findings highlight the importance of the model identification problem. This paper accounts for both Helpman's and Hummels and Levinsohn's findings by suggesting that Helpman obtained his results because IRS-based trade is important among OECD countries, whereas Hummels and Levinsohn found a similar correlation among non-OECD countries because of specialization driven by differences in factor proportions. Our results underscore the importance of both relative factor abundance and IRS as determinants of the extent of specialization and international trade, with their relative importance depending on the particular sample in question. This is in line with the results of Antweiler and Trefler (1997), who estimate IRS and CRS models nested in the Heckscher-Ohlin-Vanek factor service trade expression. They find that although a majority of industries seem to be well characterized by the assumption of CRS, there is evidence for IRS in a number of sectors.

The remainder of the paper is organized as follows. We present the gravity equation predictions of four different trade models in Section I. Section II describes in more detail how we identify a particular model in our empirical analysis. Section III discusses the data set employed in this study, and Section IV presents the empirical results. Section V presents conclusions. The sensitivity of our results is discussed in the Appendix.

I. The Gravity Equation: Theories

A. The Gravity Equation with Perfect Specialization of Production

Throughout the paper, we assume that there is balanced trade, zero trade and transport costs, no trade in intermediate goods, and two countries with identical production technologies and that consumers in both countries have identical homothetic preferences. First, we consider a typical IRS model as laid out in Helpman and Krugman (1985, chap. 8.1), where there are two countries, $c = i, j$, and two goods, $g = X, Z$. Both X and Z come in many differentiated varieties, which are identically produced with increasing returns to scale. With preferences valuing product variety, a country will demand all foreign varieties according to the country's GDP as a share of world GDP. Let the GDPs of countries i and j be denoted Y^i and Y^j, respectively, and denote the GDP of the world by Y^w. Given that IRS leads to perfect specialization of production for every variety, it can be shown that country i's imports from j, denoted M^{ij}, are given by

$$M^{ij} = \frac{Y^i Y^j}{Y^w}. \tag{1}$$

Here imports are strictly proportional to the GDPs, and so this IRS trade model is a potential candidate to explain the success of the gravity equation.[1]

Equation (1) is very general since its derivation does not require assumptions on factor price equalization, differences in factor endowments across countries, factor intensities in the production of goods X and Z, or even the number of sectors, factors, and countries (Helpman and Krugman 1985). Equation (1) holds whenever there is perfect specialization in equilibrium, all consumers face the same goods prices and have identical homothetic preferences, and trade is balanced. Also, when international differences in factor endowments are large enough across countries (compared to differences in the goods' factor input requirements), perfect specialization can result. Thus a Heckscher-Ohlin model can also generate the gravity equation (1). In the two-country, two-good, two-factor ($2 \times 2 \times 2$) case, this requires that the countries' relative endowment ratios differ by at least as much as the goods' relative input ratios, which are consistent with diversified production and factor price equalization through trade. In the empirical analysis below, we shall refer to this as the multicone Heckscher-Ohlin model.

B. Gravity Equations with Imperfect Specialization of Production

First, we present the gravity equation for a model in which one sector (Z) produces a homogeneous good under CRS, whereas a second sector (X) produces a differentiated good under IRS (Helpman [1981]; see also Helpman and Krugman [1985, chaps. 7, 8]). This model will be referred to as the IRS/unicone Heckscher-Ohlin model. There are two countries (i and j) and two factors, capital (K) and labor (L). Assume that the homogeneous good is more labor-intensive in production and that country i is capital-abundant. Let γ^c be the share of good Z in country c's GDP, $\gamma^c = Z^c/(p_x X^c + Z^c)$, where p_x is the relative price of good X. For endowments in the two countries that are sufficiently similar that factor price equalization through trade can be achieved, country i exports only the capital-intensive X varieties. Country i's share of X varieties in GDP is equal to $1 - \gamma^i$, so the value of production on which country j draws for its imports is given by $(1 - \gamma^i)Y^i$. The assumption of homothetic preferences means that country j purchases X varieties from abroad according to its share in world GDP, which is equal to Y^j/Y^w.

[1] This gravity prediction, as well as those below, is readily generalized to the case in which tradable goods make up a share λ, $0 \leq \lambda \leq 1$, of GDP that is common to all countries. In that case, one obtains $M^{ij} = \lambda(Y^i Y^j/Y^w)$.

Together with balanced trade this means that country i's imports from j are given by

$$M^{ij} = (1 - \gamma^i)\frac{Y^iY^j}{Y^w}. \tag{2}$$

For any value $\gamma^i > 0$, the level of bilateral imports is lower than in the case in which both goods are differentiated (compare eqq. [2] and [1]). Furthermore, as the share of homogeneous goods in GDP declines, the predicted level of imports rises; in the limit, as $\gamma^i \to 0$, the generalized gravity equation (2) reverts to the simple gravity equation (1) above. Therefore, the higher the volume of trade, the lower the share of homogeneous goods in GDP.[2]

Second, we present the particular form of the gravity equation in the case of a $2 \times 2 \times 2$ unicone Heckscher-Ohlin model, where both good Z and good X are homogeneous and are produced under CRS in both countries. If country i is relatively capital-abundant and good Z is relatively labor-intensive, country i's imports from j are equal to $p_x[X^i - (Y^i/Y^w)X^w]$, where X^w is the world's production of good X. Using the definition of γ and the fact that $Y^i/Y^w = 1 - (Y^j/Y^w)$ leads to

$$M^{ij} = [(1 - \gamma^i) - (1 - \gamma^j)]\frac{Y^iY^j}{Y^w} = (\gamma^j - \gamma^i)\frac{Y^iY^j}{Y^w}. \tag{3}$$

As the capital-labor ratios in the two countries converge, so do γ^j and γ^i. In the limit, when the factor proportions in i and j are equal, we have that $\gamma^j = \gamma^i$, in which case equation (3) gives the familiar result that there is no trade in a Heckscher-Ohlin model when factor proportions are identical. Equation (3) includes as a special case the prediction about the volume of imports of the multicone Heckscher-Ohlin model given in (1) because as differences in factor proportions between i and j increase, the share of good Z in country j's GDP, γ^j, approaches one, whereas the share of good Z in the GDP of country i, γ^i, tends to zero.

If we denote the prediction about the volume of imports for the case in which the production of both goods is perfectly specialized, equation (1), by M_S; the IRS/unicone Heckscher-Ohlin case with specialization for one good, but not for the other good (eq. [2]), by M_{IH}; and the unicone Heckscher-Ohlin case in which both countries produce both

[2] This finding is in part due to Heckscher-Ohlin reasons because γ^i is inversely related to country i's capital-labor ratio. A decline in γ^i therefore implies an increase in the volume of imports due to an increase in the countries' differences in factor proportions. What we emphasize is that for given differences in factor proportions, the more specialization there is, the higher the level of imports; see below.

goods, equation (3), by M_H, the following inequalities hold, ceteris paribus:

$$M_S > M_{IH} > M_H. \tag{4}$$

This confirms that the bilateral volume of imports is higher the more product specialization there is. In summary, the observable implications of these trade models are as follows. Both perfect specialization models predict that bilateral imports are directly proportional to the product of the GDPs. For the IRS/unicone Heckscher-Ohlin model, imports are less than proportional to this product, and the extent of the shortfall depends on the size of the differentiated goods sector in GDP. All else equal, the unicone Heckscher-Ohlin model predicts the smallest factor of proportionality of imports with GDP, and this factor is predicted to rise when differences in factor proportions across trade partners are larger.

II. Model Identification

In the preceding section, we discussed the specific form of the gravity equation of trade for four models. From now on, we consider only $2 \times 2 \times 2$ models. It is unlikely, however, that any of the observed trade flows are solely determined by any one of these four archetypical models. First, the data come from a world with more sectors, countries, and factors than exist in our $2 \times 2 \times 2$ models. Second, there may be positive amounts of IRS-based trade even between countries with the lowest recorded shares of IRS-based trade, and likewise for factor proportions–based trade. However, in different circumstances (such as different degrees of product differentiation in the goods that are traded), we expect that the volume of trade is better accounted for by one particular model than another. This is the basis for our model identification strategy.

Consider a cross section of country pairs in which there is little (or no) specialization due to IRS, but the absolute difference between the two countries' factor proportions, denoted FDIF, varies from one pair to another. If Heckscher-Ohlin forces are at work, then we expect, ceteris paribus, more specialization in a pair in which FDIF is larger than in another country pair in which FDIF is smaller. This observation allows us to identify the Heckscher-Ohlin motivation for specialization and the gravity prediction.

We employ the index proposed by Grubel and Lloyd (1975) to indicate the extent of IRS-based trade. As a result of trade due to IRS and product differentiation, a country simultaneously imports and exports varieties of a particular product (intraindustry trade).[3] The index, denoted GL^{ij}, measures the share of intraindustry trade in total trade:

$$GL^{ij} = 1 - \left[\frac{\sum_g |M_g^{ij} - M_g^{ji}|}{\sum_g (M_g^{ij} + M_g^{ji})}\right], \quad 0 \leq GL^{ij} \leq 1. \tag{5}$$

In the extremes, when every good g is either exported or imported (no intraindustry trade), the Grubel-Lloyd index will be equal to zero, whereas it is equal to one if a country's imports and exports of all goods are equal (only intraindustry trade).

This index is not a perfect indicator for the share of trade based on IRS. Finger (1975) has emphasized that industry classifications often contain quite different products, and clearly a high degree of disaggregation is desirable when measuring intraindustry trade. Recently, researchers have emphasized the importance of trade in intermediate goods in accounting for high values of GL^{ij} (Greenaway, Hine, and Milner 1994; Fontagne, Freudenberg, and Peridy 1998), which can in principle be due to IRS and other reasons. Moreover, a significant part of trade that is classified as interindustry trade might involve IRS, such as wide-bodied aircraft exports from the United States to most other countries.[4] After discussing this issue in some detail, Krugman (1994, p. 23) concludes that although GL^{ij} does not precisely measure the share of international trade that is due to IRS, there is no indication that there is an important upward or downward bias.

We expect that the IRS model accounts for the performance of the gravity equation in those samples in which the bilateral Grubel-Lloyd indices are relatively large, indicating that a relatively large proportion of bilateral trade is two-way trade in differentiated products. In Section IV*A*1, we examine whether in fact the prediction M_S (of the IRS model) is accurate in samples with relatively high Grubel-Lloyd indices. We emphasize that prediction M_S is common to both the multicone Heckscher-Ohlin and the IRS models. However, should we find that the prediction M_S is not rejected by the data in samples with relatively high Grubel-Lloyd indices, then it would be incorrect to interpret this finding as evidence in favor of the multicone Heckscher-Ohlin theory, which predicts no intraindustry trade.

In the following section, we briefly discuss the data that will be used.

[3] We define intraindustry trade as trade in goods with identical factor input requirements; for our empirical analysis, though, intraindustry trade is taken as two-way trade of goods in the same four-digit Standard International Trade Classification (SITC) class.

[4] The only important producer of such aircraft (made by Boeing) outside the United States is the European joint venture Airbus, so that Boeing exports to any country other than the Airbus maker countries are classified as interindustry trade. This is an example given in Krugman (1994).

III. Data

We have assembled a cross-sectional data set for 58 countries in the year 1985. The data set includes all countries with GDPs above U.S.$1 billion and in which internationally comparable estimates of capital per worker are available. These 58 countries accounted for 67 percent of world imports and 79 percent of world GDP. The data set includes nearly all the industrialized countries but relatively fewer of the less developed countries, reflecting the paucity of available capital stock estimates for the latter (see table 1).

The source of data on GDP and capital per worker is the Penn World Table (version 5.6) database (see Summers and Heston 1991). Both variables are reported in internationally comparable and purchasing power–adjusted real U.S. dollars. The total value of bilateral imports data comes from the NBER World Bilateral Trade Database (see Feenstra, Lipsey, and Bowen 1997). The quality of the data is likely to be lower for the poorer countries in the sample. This differential degree of measurement error requires caution when results for highly industrialized countries are compared with those obtained from samples that include poorer countries.

The bilateral Grubel-Lloyd indices, GL^{ij}, are calculated using data on all goods trade at the four-digit SITC, and therefore, in contrast to most reported GL^{ij} indices, we do not confine ourselves to manufacturing goods trade. The Grubel-Lloyd index can be calculated only for country pairs in which there are positive amounts of trade, which is the case for 2,870 observations, or 87 percent of the possible 3,306 bilateral import relations among 58 countries. The number of bilateral import relations varies across countries from a low of 27 for Nepal to the maximum of 57 for most industrialized countries. Furthermore, the average number of four-digit SITC categories that a country trades bilaterally varies considerably as well. At one end, Mauritius trades, on average, 36 categories with its partners, whereas at the other end, the United States trades, on average, 332 categories of goods with its trading partners.

Across each of their respective trading partners, Bolivia has the lowest average Grubel-Lloyd index, with a value of .0006; the United Kingdom has the highest value of .1495. Each country's average Grubel-Lloyd index is shown in table 1, along with the country's GDP per capita, the number of bilateral relations the average is computed from, and the average number of goods categories in which there is trade. In table 2 we report the correlation coefficients among the variables of table 1. They are all positive; in particular, the correlation of the average Grubel-Lloyd index with GDP per capita is .78, and the correlation between the average Grubel-Lloyd index and the average number of goods traded is .93. The distribution of bilateral Grubel-Lloyd indices is very skewed:

TABLE 1
SUMMARY STATISTICS

Country	GDP per Capita ($ U.S.)	GDP (Billion $ U.S.)	Capital per Worker ($ U.S.)	Number of Bilateral Relations	Average Number of Industry Classes Traded	Average Grubel-Lloyd Index
United States	16,570	3,964.9	29,925	57	332	.1371
Canada	15,589	392.3	34,535	57	190	.055
Switzerland	14,864	96.2	62,769	57	209	.1227
Norway	14,144	58.7	44,866	57	142	.0537
Australia	13,583	214	33,875	56	151	.0306
Sweden	13,451	112.3	31,326	57	198	.1163
Denmark	12,969	66.3	29,286	57	181	.0923
West Germany	12,535	765.4	47,695	57	323	.1473
Belgium-Luxembourg	12,230	116.1	33,494.5	57	221	.1347
Iceland	12,209	2.9	15,991	44	84	.0064
France	12,206	673.4	31,796	57	282	.1377
Finland	12,051	59.1	38,591	57	136	.0625
Japan	11,771	1,421.4	28,106	57	244	.064
Netherlands	11,539	167.2	29,325	57	255	.1283
New Zealand	11,443	37.4	30,511	53	106	.0189
United Kingdom	11,237	636.2	17,636	57	307	.1495
Austria	11,131	84.1	29,708	57	171	.1031
Italy	10,808	617.6	28,117	57	265	.1111
Hong Kong	10,599	57.8	11,220	57	166	.0805
Israel	8,310	35.2	22,307	52	119	.0721
Spain	7,536	29.7	21,831	57	209	.0807
Ireland	7,275	25.8	20,700	56	117	.0746
Venezuela	6,225	106.3	20,419	48	91	.0078

Greece	6,224	61.8	22,151	55	112	.0304
Mexico	5,621	420.3	14,016	52	100	.0239
Taiwan	5,449	104.9	19,194	57	167	.0569
Argentina	5,324	161.5	12,084	52	92	.0201
Portugal	5,070	51.5	9,503	56	111	.0473
Syria	4,240	43.9	15,681	41	70	.0011
Mauritius	4,226	4.3	2,393	35	36	.0106
South Korea	4,217	172.1	12,036	55	139	.0644
Iran	4,043	187.5	10,686	41	81	.0033
Panama	3,499	7.6	16,417	45	73	.0074
Chile	3,467	42	7,060	48	100	.0081
Turkey	3,077	154.8	6,989	49	103	.0195
Colombia	2,968	87.5	12,790	48	85	.0092
Ecuador	2,913	27.1	15,536	44	67	.0031
Peru	2,565	49.7	9,480	48	90	.0077
Thailand	2,463	127.3	4,051	53	114	.0286
Jamaica	2,215	5	3,725	41	47	.005
Dominican Republic	2,111	13.5	5,282	43	56	.012
Guatemala	2,090	16.6	3,985	45	57	.0079
Paraguay	2,072	7.7	849	37	42	.0021
Sri Lanka	2,045	32.4	8,238	54	66	.0186
Morocco	1,956	43.2	2,803	48	76	.0067
Bolivia	1,754	11.1	6,987	37	49	.0006
Ivory Coast	1,545	15.1	1,171	51	57	.0076
Philippines	1,542	84.4	4,087	55	82	.024
Honduras	1,387	6.1	4,507	41	51	.0028
Zimbabwe	1,216	10.2	5,750	48	52	.0033
Nigeria	1,062	88.4	1,103	44	83	.0024
India	1,050	803.4	1,712	57	115	.0305
Nepal	936	15.6	758	27	41	.0031
Sierra Leone	905	3.3	215	34	38	.0029
Zambia	808	5.5	1,785	37	53	.0053
Kenya	794	16.1	1,187	43	75	.0052
Madagascar	769	7.7	1,714	34	40	.0059

TABLE 1
(*Continued*)

Country	GDP per Capita ($ U.S.)	GDP (Billion $ U.S.)	Capital per Worker ($ U.S.)	Number of Bilateral Relations	Average Number of Industry Classes Traded	Average Grubel-Lloyd Index
Malawi	518	3.7	469	37	38	.0028
Average	6,248.55	222.31	16,214.2	49.48	123.4	.0427
Median	4,233	58.9	12,437	52	100	.0198
Standard deviation	4,934.91	563.33	14,136.79	8.05	78.57	.0469

TABLE 2
CORRELATION MATRIX

	GDP per Capita	GDP	Capital per Worker	Number of Bilateral Pairs	Average Number of Industries	Average Grubel-Lloyd Index
GDP per capita	1					
GDP	.41	1				
Capital per worker	.89	.26	1			
Number of bilateral pairs	.68	.31	.66	1		
Average number of industries	.81	.6	.73	.74	1	
Average Grubel-Lloyd index	.78	.45	.73	.71	.93	1

44 percent of all indices are equal to zero, and 78 percent of all 1,120 country pairs have a value of .05 or less. In the empirical analysis below, we shall treat these 1,120 country pairs—or 2,240 observations—as trade relations in which there is little or no IRS-based trade. The remaining 315 country pairs (where $GL^{ij} > .05$) are those in which IRS-based trade is thought to be present; they account for 87.1 percent of all the imports in our sample.

In the IRS/unicone Heckscher-Ohlin model, the bilateral imports prediction depends on whether the differentiated good is relatively capital- or labor-intensive. We follow Helpman and Krugman (1985) and others in assuming that the differentiated good is relatively capital-intensive. However, we have examined this question in some detail and concluded that the assumption may not hold empirically.[5] For this rea-

[5] Rauch (1999) has computed the extent to which an industry produces goods that are reference priced on organized exchanges, denoted *r*. It is plausible to assume that this is *inversely* related to the degree of product differentiation. We have correlated *r* with industry-level data on capital-labor intensity in production from seven major industrialized countries (from Keller [2002]) as well as with more disaggregated U.S. data (from Bartelsman and Gray [2000]), finding a zero or positive correlation. As an alternative measure of product differentiation, we have considered the standard deviation of unit import values, expecting a positive relationship (Austrian Institute for Economic Research 1999). This variable is not significantly related to the ratio of capital to labor share in value added across industries (data from Peneder [1999]). Because of limited data availability, this analysis covers only manufacturing products (details are available from the authors on request). The findings may well change when nonmanufacturing sectors and broader definitions of capital, especially including research and development, are adopted. However, within manufacturing, there are a number of highly capital-intensive industries that produce homogeneous goods (e.g., petrochemicals) as well as labor-intensive industries that produce fairly differentiated goods (e.g., apparel), explaining the mixed results.

son we also consider the alternative case in our analysis below, in which the differentiated good is relatively labor-intensive. Finally, the data on bilateral distance between countries used below are the smallest arc tan distance between capital cities, measured in kilometers (source: Haveman 1998).

IV. Empirical Results

We split the 2,870 bilateral imports observations into two subsamples according to an arbitrarily chosen critical level of intraindustry trade, denoted $\overline{GL}$. Those pairs in which $GL^{ij} \leq \overline{GL}$ belong to what is referred to as the low-Grubel-Lloyd sample, and the remaining observations are part of the high-Grubel-Lloyd sample. The samples differ in that we expect substantial amounts of trade based on product differentiation and IRS in the latter but not in the former. Given the critical level $\overline{GL}$, it will be useful to sort the data, and form classes, according to each observation's differences in factor proportions. Likewise, we sort the observations in the high-Grubel-Lloyd sample according to their level of intraindustry trade (measured by GL). Denote by V the number of classes into which the sorted observations are allocated.

In the benchmark case that is discussed first, $\overline{GL}$ is chosen to equal .05, and there are $V = 5$ classes of equal size. We then present results for two alternative cases. In one we adopt a critical level of $\overline{GL} =$ .033, which shifts more bilateral trade relations into the sample in which product differentiation/IRS-based trade is thought to be present. In the other case, we report results from regressions in which the class size is endogenously chosen so as to maximize the fit of the regression. This is complemented by a number of additional results and sensitivity analyses reported in the Appendix.

A. *Benchmark Case*

The choice of $\overline{GL} = .05$ results in 2,240 observations in the low-Grubel-Lloyd sample and the remaining 630 observations in the high-Grubel-Lloyd sample. We first discuss results for the two perfect specialization models before we turn to the two imperfect specialization models.

1. IRS-Based Perfect Specialization (High-Grubel-Lloyd Sample)

The model considered first is the IRS model, where both goods are differentiated and produced with scale economies. As discussed above, the model has a prediction of volume of imports of $M^{ij} = Y^i Y^j / Y^w$. To ensure comparability with the results for the imperfect specialization models shown below, we allocate the 630 observations into $V = 5$ classes, with GL^{ij} rising from class $v = 1, \ldots, V$. The following equation, based on (1), is estimated:

$$M_v^{ij} = \alpha_v \frac{Y_v^i Y_v^j}{Y^w} + \epsilon_v^{ij} \tag{6}$$

for each v. The disturbance ϵ_v^{ij} is assumed to have mean zero but might be heteroskedastic. In this and all following regressions, we assume that the variance of ϵ^{ij} is equal to $\sigma^2(\mathbf{Z}^{ij})^2$, where $\mathbf{Z}^{ij}$ is a vector that includes five variables: Y^i, Y^j, the sums of imports of country i and exports of country j, and the bilateral distance between i and j (denoted D^{ij}). Let the predicted values from regressing $\log|\hat{\epsilon}^{ij}|$ on a constant and $\log \mathbf{Z}^{ij}$ be given by $\log \tilde{\epsilon}^{ij}$; the parameter α is then estimated by ordinary least squares (OLS) with the scaled variables $M^{ij}/\tilde{\epsilon}^{ij}$ and $(Y^iY^j/Y^w)/\tilde{\epsilon}^{ij}$.[6] The results for these regressions can be found in table 3.

The estimated α_v's vary in the range from 0.016 to 0.139, where the largest import parameter is estimated for an intermediate level of product differentiation ($v = 3$). The table also reports standard errors (in parentheses) and the symmetric 90 percent confidence intervals of the point estimates.[7] In addition, we test the hypotheses that the import parameters of all five classes are identical and whether $\alpha_1 = \alpha_5$ (a 10 percent significance level is adopted for all hypothesis tests). Although the estimated α_v's are larger for classes with higher levels of product differentiation, it is clear that all the α_v estimates fall well short of the theoretically predicted value of one at standard levels of significance. The sixth row of table 3 shows the results of using all 630 observations. With an α estimate of 0.087, this regression confirms that the IRS model overpredicts the level of bilateral trade across countries substantially. We now turn to the multicone Heckscher-Ohlin model.[8]

[6] Moreover, the gravity equation is potentially a relationship between two endogenous variables, so the results might be affected by simultaneity bias (e.g., Saxonhouse 1989). We have also considered regressions with imports over the product of the GDPs as the dependent variable. This leads to the same qualitative results (see App. table A2). See, e.g., Harrigan (1996), who pursues this issue further.

[7] The standard errors and confidence intervals are based on bootstrap techniques (see Bickel and Freedman 1981; Efron 1982); they tend to be larger and appear to be preferable in this case to those obtained from standard asymptotic results.

[8] Equation (6) does not include a constant, which means that OLS estimates are inconsistent if $E[\epsilon_v^{ij}] \neq 0$, contrary to the assumption. While the mean of the fitted ϵ^{ij} is typically small relative to its standard deviation, we have nevertheless estimated all four models also with a constant. The pattern of how α_v varies with v remains the same (see App. table A1).

TABLE 3
BENCHMARK CASE

	IRS/Heckscher-Ohlin Model: High-Grubel-Lloyd Sample ($GL > .05$)				Heckscher-Ohlin Model: Low-Grubel-Lloyd Sample ($GL < .05$)			
	IRS Model (IRS/IRS Goods)		IRS/Unicone Heckscher-Ohlin Model (IRS/CRS Goods)		Multicone Heckscher-Ohlin Model (CRS/CRS Goods)		Unicone Heckscher-Ohlin Model (CRS/CRS Goods)	
	α_v	5% 95%	α_v	5% 95%	α_v	5% 95%	α_v	5% 95%
	Ranked by Grubel-Lloyd Index				Ranked by FDIF			
$v=1$	.016	.012	.078	.072	.039	.030	.021	.012
	(.012)	.044	(.005)	.087	(.007)	.049	(.004)	.026
$v=2$	.044	.036	.053	.047	.111	.087	.027	.025
	(.005)	.052	(.005)	.060	(.014)	.132	(.008)	.043
$v=3$	.139	.120	.117	.112	.047	.040	.058	.039
	(.013)	.164	(.009)	.141	(.005)	.056	(.008)	.066
$v=4$	.069	.049	.123	.109	.039	.034	.048	.046
	(.017)	.097	(.005)	.124	(.003)	.044	(.006)	.064
$v=5$	.099	.083	.128	.119	.039	.033	.080	.069
	(.015)	.125	(.006)	.134	(.004)	.045	(.007)	.101
All observations	.087	.076	.086	.079	.052	.047	.040	.034
	(.009)	.104	(.004)	.092	(.003)	.056	(.003)	.044
Only perfect specialization of production	yes		no		yes		no	
H_0: $\alpha_i = \alpha\ \forall i$	reject		reject		reject		reject	
H_0: $\alpha_1 = \alpha_5$	reject		reject		do not reject		reject	
Share of bilateral comparisons correct	N.A.		9/10		N.A.		9/10	

NOTE.—Standard errors are in parentheses.

2. Heckscher-Ohlin-Based Perfect Specialization (Low-Grubel-Lloyd Sample)

Recall that in the multicone Heckscher-Ohlin model, all trade is in homogeneous, perfectly specialized products. The imports prediction is given by equation (1): $M^{ij} = Y^iY^j/Y^w$. Therefore, the only parameter α_v is predicted to equal one. We estimate

$$M_v^{ij} = \alpha_v \frac{Y_v^i Y_v^j}{Y^w} + \epsilon_v^{ij} \quad (7)$$

for each v, $v = 1, \ldots, 5$. Differences in factor proportions are lowest for $v = 1$ and highest for $v = 5$; with 2,240 observations in the low-Grubel-Lloyd sample, each class has 448 observations. Table 3 shows the results.

The largest parameter, with 0.111, is estimated for class $v = 2$, whereas the α estimates for the four other classes are between 0.039 and 0.047. All parameters are less than the theoretically predicted value of one at standard levels of significance. In addition, when the entire low-Grubel-Lloyd sample is used to estimate the coefficient α, the resulting parameter estimate is 0.052 (see the sixth row of table 3). Thus, in this sample, the multicone Heckscher-Ohlin model grossly overpredicts the volume of bilateral trade.

In summary, the predictions of the multicone Heckscher-Ohlin model are rejected by the data. The evidence for the IRS model, which overpredicts bilateral trade somewhat less (with $\alpha = 0.087$ in the regression with all observations), is slightly stronger. Care must be taken, though, not to overinterpret this small difference since, as we have noted above, the data quality for the high-Grubel-Lloyd sample is higher than for the low-Grubel-Lloyd sample. The main result from our analysis thus far is that models of perfect specialization do not appear to be important in explaining the success of the gravity equation. This might in part be related to the fact that the prediction of trade volume of both of these models holds only if the production of all goods is perfectly specialized, which is unlikely. We now turn to estimating the predictions of import volume of the two models that incorporate imperfectly specialized production, with import prediction M_{IH} for the IRS/unicone Heckscher-Ohlin model and prediction M_H for the unicone Heckscher-Ohlin model.

3. IRS Model Incorporating Imperfect Specialization (High-Grubel-Lloyd Sample)

The model considered here is the IRS/unicone Heckscher-Ohlin model. The regression equation when country i is capital-abundant relative to j is

$$M_v^{ij} = (1 - \gamma_v^i)\frac{Y_v^i Y_v^j}{Y^w} + \epsilon_v^{ij}$$

for each v, $v = 1, \ldots, 5$, and

$$M_v^{ij} = (1 - \gamma_v^j)\frac{Y_v^i Y_v^j}{Y^w} + \epsilon_v^{ij}$$

if country j is relatively capital-abundant. The number of estimated parameters $(1 - \gamma_v^i)$ varies by class v. Moreover, the number of times that a given country's parameter is estimated in a given class changes as well. In table 3 we report the median of the predicted values $(\widehat{1 - \gamma_v^i})$, which is weighted to take into account the number of observations that are used to estimate any one parameter $(1 - \gamma_v^i)$. This statistic is denoted by α_v.

The median predicted value varies across classes from 0.053 to 0.128. A nonstructural interpretation of α_v is that it estimates the average size of the differentiated goods sector. This average should, first, be nonnegative, and the results confirm that. Second, in this model, higher values of the Grubel-Lloyd index (and hence class v) should be associated with higher estimates of α_v because a higher share of differentiated goods in GDP gives rise to a higher share of intraindustry trade in total trade. The α_v estimates rise from 0.078 for $v = 1$ to 0.128 for $v = 5$, with a nonmonotonicity at $v = 2$ ($\alpha_2 = 0.053$). This pattern is broadly in accordance with the model's prediction. For the two imperfect specialization models, we are interested in seeing whether α_v increases as v increases, that is, whether the point estimate is greater in classes in which the extent of specialization should be greater. Thus we report as an additional, albeit rough, indicator of model performance the number of bilateral comparisons of α_v's that are in line with these priors. That is, we ask whether α_2 exceeds α_1, whether α_3 is larger than α_1, whether α_3 exceeds α_2, and so on. As indicated in the table, only one out of 10 bilateral comparisons of α_v's is incorrect for the unicone/IRS model.

The confidence intervals suggest that some but not all differences of import parameter estimates across Grubel-Lloyd classes are also statistically significant. In particular, the median import parameter for intermediate levels of the Grubel-Lloyd index ($v = 3$) is statistically indistinguishable from that for the highest Grubel-Lloyd class. At the same time, we strongly reject both the null hypothesis that the five α_v are the same and the null that $\alpha_1 = \alpha_5$. Overall, this provides some support for the IRS/unicone Heckscher-Ohlin model. We now turn to the unicone Heckscher-Ohlin model.

4. Heckscher-Ohlin-Based Imperfect Specialization (Low-Grubel-Lloyd Sample)

We return to the sample of bilateral pairs in which the computed Grubel-Lloyd index is .05 or less. Here we estimate the unicone Heckscher-Ohlin model, where two homogeneous goods are produced in both countries. As discussed in Section I above, the model's import prediction M_H is equal to $M^{ij} = (\gamma^j - \gamma^i)(Y^iY^j/Y^w)$ if country i is capital-abundant. We estimate

$$M_v^{ij} = (\gamma_v^j - \gamma_v^i)\frac{Y_v^iY_v^j}{Y^w} + \epsilon_v^{ij}$$

for each v, $v = 1, \ldots, 5$, and

$$M_v^{ij} = (\gamma_v^i - \gamma_v^j)\frac{Y_v^i Y_v^j}{Y^w} + \epsilon_v^{ij}$$

if country j is capital-abundant. In each class v, there are up to 58 countries. Out of 58 parameters γ, only 57 are identified. We arbitrarily set the γ for the country with the lowest capital-labor ratio in the sample (Sierra Leone) equal to one in each class v. Because the composition of the classes varies in terms of countries, we also report here the median of the predicted (net) import parameter, given by $\widehat{\gamma_v^j - \gamma_v^i}$ and denoted by α_v. A nonstructural interpretation is that it represents the difference between the size of the labor-intensive sector in the labor- and in the capital-abundant country. We expect that α_v is positive and that it tends to rise when bilateral differences in factor proportions increase.

Table 3 shows that α_v for the unicone Heckscher-Ohlin model is estimated to lie between 0.021 and 0.080, which together with the confidence intervals suggests that α_v is positive for all classes. The median value of $\widehat{\gamma^j - \gamma^i}$ when all 2,240 observations are used in the estimation is 0.04 (sixth row). Further, α_v is generally rising with differences in factor proportions across classes, although there is one nonmonotonicity at class $v = 4$. Nine out of 10 bilateral comparisons are correctly predicted by the unicone Heckscher-Ohlin model. The confidence intervals also allow us to reject the null hypothesis that all α_v's are identical as well as the hypothesis that $\alpha_1 = \alpha_5$. Thus we find support in favor of the unicone Heckscher-Ohlin model. Interestingly, although we find a nonmonotonicity for α_v at a relatively high FDIF level ($v = 4$), the unicone Heckscher-Ohlin model works well even for the class with the largest factor proportion differences. A priori, one might have thought that the assumption of a common cone of production diversification is then implausible. We shall take up this issue again in subsection 6.

In summary, from this analysis of these imperfect specialization models, we find evidence supportive of both the IRS/unicone Heckscher-Ohlin model and the unicone Heckscher-Ohlin model.

5. Comparing the Perfect and Imperfect Specialization Models in Terms of Fit

In this subsection we examine whether the perfect and imperfect specialization models differ considerably in terms of empirical fit. Table 4 shows a number of widely used model selection criteria for the benchmark case. The R^2 is computed using the expression $\hat{y}'\hat{y}/y'y$, where $\hat{y}$ and y are the deviations from the means of the predicted values of the

TABLE 4
MEASURES OF FIT FOR THE BENCHMARK CASE

	IRS/HECKSCHER-OHLIN MODEL: HIGH-GRUBEL-LLOYD SAMPLE ($GL > .05$)						HECKSCHER-OHLIN MODEL: LOW-GRUBEL-LLOYD SAMPLE ($GL < .05$)					
	IRS Model (IRS/IRS Goods)			IRS/Unicone Heckscher-Ohlin Model (IRS/CRS Goods)			Multicone Heckscher-Ohlin Model (CRS/CRS Goods)			Unicone Heckscher-Ohlin Model (CRS/CRS Goods)		
	ln ($e'e$)	AIC	R^2	ln ($e'e$)	AIC	R^2	ln ($e'e$)	AIC	R^2	ln ($e'e$)	AIC	R^2
	Ranked by Grubel-Lloyd Index						Ranked by FDIF					
$v=1$	44.61	39.79	1.083	43.74	39.30	.832	43.36	37.26	.238	43.24	37.37	.712
$v=2$	45.21	40.39	.456	44.93	40.46	.619	45.61	39.51	.455	45.25	39.38	.605
$v=3$	48.81	43.99	.958	48.75	44.26	.674	44.72	38.62	.417	44.24	38.38	.484
$v=4$	47.17	42.34	.917	46.45	41.93	.869	44.45	38.35	1.963	43.88	38.03	.793
$v=5$	50.61	45.79	.214	49.61	45.04	.934	44.32	38.22	.596	43.84	37.95	.690
All observations	50.91	44.46	.280	51.03	44.69	.105	46.74	39.03	.451	46.96	39.30	.195
Only perfect specialization of production		yes			no			yes			no	

independent variable and the actual value of the independent variable, respectively. Akaike's Information Criterion (AIC) is defined as

$$\text{AIC} = \ln\left(\frac{e'e}{N}\right) + \frac{2Q}{N},$$

where $e'e$ is the residual sum of squares, N is the number of observations, and Q is the number of estimated parameters. Also the unadjusted log of $e'e$ is reported. In the regression with all 630 observations, the R^2 for the IRS model is .280, whereas for the IRS/unicone Heckscher-Ohlin model, it is .105.[9] Also, according to the AIC, the perfect specialization model is preferred to the imperfect specialization model in the high-Grubel-Lloyd sample. The same qualitative results are contained for the two other models in the low-Grubel-Lloyd sample.

Overall, it appears that our analysis of empirical fit does not lead to major conclusions, except perhaps that the reduction in the residual sum of squares that the imperfect specialization models generate is small relative to the additional number of parameters that are estimated. We conclude the analysis of the benchmark case by examining the country-specific estimates of the two imperfect specialization models.

6. Country-Specific Estimates of the Imperfect Specialization Models

In this subsection we discuss the estimates of $1 - \gamma^i$ and γ^i from the IRS/unicone Heckscher-Ohlin and unicone Heckscher-Ohlin models, respectively. The results are collected in table 5. In the case of the IRS/unicone Heckscher-Ohlin model, 35 country-specific parameters are estimated. The remaining 23 countries have no trade relations in which the level of intraindustry trade exceeds the critical value of $\overline{GL} = .05$. These countries are primarily developing countries, as table 5 shows. The parameter $1 - \gamma^i$ denotes the share of the differentiated goods sector in GDP in the capital-abundant country, which should be bounded between zero and one. Moreover, under the assumption that the differentiated good is relatively capital-intensive, the model implies that the share of differentiated goods in GDP increases with the relative abundance of capital to labor, ceteris paribus. Under these conditions, one expects the correlation of K^i/L^i and $1 - \gamma^i$ to be positive. As reported in table 5, 32 out of 35, or 91 percent, of the $1 - \gamma^i$ estimates satisfy the restriction of being between zero and one. We estimate values above

[9] It is sometimes suggested to compute y and $\hat{y}$ as deviations from zero when the regression does not contain a constant (e.g., Kennedy 1992, p. 27). This changes little in the present context, though.

TABLE 5

Country Parameter Estimates for the Imperfect Specialization Heckscher-Ohlin Models

			IRS/Unicone Heckscher-Ohlin Model			Unicone Heckscher-Ohlin Model		
Country	GDP per Capita (1)	K/L (2)	$1-\gamma$ (3)	Standard Error (4)	p-Value (5)	γ (6)	Standard Error (7)	p-Value (8)
Sierra Leone	905	215				1.000	N.A.	N.A.
Malawi	518	469				1.002	.018	1.00
Nepal	936	758				.962	.018	.02
Paraguay	2,072	849				1.001	.016	1.00
Nigeria	1,062	1,103				1.007	.020	1.00
Ivory Coast	1,545	1,171				1.008	.018	1.00
Kenya	794	1,187				.999	.015	.96
India	1,050	1,712				.977	.015	.00
Madagascar	769	1,714				.980	.014	.14
Zambia	808	1,785	.022	.006	.00	1.001	.045	1.00
Mauritius	4,226	2,393				1.027	.032	1.00
Morocco	1,956	2,803				.976	.015	.04
Jamaica	2,215	3,725				1.009	.017	1.00
Guatemala	2,090	3,985				.986	.016	.36
Thailand	2,463	4,051				.976	.014	.02
Philippines	1,542	4,087				.976	.015	.00
Honduras	1,387	4,507	2.647	.685	1.00	1.009	.021	1.00
Dominican Republic	2,111	5,282	.211	.001	1.00	.986	.016	.28
Zimbabwe	1,216	5,750	6.415	1.521	1.00	.976	.016	.02
Bolivia	1,754	6,987				1.015	.050	1.00
Turkey	3,077	6,989				.981	.014	.22
Chile	3,467	7,060				1.005	.015	1.00
Sri Lanka	2,045	8,238	.017	.006	.00	.983	.016	.16
Peru	2,565	9,480				.982	.015	.12
Portugal	5,070	9,503				.969	.020	.04
Iran	4,043	10,686				.977	.020	.28
Hong Kong	10,599	11,220	.111	.020	1.00	.940	.033	.02
Republic of Korea	4,217	12,036	1.497	.227	1.00	.975	.020	.24
Argentina	5,324	12,084	.206	.006	1.00	.955	.021	.00
Colombia	2,968	12,790	.140	.018	1.00	.971	.015	.00
Mexico	5,621	14,016	.037	.008	.00	.973	.014	.00
Ecuador	2,913	15,536	.391	.025	1.00	.971	.018	.04
Syria	4,240	15,681				.964	.015	.00
Iceland	12,209	15,991				1.018	.034	1.00
Panama	3,499	16,417	.406	.163	1.00	.975	.034	.54
United Kingdom	11,237	17,636	.048	.006	.00	.903	.020	.00
Taiwan	5,449	19,194	.111	.020	1.00	.959	.019	.02
Venezuela	6,225	20,419	.015	.001	.00	.952	.020	.00
Ireland	7,275	20,700	.198	.035	1.00	.934	.023	.00
Spain	7,536	21,831	.049	.007	.00	.933	.016	.00
Greece	6,224	22,151	.076	.009	.04	.968	.015	.00
Israel	8,310	22,307	.030	.005	.00	.910	.023	.00
Japan	11,771	28,106	.072	.008	.06	.954	.016	.00
Italy	10,808	28,117	.021	.003	.00	.932	.017	.00
Denmark	12,969	29,286	.101	.018	1.00	.933	.017	.00
Netherlands	11,539	29,325	.065	.012	.08	.856	.024	.00
Austria	11,131	29,708	.080	.013	.38	.943	.015	.00
United States	16,570	29,925	.042	.006	.00	.953	.014	.00
New Zealand	11,443	30,511	.178	.024	1.00	.910	.022	.00
Sweden	13,451	31,326	.089	.012	.74	.923	.021	.00
France	12,206	31,796	.054	.004	.00	.925	.019	.00

TABLE 5 *(continued)*
COUNTRY PARAMETER ESTIMATES FOR THE IMPERFECT SPECIALIZATION HECKSCHER-OHLIN MODELS

			IRS/UNICONE HECKSCHER-OHLIN MODEL			UNICONE HECKSCHER-OHLIN MODEL		
COUNTRY	GDP PER CAPITA (1)	K/L (2)	$1-\gamma$ (3)	Standard Error (4)	*p*-Value (5)	γ (6)	Standard Error (7)	*p*-Value (8)
Belgium-Luxembourg	12,230	33,495	.098	.010	1.00	.855	.027	.00
Australia	13,583	33,875	.107	.010	1.00	.930	.019	.00
Canada	15,589	34,535	.076	.007	.10	.946	.016	.00
Finland	12,051	38,591	.143	.020	1.00	.930	.017	.00
Norway	14,144	44,866	.196	.026	1.00	.934	.019	.00
West Germany	12,535	47,695	.086	.005	.30	.906	.018	.00
Switzerland	14,864	62,769	.091	.010	N.A.	.874	.018	.00

one, which would mean a negative share of the homogeneous goods sector, for Honduras, Zimbabwe, and the Republic of Korea.

The table also reports the standard errors of the $1 - \gamma^i$ as well as *p*-values, both of which are computed using bootstrap techniques. The *p*-values indicate the probability that a given country's $1 - \gamma$ exceeds that of Switzerland, the country with the highest relative capital endowment in the sample. The estimate for Switzerland is equal to .091. The $1 - \gamma^i$ estimates often have relatively small standard errors; however, it is not the case that most of these estimates are smaller than 0.091: at a 10 percent level, only 13 out of 34, or 38 percent, of the parameters are significantly smaller than that of Switzerland. Thus it is not surprising that we do not find the positive correlation between $1 - \gamma^i$ and K^i/L^i that the model predicts. As reported in table 6, the simple correlation of $1 - \gamma^i$ and K^i/L^i is negative. The result is further confirmed by regressing the $1 - \gamma^i$ on a constant and K^i/L^i using OLS: the *t*-statistic of the slope parameter in this regression is equal to -2.12.[10] The analogous correlations with GDP per capita are not much different.

These estimates do not provide support for the IRS/unicone Heckscher-Ohlin model. Given our difficulty of establishing whether the differentiated good is relatively capital-intensive or not, we have also considered the alternative assumption, that the homogeneous good is relatively capital-intensive. It can be shown that, on the maintained assumption that country i is relatively capital-abundant, the import prediction becomes $M^{ij} = (1 - \gamma^j)(Y^iY^j/Y^w)$. That is, the regression parameter estimates country j's share of the differentiated good in GDP instead

[10] We have also weighted each observation by the inverse of the standard error of $1 - \gamma^i$; the results of that regression do not provide more support for the model. When the three cases in which $1 - \gamma^i$ exceeds one are dropped from the regression since they do not satisfy the restriction that $1 - \gamma^i$ is a share, the *t*-statistic drops to -0.60.

TABLE 6
Relation of Capital-Labor Endowment Ratio to $1-\gamma$ and γ

	IRS/Unicone Heckscher-Ohlin Model			Unicone Heckscher-Ohlin Model		
	Correlation	OLS *t*-Statistic	R^2	Correlation	OLS *t*-Statistic	R^2
Capital-labor endowment ratio (K/L)	−.35	−2.12	.12	−.80	−9.65	.63
GDP per capita	−.40	−2.54	.16	−.74	−8.03	.64

of country *i*'s share. However, estimating this alternative prediction also fails to confirm the theoretical relationship in the IRS/unicone Heckscher-Ohlin model between the size of the differentiated goods sector and factor endowments.[11]

In the case of the unicone Heckscher-Ohlin model, we estimate 57 parameters γ^i, the share of the labor-intensive good in GDP. The assumption of imperfect specialization means that γ should be greater than zero for all countries, and our estimates are consistent with that (conditional on the set value of $\gamma = 1$ for Sierra Leone). The range of the estimates is relatively small, from a low of 0.855 for Belgium-Luxembourg to a high of 1.027 for Mauritius. Because some international differences in the size of the homogeneous goods sector in GDP are likely to be larger than (approximately) 17 percentage points, the γ estimates should not be interpreted too literally. Nevertheless, it is interesting to see how different the point estimates are relative to the set value for Sierra Leone. The *p*-values in column 8 of table 5 report the probability that the γ of a given country is at least as large as that of Sierra Leone. For 36 out of 57 countries, or 63 percent, we can reject the null hypothesis that $\gamma^i = 1$. Moreover, none of the γ^i are estimated significantly higher than one at standard levels.

We now examine the correlation of relative capital endowments and the estimated γ's. All else equal, the higher a country's relative capital endowment, the more it will specialize in the production of the capital-intensive good, implying a lower share of the labor-intensive goods sector in GDP. Thus we expect a negative correlation of γ and K/L. As reported in table 6, this is what we obtain, with a correlation of −.80. The result is strongly confirmed by an OLS regression of γ on a constant and

[11] We estimate 47 $(1-\gamma^j)$ parameters, all positive, and three parameters exceed one (for Malawi, Iceland, and West Germany). If the homogeneous good Z is relatively capital-intensive, the model predicts a positive correlation of γ, Z's share in GDP, and K/L (or, equivalently, a negative correlation of $1-\gamma^j$ and K^j/L^j). Regressing $1-\gamma^j$ on a constant and K^j/L^j, we find a *t*-statistic for the slope parameter of −0.35, which changes to 2.06 once the three estimates above one are dropped.

K/L, which results in a t-statistic of the slope parameter of -9.65 and accounts for 63 percent of the variation in the estimated γ's. We have also examined the correlation of γ and K/L for different levels of differences in factor proportions, as indexed by class v, to see whether the negative correlation holds for all levels of FDIF to the same extent. While the correlation is always found to be negative, the differences in K/L account for about 70 percent in the variation of γ's for the classes $v = 1$ and $v = 2$, where FDIF is relatively low, but only for less than 10 percent for the classes $v = 4$ and $v = 5$, where differences in factor proportions are higher. These results suggest that the predicted link between factor proportions and trade is stronger for countries with small to intermediate differences in factor proportions, where the assumption that such countries share the same cone of production diversification is more plausible.

Overall, these results suggest that the factor proportions–driven pattern of specialization and trade that the unicone Heckscher-Ohlin model predicts has indeed an important effect on the volume of bilateral trade. We now briefly analyze the robustness of the major findings by considering (*a*) the case of endogenous class formation and (*b*) the case of a different critical level that separates the high- and low-Grubel-Lloyd samples.

B. Alternative Specifications

1. Endogenous Class Size

We allow for a varying number of observations in each class or, equivalently, for the endogenous formation of classes according to the variables FDIF^{ij} and GL^{ij}, respectively. We choose that allocation of observations across five classes that minimizes the AIC. Another interpretation of this analysis is that we estimate the five α_v's together with the four class breakpoint values of FDIF (or Grubel-Lloyd index, respectively) by maximum likelihood estimation (MLE).[12] Table 7 shows the results.

For the IRS model, we estimate α_v between 0.016 (for the lowest Grubel-Lloyd sample) and 0.133 (for the second-highest Grubel-Lloyd sample). As in the benchmark case, all α_v estimates are significantly lower than the theoretically predicted value of one. Inspection of the confidence intervals reveals that both the hypothesis that $\alpha_v = \alpha$, for all v, and the hypothesis that $\alpha_1 = \alpha_5$ are rejected, even though in the top end of the distribution there seems to be relatively little difference in the α_v's for $v = 2$ to $v = 5$ (see the ninety-fifth percentile estimates).

[12] In order to ensure the estimability of the α's in each class, we have to use certain minimum numbers of observations per class. Our estimates are therefore only approximations to maximum likelihood estimates.

TABLE 7
Endogenous Class Size

	IRS/Heckscher-Ohlin Model (*GL* > .05)				Heckscher-Ohlin Model (*GL* < .05)			
	IRS Model (IRS/IRS Goods)		IRS/Unicone Heckscher-Ohlin Model (IRS/CRS Goods)		Multicone Heckscher-Ohlin Model (CRS/CRS Goods)		Unicone Heckscher-Ohlin Model (CRS/CRS Goods)	
	α_v	5% 95%	α_v	5% 95%	α_v	5% 95%	α_v	5% 95%
	Ranked by Grubel-Lloyd Index				Ranked by FDIF			
$v=1$	.016	.011	.076	.053	.039	.024	.015	.008
	(.015)	.058	(.030)	.144	(.014)	.056	(.007)	.031
$v=2$	.042	.034	.059	.047	.111	.074	.030	.019
	(.060)	.209	(.026)	.126	(.026)	.157	(.012)	.051
$v=3$	.027	.037	.089	.038	.029	.025	.047	.034
	(.051)	.208	(.037)	.150	(.033)	.123	(.017)	.084
$v=4$	.133	.040	.106	.050	.058	.026	.079	.043
	(.039)	.152	(.064)	.270	(.018)	.076	(.013)	.085
$v=5$	.102	.080	.103	.092	.037	.026	.058	.052
	(.044)	.192	(.033)	.192	(.015)	.062	(.010)	.076
Only perfect specialization of production	yes		no		yes		no	
H_0: $\alpha_i=\alpha\ \forall i$	reject		do not reject		reject		reject	
H_0: $\alpha_1=\alpha_5$	reject		do not reject		do not reject		reject	
Share of bilateral comparisons correct	N.A.		8/10		N.A.		9/10	

Note.—Standard errors are in parentheses. The entries for all observations are the same as in table 3.

With 630 high-Grubel-Lloyd observations, the best fit tends to be obtained with a relatively high number of observations in the $v = 1$ and $v = 5$ classes.[13]

In the case of the multicone Heckscher-Ohlin model, the α_v estimates range from 0.029 to 0.111, all significantly lower than one. As in the benchmark case, we find that the highest estimate is not in the class with the largest differences in factor proportions, but in the $v = 2$ class. Moreover, the estimate of α_2 is significantly larger than α_5 at a 10 percent level, and we cannot reject the null hypothesis that $\alpha_1 = \alpha_5$. The best fit tends to be obtained when the 2,240 low-Grubel-Lloyd observations are allocated more or less evenly across the five classes; only the $v = 5$ class tends to be relatively large. We now turn to the imperfect specialization models.

In the high-Grubel-Lloyd sample, the results for the IRS/unicone

[13] Details on the class breakpoint analysis are available from the authors on request.

Heckscher-Ohlin model differ from those of the benchmark case discussed above. While the point estimates of the median $(\widehat{1-\gamma_v^i})$ still tend to increase with the Grubel-Lloyd index and eight out of 10 bilateral comparisons are correctly predicted, the standard errors of the estimates are fairly large. Consequently, one cannot reject the hypothesis that $\alpha_v = \alpha$, for all v, nor the hypothesis that the estimates for the first and the fifth classes are equal. The reason for the diverging MLE results relative to those of the benchmark case has to do with the different allocation of the 630 observations into the five classes. Here there are approximately 78 observations each in classes $v = 1$ to $v = 4$, and the remaining approximately 316 observations are in class 5.

The MLE estimates of the median $(\widehat{\gamma^j - \gamma^i})$ for the unicone Heckscher-Ohlin model range from 0.015 for $v = 1$ to 0.079 for $v = 4$. As in the benchmark case, the α_v's are broadly increasing with higher differences in factor proportions (nine of the bilateral comparisons correct). The allocation of the total observations across classes varies from that in the benchmark case: the classes $v = 1$ to $v = 4$ tend to have about 360 observations, and the remaining approximately 800 are in the class with the higher differences in factor proportions. However, in contrast to the IRS/unicone Heckscher-Ohlin model, we can reject both the null hypothesis that $\alpha_v = \alpha$, for all v, and the hypothesis that $\alpha_1 = \alpha_5$ at standard levels of significance.

In sum, while noting both the generally larger confidence intervals and the weaker results for the IRS/unicone Heckscher-Ohlin model, we find that the MLE results are broadly similar to those that we have obtained in the benchmark case.

2. A Different Separation into High- and Low-Grubel-Lloyd Samples

Table 8 shows the results when the critical level $\overline{GL}$ to separate the high- and low-Grubel-Lloyd samples is set at .033; this allocated 740 observations to the high-Grubel-Lloyd sample and 2,130 observations to the low-Grubel-Lloyd sample.

The α_v estimates for the IRS model lie now between 0.019 and 0.141, and the regression with all 740 observations gives an α of 0.085. These results are qualitatively identical to those of the benchmark case; in particular, all import parameter estimates are significantly below the value predicted by the model, which is one. This is also the case for the multicone Heckscher-Ohlin results, where the regression with all 2,130 observations leads to an α of 0.052; this is identical to the corresponding value in the benchmark case.

Among the imperfect specialization models, the IRS/unicone Heckscher-Ohlin model fares worse in that now only six bilateral com-

TABLE 8
Alternative Critical Grubel-Lloyd Level ($\overline{GL}$ = .033)

	IRS/Heckscher-Ohlin Model: High-Grubel-Lloyd Sample ($GL > .05$)				Heckscher-Ohlin Model: Low-Grubel-Lloyd Sample ($GL < .05$)			
	IRS Model (IRS/IRS Goods)		IRS/Unicone Heckscher-Ohlin Model (IRS/CRS Goods)		Multicone Heckscher-Ohlin Model (CRS/CRS Goods)		Unicone Heckscher-Ohlin Model (CRS/CRS Goods)	
	α_v	5% 95%	α_v	5% 95%	α_v	5% 95%	α_v	5% 95%
	Ranked by Grubel-Lloyd Index				Ranked by FDIF			
$v=1$	.057	.047	.070	.057	.052	.040	.015	.013
	(.010)	.072	(.007)	.079	(.008)	.067	(.004)	.025
$v=2$	.019	.014	.068	.061	.123	.096	.026	.020
	(.010)	.045	(.005)	.077	(.020)	.152	(.008)	.044
$v=3$	.141	.121	.124	.111	.049	.041	.056	.035
	(.015)	.168	(.006)	.134	(.005)	.057	(.008)	.059
$v=4$	.103	.078	.117	.106	.034	.028	.049	.044
	(.019)	.139	(.005)	.123	(.004)	.041	(.006)	.062
$v=5$	.091	.078	.100	.097	.062	.054	.069	.055
	(.012)	.115	(.016)	.154	(.006)	.072	(.007)	.076
All observations	.085	.073	.087	.075	.052	.047	.033	.031
	(.007)	.095	(.005)	.092	(.003)	.056	(.003)	.040
Only perfect specialization of production	yes		no		yes		no	
H_0: $\alpha_i = \alpha\ \forall i$	reject		reject		reject		reject	
H_0: $\alpha_1 = \alpha_5$	reject		reject		do not reject		reject	
Share of bilateral comparisons correct	N.A.		6/10		N.A.		9/10	

Note.—Standard errors are in parentheses.

parisons are correctly predicted, as opposed to nine in the benchmark case. This confirms the MLE results for this model and suggests that the differences between the $\widehat{1-\gamma_v^i}$ estimates for the higher Grubel-Lloyd classes should not be overemphasized. Nevertheless, it remains the case that one can reject both the null hypothesis that all α_v's are the same and the null hypothesis that $\alpha_1 = \alpha_5$ according to the confidence intervals reported in table 8. For the unicone Heckscher-Ohlin model, the pattern of the estimates is the same as in the benchmark case. Moreover, we reject both the hypothesis that all α_v's are identical and the hypothesis that α_1 is as large as α_5. Thus the performance of the unicone Heckscher-Ohlin model appears to be robust.

In some unreported analysis we have also examined both the goodness of fit of the models and the country-specific parameter estimates for the imperfect specialization models. This produced results very similar to those in the benchmark case. In particular, the correlation of $1 - \gamma^i$ and K^i/L^i estimated from the IRS/unicone Heckscher-Ohlin model is nonpositive, which differs from the positive correlation predicted by the model. In contrast, the correlation of the γ^i's estimated from the unicone Heckscher-Ohlin model and data on K^i/L^i remains strong and has the correct sign (which is negative).

Overall, the analysis of these two alternative specifications suggests that the qualitative results that we have obtained in the benchmark case are by and large robust.

C. The Effect of Distance and Other Sensitivity Analyses

One element that we have not controlled for so far is differences in the distance between countries. It is well established empirically that bilateral trade volumes fall as distance increases (the second pillar of the gravity equation), raising the possibility that the different parameter estimates are driven by differences in bilateral distance among trading partners.[14] Even though the predictions of import volume we use do not incorporate the effect of trade costs, at an empirical level, it is important to see whether our results depend on the omission of bilateral distance from our gravity regressions. Our analysis in section *A* of the Appendix shows that this is not the case. While this cannot settle the theoretical question of what the appropriate predictions for bilateral trade volumes are in the presence of trade costs, it does mean that our earlier findings are robust to incorporating the effects of distance.

We also report in the Appendix the results from a range of additional sensitivity analyses. This encompasses alternative specifications, outlier analysis, as well as different sources and definitions of the data, which are intended to address problems resulting from measurement error in the data that we have used. The broad conclusion that emerges from this is that our results are not sensitive to these factors.

V. Conclusions

Although it is well known that the gravity equation can explain much of the variation in bilateral trade volumes, to date there has been little agreement about which theory or theories underpin its success. In this paper we have employed information on the extent of intraindustry trade and on the differences in factor endowments in a given sample

[14] Note, however, the recent paper by Krishna (2000), who argues that the importance of distance has been exaggerated in previous research.

to help identify which of four models might be driving the gravity equation.

Our approach relies on identifying intraindustry trade with IRS-based trade. Intraindustry trade, though, can also be caused by other reasons, in particular by Ricardian technology differences (Davis 1995). However, the lack of data on production technologies at the goods level makes it difficult to evaluate this alternative hypothesis. In addition, we have not modeled other potentially important determinants of trade flows, especially transportation and trade costs. Although we find that controlling for distance does not overturn our principal findings, this analysis is far from complete; moreover, it does not address how trade impediments alter international trade flows in terms of theory.

We find that specialization and trade are increasing as the share of intraindustry trade in total trade rises. This provides support for the product differentiation cum IRS model of trade to play a role in North-North trade. However, the perfect specialization version of this model overpredicts the amount of bilateral trade by a large margin. Adding a factor abundance element leads to some imperfectly specialized production. In line with the predictions of this IRS/unicone Heckscher-Ohlin model, we find that the size of the differentiated goods sector and the share of intraindustry trade move together. This suggests that scale economies and product differentiation are important in explaining the volume of North-North bilateral trade. At the same time, the size of the differentiated goods sector is not related to relative factor abundance in the way predicted by the IRS/unicone Heckscher-Ohlin model.

Before discussing our findings in samples in which intraindustry trade is only a small part of overall trade and comparing them with those obtained for the high-Grubel-Lloyd sample, we caution again that the data quality for the low-Grubel-Lloyd sample is worse than for the high-Grubel-Lloyd sample. That being said, in samples with little intraindustry trade, the prediction of the multicone Heckscher-Ohlin model is rejected. Thus there is no support for the *perfect* specialization Heckscher-Ohlin model. Our other estimates suggest that *imperfect* specialization due to differences in factor proportions is important for explaining bilateral trade. Generally, trade volumes rise with greater differences in factor proportions. Moreover, as predicted, the relative size of the labor-intensive goods sector is negatively correlated with the relative capital endowment across countries. This suggests that the factor abundance motive is important in explaining the volume of North-South trade.

In the light of this paper, there is little support for trade models that predict perfect specialization for all goods: both the IRS model and the multicone Heckscher-Ohlin model are rejected in our work. Therefore, the analysis of perfect specialization models in Hummels and Levinsohn (1995), which appears to throw doubt on the empirical relevance of

IRS trade theory, is misplaced. Imperfect specialization due to differences in factor proportions is likely to account for the regression results that these authors obtain from a sample of trade relations between non-OECD nations, where little product differentiation/IRS was expected. Our results suggest that IRS and product differentiation are empirically relevant elements in helping to explain the volume of bilateral trade. At the same time, they stop well short of a full endorsement of the IRS models that we have considered. Specifically, when product differentiation in trade is important, differences in relative factor abundance across countries seem to matter less for specialization than according to the synthesis of factor abundance and IRS motives for trade proposed by Helpman (1981). In part, our mixed results for the IRS/unicone Heckscher-Ohlin model might be due to the fact that every differentiated good is neither more capital-intensive nor more labor-intensive than homogeneous goods, in contrast to the clear-cut distinction in Helpman's model. Further, once product differentiation in trade is high (and bilateral differences in factor proportions are small), perhaps it is too much to ask of the data that the size of the differentiated goods sector be related to relative factor abundance across countries. This corresponds to the widely held expectation that the influence of factor abundance is muted in North-North trade relationships. Underlying all of this is the strong suspicion that this two-good, two-factor IRS/unicone Heckscher-Ohlin model is too simple to perform well empirically.

While we reject the multicone Heckscher-Ohlin model, the support for the unicone Heckscher-Ohlin model is perhaps surprisingly strong. Of course, this support does not mean that the world's countries indeed occupy the same cone of diversification; data on factor prices as well as on factor input ratios in production cast doubt on this notion. Moreover, rejecting a model in which each country occupies a distinct cone does not mean that perfect specialization due to differences in factor proportions is irrelevant; it leaves open the possibility of a relatively small number of, say, three or four diversification cones. Future research should develop multicountry IRS/Heckscher-Ohlin models in which production is partly specialized—because of both factor abundance and IRS—and partly diversified and in which bilateral trade volumes are determined. We take the relative success of the unicone Heckscher-Ohlin model to suggest that the bilateral trade flow predictions of that multicountry model will probably turn out to be closely related to that of the unicone Heckscher-Ohlin model that we have analyzed here.

Appendix

A. *Sensitivity Analysis I*

Results on the first and more extensive set of sensitivity analyses are presented in table A1; the benchmark case is listed again for reference. The specifications

TABLE A1
Robustness Analysis I

	IRS	IRS/Unicone	Multicone	Unicone
		Benchmark Specification		
H_0: $\alpha_i = \alpha \ \forall i$	reject	reject	reject	reject
H_0: $\alpha_1 = \alpha_5$	reject	reject	not reject	reject
Share of bilateral comparisons correct	N.A.	9/10	n/a	9/10
t-statistic*	N.A.	−2.12	N.A.	−9.65
		$\overline{GL}$ = .075		
H_0: $\alpha_i = \alpha \ \forall i$	reject	reject	reject†	reject
H_0: $\alpha_1 = \alpha_5$	reject	reject	reject	reject
Share of bilateral comparisons correct	N.A.	9/10	N.A.	8/10
t-statistic	N.A.	−.91	N.A.	−10.49
		With Distance		
H_0: $\alpha_i = \alpha \ \forall i$	reject	reject	reject	reject
H_0: $\alpha_1 = \alpha_5$	reject	reject	not reject	reject
Share of bilateral comparisons correct	N.A.	6/10	N.A.	9/10
t-statistic	N.A.	−2.08	N.A.	−9.69
		With Constant		
H_0: $\alpha_i = \alpha \ \forall i$	reject	reject	reject	reject
H_0: $\alpha_1 = \alpha_5$	reject	reject	not reject	reject
Share of bilateral comparisons correct	N.A.	7/10	N.A.	8/10
t-statistic	N.A.	1.00	N.A.	−2.07

* The *t*-statistic is the ratio of the slope coefficient to its standard error in an OLS regression of $1-\gamma^i$ (for the IRS/unicone model) and γ^i (for the unicone Heckscher-Ohlin model) on K^i/L^i.

† The coefficient α_1 is significantly larger than α_5.

are as follows: (1) The first specification includes a higher critical level that separates the low- and the high-Grubel-Lloyd samples, with $\overline{GL} = .075$. This complements the analysis in the text, where results for the cases of $\overline{GL} = .050$ (the benchmark case) and $\overline{GL} = .033$ are discussed. (2) The second specification incorporates bilateral distance between countries i and j; it is

$$M_v^{ij} = \alpha_v \frac{Y_v^i Y_v^j}{Y^w d^{ij}} + \epsilon_v^{ij} \tag{A1}$$

for all v, where d^{ij} is the log distance between countries i and j. (3) The third specification includes a constant term, α_0:

$$M_v^{ij} = \alpha_0 + \alpha_v \frac{Y_v^i Y_v^j}{Y^w} + \epsilon_v^{ij}. \tag{A2}$$

For each of these three specifications, four statistics that summarize the performance of the models are reported: First, we test whether the α_v's across all five classes are equal. Second, we test whether the values for the first (lowest Grubel-Lloyd or FDIF class, respectively) and the fifth classes are equal. Third, for the two imperfect specialization models, we report how many bilateral comparisons of α_v point estimates are correctly predicted for a given model. All results are based on bootstrapped α_v and their symmetric confidence intervals; the tests are based on a significance level of 10 percent. Fourth, table A1 reports the *t*-statistic from an OLS regression of γ^i (for the unicone Heckscher-Ohlin model) and $1 - \gamma^i$ (IRS/unicone model) on the relative capital endowments, K^i/L^i. All three specifications confirm the benchmark results for the two hypothesis tests, with one exception: for the multicone Heckscher-Ohlin model in the $\overline{GL} = .075$ case, we find that α_1 is significantly larger than α_5. This does not provide more evidence in support of this model, though, since even the α_1 estimate remains smaller than the theoretically predicted value of one at standard levels of significance.

We now turn to the correlation of the relative capital endowments with the country-specific estimates from the imperfect specialization models. While there are no major differences relative to the benchmark case for $\overline{GL} = .075$ and the distance specification, including a constant leads to the following changes. The correlation of $1 - \gamma^i$ and K^i/L^i is now positive, which is what one expects on the basis of the assumption that the differentiated good is capital-intensive. This correlation is not very strong, though. In consequence, this cannot change our general assessment of the performance of the IRS/unicone model, which has been mixed. For the unicone Heckscher-Ohlin model, the correlation between γ^i and K^i/L^i is negative, as predicted by the model. However, the substantially lower *t*-statistic (-2.07 vs. -9.65 for the benchmark case) indicates that the relation between production specialization and factor proportions is weakened. In none of these three alternative specifications, though, do we obtain qualitatively different results relative to the benchmark case: for the IRS/unicone model, the predicted correlation is weak or has the wrong sign, whereas for the unicone Heckscher-Ohlin model, the correlation has the correct sign, and it is generally also quite strong.

While table A1 shows that controlling for within-class bilateral distance does not affect the results substantially, we have extended this analysis to control for between-class differences in bilateral distance. It turns out that the bilateral distance between countries does not vary much across FDIF classes. However, the bilateral distance across the five classes in the high-Grubel-Lloyd sample varies substantially; in particular, the median bilateral distance is falling from 6,357 kilometers for $v = 1$ to 1,066 kilometers for $v = 5$. Because it is well known that, ceteris paribus, there is a higher level of trade over a shorter distance, it is important to see whether our result that α_v is rising with v for the IRS/unicone model in the high-Grubel-Lloyd sample holds up even when we control for the differences in bilateral distance across classes. Our analysis (not reported, but available on request) shows that this is the case.

B. *Sensitivity Analysis II: Measurement Error and Other Problems*

In table A2, we report the results of further sensitivity analysis for the two imperfect specialization models. The alternative specifications are as follows: (1) $V = 2$ classes instead of five; (2) $V = 3$ classes; (3) endogenous formation of three Grubel-Lloyd and FDIF classes as a function of model fit (denoted as three-class MLE in table A2); (4) modified Grubel-Lloyd indices. Instead of using

TABLE A2
Robustness Analysis II

Specification	Share of Bilateral Comparisons Correct (out of x)	
	IRS/Unicone	Unicone
Benchmark	90% (10)	90% (10)
Two classes ($V=2$)	100% (1)	100% (1)
Three classes ($V=3$)	66% (3)	100% (3)
Three-class MLE	100% (3)	100% (3)
Modified Grubel-Lloyd	70% (10)	70% (10)
Dependent variable $M^{ij}/(Y^iY^j/Y^w)$	100% (10)	40% (10)
GDP data from WDI	70% (10)	90% (10)
Imports data from IMF	70% (10)	100% (10)
Average	83%	86%

the GL^{ij} values as computed in point 3 above, we use modified Grubel-Lloyd values in which the idiosyncratic component for a given bilateral pair is eliminated; this series is constructed by running a linear regression of GL^{ij} on country effects and then using the fitted values from this as the modified Grubel-Lloyd series. The correlation between the two Grubel-Lloyd series is equal to .70. (5) The dependent variable is imports divided by the GDP term, $M^{ij}/(Y^iY^j/Y^w)$. If simultaneity between imports and GDPs is a major concern or there is mismeasurement in the GDP terms, this specification should help detecting it. For the perfect specialization models, the α_v's are given by the ratio of the mean of M^{ij} and the mean of Y^iY^j/Y^w. (6) GDP data come from the World Bank's (2000) *World Development Indicators* (WDI) database instead of from Summers and Heston (1991). Our preferred measure is the GDP data from Summers and Heston, which are based on purchasing power parity exchange rates, whereas the WDI GDP data employed here are computed using actual exchange rates. (7) Bilateral imports data are taken from the International Monetary Fund's *Directions of Trade Statistics* (1995 edition) instead of from Feenstra et al. (1997). These data are based on the reports of national treasuries to the IMF and differ from the trade flows reported by customs officials, which form the basis of the revised U.N. trade data of Feenstra et al.

Table A2 indicates that the results are by and large robust. For both imperfect specialization models, we see that the a priori expectation that α_v is rising with v is broadly confirmed. We have also considered the specifications 1–7 for the multicone Heckscher-Ohlin and the IRS models. Also, when these perfect specialization models are employed, the results are similar to the benchmark results, and in particular, the import parameter estimate α tends to be much smaller than one.

Finally, a number of unreported sensitivity checks regarding the following have been conducted: (1) computing world GDP with Summers-Heston data as the sum of the GDP of all 152 countries for which the data are available, as opposed to the 58 countries in the sample that we use; (2) using the WDI's GDP data on 58 countries to compute world GDP; (3) employing Summers-Heston data on GDP per capita as an alternative proxy of the relative capital endowment, K^i/L^i, for the analysis of the multicone and unicone Heckscher-Ohlin models; (4) analyzing outlier and influential points in the form of (*a*) dropping each observation one at a time and reestimating and (*b*) discarding small groups of observations in which the absolute value of the residual exceeds some specified value and reestimating; and (5) making adjustments using different as well as simpler heteroskedasticity (scaling). The analysis of these cases has always led to results similar to those for the benchmark case.

References

Anderson, James. "A Theoretical Foundation for the Gravity Equation." *A.E.R.* 69 (March 1979): 106–16.

Antweiler, Werner, and Trefler, Daniel. "Increasing Returns and All That: A View from Trade." Manuscript. Toronto: Univ. Toronto, August 1997.

Austrian Institute for Economic Research. *Competitiveness Report 1999.* Vienna: Austrian Inst. Econ. Res. (for European Comm.), 1999.

Bartelsman, Eric, and Gray, Wayne. NBER Manufacturing Productivity Database. Cambridge, Mass.: NBER, March 2000. http://www.nber.org.

Bergstrand, Jeffrey H. "The Heckscher-Ohlin-Samuelson Model, the Linder Hypothesis and the Determinants of Bilateral Intra-industry Trade." *Econ. J.* 100 (December 1990): 1216–29.

Bickel, Peter J., and Freedman, David A. "Some Asymptotic Theory for the Bootstrap." *Ann. Statis.* 9 (November 1981): 1196–1217.

Davis, Donald R. "Intra-industry Trade: A Heckscher-Ohlin-Ricardo Approach." *J. Internat. Econ.* 39 (November 1995): 201–26.

Deardorff, Alan V. "Determinants of Bilateral Trade: Does Gravity Work in a Neoclassical World?" In *The Regionalization of the World Economy,* edited by Jeffrey A. Frankel. Chicago: Univ. Chicago Press (for NBER), 1998.

Eaton, Jonathan, and Kortum, Samuel. "Technology and Bilateral Trade." Manuscript. Boston: Boston Univ., April 1997.

Efron, Bradley. *The Jackknife, the Bootstrap and Other Resampling Plans.* Philadelphia: Soc. Indus. and Appl. Math., 1982.

Feenstra, Robert C.; Lipsey, Robert E.; and Bowen, Harry P. "World Trade Flows, 1970–1992, with Production and Tariff Data." Working Paper no. 5910 and accompanying CD-ROM. Cambridge, Mass.: NBER, January 1997.

Finger, J. Michael. "A New View of the Product Cycle Theory." *Weltwirtschaftliches Archiv* 111, no. 1 (1975): 79–99.

Fontagne, Lionel; Freudenberg, Michael; and Peridy, Nicolas. "Intra-industry Trade and the Single Market: Quality Matters." Working Paper no. 1959. London: Centre Econ. Policy Res., 1998.

Greenaway, David; Hine, Robert; and Milner, Chris. "Country-Specific Factors and the Pattern of Horizontal and Vertical Intra-industry Trade in the UK." *Weltwirtschaftliches Archiv* 130, no. 1 (1994): 77–100.

Grubel, Herbert G., and Lloyd, Peter J. *Intra-industry Trade: The Theory and Measurement of International Trade in Differentiated Products.* London: Macmillan, 1975.

Harrigan, James. "Openness to Trade in Manufactures in the OECD." *J. Internat. Econ.* 40 (February 1996): 23–39.

Haveman, Jon. Geographic Distance Data. 1998. http://www.macalester.edu/

research/economics/PAGE/HAVEMAN/Trade.Resources/TradeData.html#Gravity.

Helpman, Elhanan. "International Trade in the Presence of Product Differentiation, Economies of Scale and Monopolistic Competition: A Chamberlin-Heckscher-Ohlin Approach." *J. Internat. Econ.* 11 (August 1981): 305–40.

———. "Imperfect Competition and International Trade: Evidence from Fourteen Industrial Countries." *J. Japanese and Internat. Econ.* 1 (March 1987): 62–81.

———. "The Structure of Foreign Trade." Working Paper no. 6752. Cambridge, Mass.: NBER, October 1998.

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

Hummels, David, and Levinsohn, James. "Monopolistic Competition and International Trade: Reconsidering the Evidence." *Q.J.E.* 110 (August 1995): 799–836.

International Monetary Fund. *Directions of Trade Statistics.* 1995 yearbook. Washington: Internat. Monetary Fund, 1995.

Keller, Wolfgang. "Geographic Localization of International Technology Diffusion." *A.E.R.* 92 (March 2002).

Kennedy, Peter. *A Guide to Econometrics.* 3d ed. Cambridge, Mass.: MIT Press, 1992.

Krishna, Pravin. "Are Regional Trading Partners 'Natural'?" Working paper. Providence, R.I.: Brown Univ., 2000.

Krugman, Paul R. "Empirical Evidence on the New Trade Theories: The Current State of Play." In *New Trade Theories: A Look at the Empirical Evidence.* Conference Report. London: Centre Econ. Policy Res., 1994.

Leamer, Edward E., and Levinsohn, James. "International Trade Theory: The Evidence." In *Handbook of International Economics,* vol. 3, edited by Gene M. Grossman and Kenneth Rogoff. Amsterdam: Elsevier, 1995.

Markusen, James R., and Wigle, Randall M. "Explaining the Volume of North-South Trade." *Econ. J.* 100 (December 1990): 1206–15.

Peneder, Michael. "Intangible Investment and Human Resources: The New WIFO Taxonomy of Manufacturing Industries." WIFO Working Paper no. 114. Vienna: Austrian Inst. Econ. Res., May 1999.

Rauch, James E. "Networks versus Markets in International Trade." *J. Internat. Econ.* 48 (June 1999): 7–35.

Saxonhouse, Gary R. "Differentiated Products, Economies of Scale, and Access to the Japanese Market." In *Trade Policies for International Competitiveness,* edited by Robert C. Feenstra. Chicago: Univ. Chicago Press (for NBER), 1989.

Summers, Robert, and Heston, Alan. "The Penn World Table (Mark 5): An Expanded Set of International Comparisons, 1950–1988." *Q.J.E.* 106 (May 1991): 327–68.

World Bank. *World Development Indicators 2000.* CD-ROM. Washington: World Bank, 2000.