On Theories Explaining the Success of the Gravity Equation*

Simon J. Evenett†

Rutgers University and The Brookings Institution

Wolfgang Keller‡

University of Wisconsin and NBER

May 1998

Abstract

We analyze two main theories of international trade, the Heckscher-Ohlin theory and the Increasing Returns trade theory, by examining whether they can account for the empirical success of the so-called Gravity Equation. Since versions of both models can generate this prediction, we tackle the model identification problem by conditioning bilateral trade relations on factor endownment differences and the share of intra-industry trade, because only for large factor endowment differences does the Heckscher-Ohlin model generate specialization of production and the Gravity Equation, and it predicts inter-, not intra-industry trade. There are three major findings: First, little production is perfectly specialized due to factor endowment differences, making the perfect specialization version of the Heckscher-Ohlin model an unlikely candidate to explain the empirical success of the Gravity Equation. Second, increasing returns are important causes for perfect product specialization and the Gravity Equation, especially among industrialized countries. Third, to the extent that production is not perfectly specialized across countries, we find support for both Heckscher-Ohlin

*First version: November 1996. We have benefited from comments at seminar presentations at Berlin (Humboldt), Columbia, Geneva (Graduate Institute), Michigan, Penn, Toronto, and Wisconsin, as well as from participants at the 1997 NBER Summer Institute, the August 1997 EEA meeting at Toulouse, the September 1997 ERWIT/CEPR conference in Helsinki, and the Empirical Trade Conference 1996 at Purdue University. Special thanks go to Don Davis, Mahmoud El-Gamal, Gordon Hanson, and Elhanan Helpman. All remaining errors are our own.

†Department of Economics, 75 Hamilton Street, New Brunswick NJ 08901-1248; email: evenett@fas-econ.rutgers.edu.

‡Department of Economics, 1180 Observatory Drive, Madison, WI 53706; email: wkeller@ssc.wisc.edu, web: www.ssc.wisc.edu/~wkeller.

and Increasing Returns models. Based on these findings, we argue that both models explain different components of the international variation of production patterns and trade volumes, with important implications for productivity growth, labor, and macro economics.

Keywords: Gravity equation, Heckscher-Ohlin trade theory, increasing returns trade theory, model identification, resampling, non-parametric analysis.

JEL Classification: F11, F12, C14

1. Introduction

The so-called Gravity Equation of trade predicts that the volume of bilateral trade is positively related to the product of the countries' gross domestic products (GDPs) and negatively related to trade barriers between trade partners. Empirical research has found that various versions of the Gravity Equation well describe the variation in the volume of trade across country-pairs as well as over time (see Leamer and Levinsohn 1995).

Since Anderson (1979) it has been increasingly recognized that the gravity equation prediction can be derived from very different structural models, including Ricardian models, Heckscher-Ohlin (H-O) models, and increasing returns to scale (IRS) models.[1] When consumers have both identical homothetic preferences and access to the same goods prices, a sufficient condition for obtaining a gravity equation is perfect product specialization, in the sense that each commodity is produced in only one country. The three types of trade models differ in the way product specialization is obtained in equilibrium: Technology differences across countries in the Ricardian model, factor proportions outside the vector space-of-diversification in the H-O model, and increasing returns at the firm level in the IRS model. Indeed, Deardorff (1998) has recently argued that the gravity prediction per se cannot be used to test any of these trade theories. Yet, the gravity prediction constitutes, perhaps along with the H-O-Vanek factor service trade prediction (see recent work by Trefler 1995, Davis et al. 1997), the most important result regarding the volume of international trade. Therefore, major insights are to be gained if it could be determined which model generates gravity-like trade volumes in a given sample of data, a problem we refer to as a model identification issue. In addition, it would allow us to assess the suggestion by Hummels and Levinsohn (1995, 828) that a theory other than IRS is responsible for the empirical success of the gravity equation.

In this paper, first we address this identification problem by noting that, on the one hand, in

[1] Bergstrand (1990), Deardorff (1998), Eaton and Kortum (1997), Helpman and Krugman (1985), Hummels and Levinsohn (1995), Leamer (1992), and Markusen and Wigle (1990).

a constant returns (CRS) H-O world, bilateral factor proportions differences must be very large in order to ensure the economies lie outside a common space of diversification and to generate product specialization (see Figure 1). On the other hand, when product specialization is the result of IRS, the gravity prediction can be obtained even when there are no factor proportions differences. This difference has the following implication for the type of trade in the two models: In the H-O model, trade is exclusively inter-industry trade, defined as trade in goods with different factor intensities. However for the IRS model at least some, and potentially all, trade is intra-industry trade. Consequently, in the analysis below, we will take samples with low shares of intra-industry trade in total trade together with large factor proportions differences as those where a model of H-O-based product specialization might be behind the Gravity Equation; and samples with high shares of bilateral intra-industry trade (irrespective of factor proportions differences) as those where IRS-based product specialization might drive the Gravity Equation.

This might suggest a country-by-country reconciliation of the perfect specialization models of the H-O and IRS-based trade theories. Along these lines, the perfect specialization H-O theory would be responsible for the gravity prediction's success in explaining the variation in bilateral trade flows among countries with large factor proportions differences and low shares of intra-industry trade (the so-called 'North-South' trade). At the same time, the IRS theory would account for the success of the gravity prediction in trade between countries where product differentiation and intra-industry trade is prevalent (the so-called 'North-North' trade.) However, no such reconcilation emerges from our analysis of a large and heterogeneous set of bilateral trade relations in the year 1985. We present evidence that perfect product specialization due to factor proportions differences is not a major part of explaining the success of the gravity prediction. In contrast, there is robust evidence that IRS-based theories of trade are an important reason why the gravity equation fits trade flows among industrialized countries well.

Secondly, the paper assesses the empirical relevance of the IRS-based trade models in general, with their far-reaching implications for productivity growth, labor, and open macroeconomics. This is more important now than ever before, as it has been demonstrated that there is no need to resort to IRS models to explain intra-industry trade (Davis 1995), or large trade volumes between countries with similar factor endowments (Chipman 1992, Davis 1997). Even at an empirical level, Helpman's (1987) work showing that key implications of the IRS-based model are consistent with OECD countries' trade data was followed up by Hummels and Levinsohn (1995), who repeated Helpman's analysis with a set of non-OECD countries, the trade between which was not expected to contain much IRS-based trade. They showed that many correlations found by Helpman continue to hold in their non-OECD sample. To take Hummels and Levinsohn's result as evidence against the IRS-based trade models would be wrong, however, because their finding is just an expression of the model identification problem. In this paper, we account for both Helpman's (1987) and Hummels and Levinsohn's (1995) findings by showing that the former obtained his results because IRS-based trade is prevalent among OECD countries, whereas the latter found a similar correlation among non-OECD countries because there is some (but not perfect) product specialization driven by factor proportions differences.

We present strong evidence that the volume of international trade is determined by the extent of product specialization, which in turn is due to IRS and factor proportions differences. Where there is little or no two-way trade between nations, we find that a H-O model that predicts imperfect specialization better explains trade flows than a H-O model with perfect specialization. Also where there is two-way trade among nations, international trade flows are better explained by a model that incorporates both factor proportions differences and IRS than by a model where IRS alone generates product specialization. These findings highlight the significance of both factor proportions differences and IRS as determinants of the extent of specialization and international trade flows. It parallels recent results by Antweiler and Trefler (1997), who nest IRS and CRS models in the H-O-Vanek factor service trade expression to estimate scale parameters from these input demand equations. These authors find that although a majority of industries seems to be well-characterized by the assumption of CRS, there is evidence for IRS in a number of sectors.[2]

There are several important caveats to our approach: First of all, in our analysis of the gravity equation we focus on the proportionality of the volume of trade to the countries' GDPs rather than its relationship to trade resistance. This corresponds to the fact that to date there is relatively little theory on what form the gravity equation takes in the presence of transport costs.[3] Although we take up the influence of trade resistance in section 4.4 below and conclude that our major findings are not sensitive to that omission, we plan to address the issue more comprehensively in future work.

Our approach relies on identifying intra-industry trade with IRS-based trade. Therefore, the second caveat is that instead of IRS we cannot rule out that Ricardian technology differences are what is really behind intra-industry trade (see Figure 1).[4] To make progress on this issue requires product-level estimates of production technologies across countries, which are unavailable. Thirdly, we note that also in the H-O model it is possible to generate trade in goods of identical factor intensity if one follows Armington (1969) and assumes that products are differentiated by country of production. A direct test of the Armington assumption would require data on people's perceptions of the differences (if any) between domestic and foreign-produced goods. We are not aware of data allowing us to test this assumption. In addition, by relying on the assumption of identical homothetic preferences across countries throughout, we do not give any role to demand as, for instance, emphasized by Markusen (1986). Lastly, we ignore trade imbalances in the analysis below. Those effects play, as in Helpman (1987) and Hummels and Levinsohn (1995), a quantitatively very minor role in what follows.

The remainder is as follows. In the following section, we derive the gravity equation predictions of four different trade models, and describe in more detail how we attempt to identify a particular model in our empirical analysis. Section 3 discusses the data set employed in this study, whereas section 4

[2] Also see Davis and Weinstein (1996, 1997) who test for the presence of IRS by exploiting the 'home-market' effect prediction (strong idiosyncratic demand for a certain good means a country will be exporting the good in the presence of IRS and transport costs); their results so far are mixed.

[3] It includes, however, Anderson (1979), Bergstrand (1990), and Deardorff (1998).

[4] Also for that reason, we cannot rule out that trade within an industry is driven by specialization across countries due to external effects at the sub-industry level, even though there are CRS at the product level.

presents the key empirical results. Section 5 concludes. The sensitivity of our results is discussed in the appendix.

2. Derivation of the Estimating Equations

2.1. The Gravity Equation with Perfect Product Specialization

Throughout the paper, we assume balanced trade, zero trade and transport costs, and that consumers in all countries have identical homothetic preferences. First, we consider a typical model as laid out in Helpman and Krugman (1985, Ch. 8.1), where there are two countries, $c = i$, j, and two goods, $g = X, Z$ (see Appendix A for more details). Both X and Z come in many symmetric differentiated varieties which are identically produced with increasing returns to scale. With preferences valuing product variety, both countries will demand all foreign varieties according to the countries' GDP as a share of world GDP, denoted s^c. Given that IRS leads to perfect product specialization for every variety, country i's imports from j, denoted M^{ij}, will be

$$M^{ij} = s^i \left[p_x n_x^j x^j + n_z^j z^j \right]. \tag{2.1}$$

Here, n_g^c is the number of varieties of good g in country c, and p_x is the relative price of good x. Country j's imports from i, M^{ji}, will be

$$M^{ji} = s^j \left[p_x n_x^i x^i + n_z^i z^i \right]. \tag{2.2}$$

The terms in the hard brackets of (2.1) and (2.2) are equal to the GDP of country j (denoted Y_j) and country i (denoted Y_i) respectively. Substitution of Y_i and Y_j yields

$$M^{ij} = s^i Y^j = \frac{Y^i Y^j}{Y^w} = s^j Y^i = M^{ji}. \tag{2.3}$$

Equation (2.3) is the gravity prediction referred to above,[5] and hence, this IRS trade model is potentially a candidate to explain the success of the gravity equation.

The equation (2.3) is very general, as its derivation does not require assumptions on factor price equalization, factor endowment differences across countries, or factor intensities in the production of goods X and Z (Helpman and Krugman 1985). Equation (2.3) holds whenever there is perfect product specialization in equilibrium, all consumers face the same goods prices and have identical homothetic preferences, and trade is balanced. Therefore, if instead one assumes that goods X and Z are homogeneous and produced with CRS, one can still derive (2.3) when perfect product specialization is obtained in equilibrium. For the following exposition, we assume without loss of generality that country i is capital-abundant and good X capital-intensive. Consider a model where factor endowments between the countries differ by so much that each country specializes in the production of one good. Let k^x and k^z denote the equilibrium capital-labor ratios which are employed in industry X and Z, respectively, in the integrated equilibrium (see Helpman and Krugman 1985, Ch.1). Then, if factor proportions in countries i and j lie outside that range,

$$\frac{K^i}{L^i} \geq k^x > k^z \geq \frac{K^j}{L^j}, \tag{2.4}$$

at least one country will specialize in the production of only one good (if the capital-abundant country i specializes, it produces the capital-intensive good X, and if the labor-abundant country j specializes, it does so producing the labor-intensive good Z, by comparative advantage). For the generic case– obtained with sufficiently similar country size and no strong taste bias–, both countries specialize in producing only their comparative-advantage good. In that case, if X^c and Z^c denote the production of good X and Z, respectively, we have $X^i = X^w$ and $Z^j = Z^w$; further, it is clear that the value of

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

good $X-$production in country i is equal to its GDP, i.e., $p_x X^i = Y^i$, and similarly, $Z^j = Y^j$. This means that

$$M^{ij} = s^i Z^j = s^i Y^j = \frac{Y^i Y^j}{Y^w}, \qquad M^{ji} = s^j p_x X^i = s^j Y^i = \frac{Y^i Y^j}{Y^w}, \tag{2.5}$$

which is identical to the gravity equation (2.3) above.

As discussed by Helpman and Krugman (1985), this analysis generalizes to the multi-sector, multi-factor, and multi-country settings as long as specialization is the equilibrium outcome in all sectors of all countries. In the IRS-based model from equation (2.1), if there more countries and/or more goods, country i will import the same share s^i of total production of each, in the same way country i imports that share of country-j's varieties X and Z. Product specialization for all goods due to factor proportions differences will not necessarily occur when there are more than two countries (or goods), but only two factors of production. With more than two factors, factor-specificity can generate product specialization. In an extreme case where the production of any good in any country requires inputs that are specific to that good in that country (such as technological know-how), equation (2.3) will reappear. Therefore, for an appropriately defined metric of factor endowment differences, also the H-O rationale for the gravity equation generalizes to a multi-sector, multi-country, and multi-factor setting.

So far our discussion has left open the question of whether one can derive gravity-type import volume predictions for the case where at least some goods are produced in more than one country. While product specialization has recently emphasized to be an important phenomenon of the patterns of world production (e.g., Haveman and Hummels 1997), no doubt there are many commodities which are produced in several countries. We now turn to two of those trade models with imperfect product specialization for the $2 \times 2 \times 2$ case: the simplest, textbook Heckscher-Ohlin model, as well as its generalization to include a IRS sector due to Helpman (1981).

2.2. Gravity Equations with Imperfect Product Specialization

First, we derive a gravity equation for the generalization of the H-O model where 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, Ch.7,8). There are two countries (i and j) and two factors (K and L). For endowments inside the factor price equalization set, the volume of bilateral trade, defined as the sum of a country's exports and imports, is given by

$$VT^{ij} = s^j p_x X^i + s^i p_x X^j + \left(Z^j - s^j Z^w \right), \tag{2.6}$$

where the first term on the right hand side gives country i's exports (M^{ji}), and the remaining two terms are country i's imports (M^{ij}).[6]

Let γ^c be the share of good Z in country c's GDP, $\gamma^c = \frac{Z^c}{p_x X^c + Z^c}$. It is easy to show that the assumption that country i is capital-abundant implies that $0 \leq \gamma^i < \gamma^j$. The following reproduces a result from Keller (1998).

Proposition 1. *If good X is capital-intensive and produced under IRS, good Z labor-intensive and produced under CRS, and country i is relatively capital-abundant, then country i's imports from country j are given by the following gravity equation:*

$$M^{ij} = \left(1 - \gamma^i\right) \frac{Y^i Y^j}{Y^w}. \tag{2.7}$$

Proof. With balanced trade, $M^{ij} = M^{ji}$; $M^{ji} = s^j p_x X^i$, and using the definition of γ^i gives $M^{ji} = s^j \left(1 - \gamma^i\right) Y^i = \left(1 - \gamma^i\right) \frac{Y^i Y^j}{Y^w}$. ■

This proposition states that for any value $\gamma^i > 0$, the level of bilateral imports is lower than in

[6]With zero trade costs, Z being homogeneous and produced under CRS, and factor price equalization, consumers are indifferent with regard to Z's country of origin and producers are indifferent with respect to who buys it. Therefore, underlying equation (2.6) is a "minimum trade-volume" assumption, that only a country's excess demand for a good is imported from abroad. It is well-known that the H-O model determines only net, but not gross, trade flows.

the case where both goods are differentiated (compare 2.7 to the gravity equation 2.3). Furthermore, as the share of homogeneous good production in GDP declines, the predicted level of imports rises, and in the limit, as $\gamma^i \to 0$, the generalized gravity equation (2.7) reverts back to the simple gravity equation (2.3) above. In a sense, therefore, the volume of trade is higher, the lower is the share of homogeneous goods in GDP.[7] It is also interesting to note that except for country j's GDP, only the characteristics of the capital-abundant country i enter the bilateral imports prediction (2.7).

Second, we show the particular form of the gravity equation in the case of a simple $2 \times 2 \times 2$ H-O model; that is, now both good Z and good X are homogeneous and produced under CRS. The volume of trade between countries i and j is given by

$$VT^{ij} = p_x \left(X^i - s^i X^w\right) + \left(Z^j - s^j Z^w\right), \tag{2.8}$$

where, according to the H-O theorem, the capital-abundant country i exports the capital-intensive good X and imports the labor-intensive good Z.[8] For this model, the following proposition, reproduced from Keller (1998), states the corresponding gravity-type relationship.

Proposition 2. *If both goods are homogeneous and produced under CRS, with country i being relatively capital-abundant and good X being relatively capital-intensive, then country i's imports from country j are given by the following gravity equation:*

$$M^{ij} = \left(\gamma^j - \gamma^i\right) \frac{Y^i Y^j}{Y^w}. \tag{2.9}$$

Proof. See Appendix B.

[7] Note that this finding is in part due to H-O 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' factor proportions differences (triggering a production response a la Rybczynski). What we emphasize is that for given factor proportions differences, the more product specialization there is the higher is the level of imports; see below.

[8] The "minimal trade-volume" rule has again been used to derive (2.8).

The gravity equation in the imperfect specialization H-O case, equation (2.9), depends not only on the product of the GDP's in the familiar way, but also on γ^j and γ^i, which are characteristics of both countries. Note that 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 (2.9) gives the familiar result that there is no trade in a H-O model when factor proportions are identical across countries. Equation (2.9) includes the volume of imports prediction of the multi-cone H-O model given in (2.5) as a special case, because as factor proportions differences between i and j increase, the share of GDP derived from good Z in country j, γ^j, approaches one, whereas the share of good Z in the GDP of country i, γ^i, tends to zero. Indeed, when $\gamma^j = 1$ and $\gamma^i = 0$, equation (2.9) reverts to $M^{ij} = \frac{Y^i Y^j}{Y^w}$,the gravity equation for the perfect specialization model (see 2.5).[9] Also, we can rewrite equation (2.9) in the following way:

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

Therefore, if we denote the volume of imports prediction for the case where the production of both goods is specialized across countries, equations (2.3) and (2.5), by M_S; the generalized H-O case with specialization for one good, but not the other good (equation 2.7) with M_G; and the H-O case in which both countries produce all goods, equation (2.9), with M_H, the following inequalities hold, ceteris paribus:

$$M_S > M_G > M_H. \tag{2.10}$$

This confirms that the bilateral volume of imports is higher, the more product specialization there is.

[9] This assumes the generic case where *both* countries specialize in production; see above.

2.3. Model Identification

In the preceding section, we have derived the specific form of the gravity equation of trade for four models: (1) Multi-cone H-O: factor endowment differences lead to perfect product specialization; (2) Pure IRS: IRS leads to perfect product specialization; (3) Generalized H-O: with one good produced using IRS technology and perfectly specialized, the other good being produced with CRS technology and not specialized, and (4) Uni-cone H-O: where both goods are produced using CRS technology and are not specialized. If we restrict ourselves to $2 \times 2 \times 2$ models, then M_S is the import prediction for models (1) and (2), M_G is the prediction for model (3), and M_H is the prediction for model (4).

It is unlikely, however, that any of the observed trade flows are solely determined by any one of these four archetypal models. First of all, the data comes from a world with more sectors, countries, and factors than our $2 \times 2 \times 2$ models. Secondly, there may be positive amounts of IRS-based trade even between countries with the lowest recorded shares of IRS-based trade. Similarly, we expect there to be some factor-proportions-based trade even among countries where the share of IRS-based trade is highest.[10] Observed bilateral trade among countries is likely to be the result of the combination of the determinants of trade flows formalized in the four models considered here, and perhaps by other determinants which we have not addressed. However, in different circumstances (such as different distributions of factor endowments across trading partners) we expect different trade models to account for different proportions of the observed trade flows. Our inferences are based on whether each of the four trade models actually performs better in the very sample(s) (which are identified using explicitly stated criteria) where one expects the theory to perform better.

Consider a cross-section of country-pairs where there is little (or no) product specialization due to IRS, but the absolute difference between the two countries' factor proportions, denoted $FDIF$, differs from one pair to another.[11] From equation (2.4), we expect *ceteris paribus* more product specialization

[10] Note that this point does not rule out the generalized H-O model (3).

[11] We treat from now on a country-pair observation drawn from the actual multilateral trading world as resembling the

in a pair where $FDIF$ is larger relative to another country-pair where factor proportions differences are smaller. Consequently, we expect the gravity prediction (2.3), i.e., prediction M_S, to be more accurate for the pair where $FDIF$ is larger, compared to the other with the lower value of $FDIF$. This observation allows to identify the H-O motivation for product specialization and the gravity prediction: If factor proportions differences are important in explaining the success of the gravity equation, then the prediction M_S should fit better in samples where the observed factor proportions differences $FDIF$ are higher.[12]

Moreover, we can test the multi-cone H-O model against the uni-cone H-O model for different values of factor proportions differences: If product specialization is principally the result of factor proportions differences, then the uni-cone H-O model should be preferred to the multi-cone H-O model in samples where the observed factor proportions differences $FDIF$ are lower.

We will employ the index proposed by Grubel and Lloyd (1975) to control for the extent of IRS-based trade. Trade due to IRS and product differentiation can result in a country simultaneously importing and exporting varieties of a particular product (intra-industry trade).[13] The index, denoted GL^{ij}, measures the share of intra-industry trade in the total trade

$$GL^{ij} = 1 - \left(\frac{\sum_g \left| M_g^{ij} - M_g^{ji} \right|}{\sum_g \left(M_g^{ij} + M_g^{ji} \right)} \right), \quad 0 \leq GL^{ij} \leq 1. \tag{2.11}$$

In the extreme case where a given good g is either exported or imported (no intra-industry trade),

two-country world observations of models (1)-(4). There is an empirical and a theoretical concern, only the first of which applies to the models with perfect product specialization. First, trade between nations is often not balanced. Trade imbalance effects, however, are too small to affect our qualitative results. Second, if a homogeneous good is produced in two or more countries, then bilateral trade volumes may become indeterminate (this led to the "minimal-trade volume rule" above). We cannot be sure that, in a multi-country model, bilateral trade volumes are determined by the same rule. For a given rule, though, we conjecture that the two-country predictions are closely related to the bilateral predictions from a multilateral model, and plan on working to show this rigorously in the future.

[12] This result holds *ceteris paribus*: As discussed above, perfect specialization in homogeneous good production can be obtained not only through large factor proportions differences. Our empirical strategy accounts for that, see below.

[13] We define intra-industry trade as trade in goods with identical factor input requirements; for our empirical analysis, though, intra-industry trade is taken as two-way trade of goods in the same four-digit SITC class. The two concepts need not be the same in a deterministic sense, but our analysis remains valid as long as they are the same on average; see below.

the Grubel-Lloyd index will be equal to zero. With positive amounts of intra-industry trade, it will be between zero and one;[14] it is generally higher when the share of intra-industry trade in total trade is higher.

The share of intra-industry trade is not a perfect indicator for the share of trade based on IRS. Finger (1975) has argued that intra-industry trade is found because products which are actually different are classified to the same industry. Clearly, a high degree of disaggregation is desirable when studying intra-industry trade. Recently, researchers have emphasized the importance of trade in intermediate goods (vertical differentiation) in accounting for a high values of GL (Greenaway, Hine, and Milner 1994, and Fontagne, Freudenberg and Peridy 1997); also this can in principle be due to IRS *or* other reasons. Moreover, a significant part of trade which is classified as inter-industry trade might be based on IRS, such as the example of wide-bodied aircraft exports from the U.S. to most other countries.[15] At the macro level, the presumption that GL falls with larger factor proportions differences across countries has generally been confirmed in the literature.[16] At the micro level, the evidence on whether GL correlates positively with economies of scale at the industry level is mixed,[17] which in part is due to obtaining good measures which can identify IRS. After discussing the issue in some detail, Krugman (1994, 23) concludes that although GL does not exactly measure the share of international trade which is due to IRS, this true share might be higher *or* lower. In this study, we use the GL index to infer the importance of IRS-based trade in total trade using a methodology which accounts for the fact that GL is an imperfect measure of that.

[14] In the pure IRS model (1), the maximum value of $GL^{ij} = 1$ is obtained if the share of GDP derived from producing X-varieties is the same in both countries; in the $2 \times 2 \times 2$ model, this will happen if factor proportions are identical in the two countries.

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

[16] See Helpman (1987), Bergstrand (1990), and Fontagne, Freudenberg, Peridy (1997); the results in Hummels and Levinsohn (1995) are mixed. The presumption can be established rigorously in the two country, two factor, two industry case, see Helpman (1981).

[17] An example of a recent study is Fontagne, Freudenberg and Peridy (1997), who find a strong positive correlation of the share of two-way trade and a proxy of IRS at the industry-level, but this is not true for all recent studies, see their references, and also Krugman (1994).

The Grubel-Lloyd index is employed to identify the samples where the pure IRS model is likely to determine trade flows. Specifically, we expect that the IRS model accounts for the performance of the gravity equation in those samples where the bilateral Grubel-Lloyd indices are larger, indicating that a larger proportion of bilateral trade is two-way trade in perfectly specialized differentiated products. In section 4.1.2 we examine whether in fact the prediction M_S (of the IRS model) is more accurate in samples with higher Grubel-Lloyd indices.

We emphasize that prediction M_S is common to both the multi-cone H-O and the pure IRS models. However, if we were to find that the prediction M_S is less at odds with the data in samples with higher Grubel-Lloyd indices, it would be incorrect to interpret this finding as evidence in favor of the multi-cone H-O theory.[18] Furthermore, we can test the generalized H-O model against the pure IRS model. We expect the latter model to be preferred to the former model in samples where there is a greater share of intra-industry trade, as measured by the Grubel-Lloyd index.

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

3. Data

We have developed a cross-sectional data set for fifty-eight countries for the year of 1985. The data set includes all countries with both GDPs above 1 billion US dollars and where internationally comparable capital-per-worker estimates (measured in US dollars) are available. These fifty-eight countries account for 67% of world imports and 79% of world GDP in 1985. The countries are listed in Table 1A. The data set includes nearly all industrialized countries, but relatively few of the less developed countries. This reflects the paucity of capital stock estimates available from the latter.

The source for data on GDP and capital-per-worker is the current Summers and Heston (Version 5.6) database (see Summers and Heston 1991 for a description of this database). Both variables are

[18]This is so unless there was a perfect positive correlation across country pairs between the observed Grubel-Lloyd indices and factor endowment differences–which there is not.

in internationally comparable (purchasing power adjusted), real US dollars for the year 1985.[19] Each country's GDP and capital-per-worker data are reported in Table 1A. The trade data comes from the NBER World Bilateral Trade Database, see Feenstra et al. (1997). The overall imports of country i from country j (M^{ij}) are immediately available from this source. We have used the values reported by the importing country, as these are known to be more accurate than trade reported by the exporting country. With fifty-eight countries, we have $58 \times 57 = 3306$ bilateral import relations.

The bilateral Grubel-Lloyd indices GL^{ij} are also derived from data in Feenstra et al. (1997). These indices are computed using all goods at the four-digit SITC classification; that is, we do not confine ourselves to the manufacturing sector as is frequently done when Grubel-Lloyd indices are calculated. The Grubel-Lloyd index (2.11) can only be computed for country-pairs where there are positive amounts of trade. This is the case for 87% of all bilateral relations, so that our sample consists of 2870 observations.[20] The number of bilateral trade relations with partner countries varies by country from a low of 27 for Nepal to the maximum of 57 for most industrialized countries. Furthermore, the number of different goods classes traded with a partner country varies from country to country. At one end, Mauritius trades on average 36 types of goods with its partners, whereas at the other end, the U.S. trades on average 332 types of goods with its trade partners.

Keeping these facts in mind, the average Grubel-Lloyd indices we compute are presented in Figure 2. Across each of their respective trading partners, Bolivia has the lowest average Grubel-Lloyd index in this sample, with a value of 0.0006; and the U.K. has the highest value of 0.1495. Each nation's average Grubel-Lloyd index is shown in Table 1A, along with national GDP per capita, the number of bilateral relations the average is computed from, and the average number of industry classes where there is positive trade. In Table 1B we show the correlation among the variables of Table 1A. All

[19]One reason of using data for 1985 is that this is the benchmark year for inflation and PPP adjustments in the Summers and Heston dataset.

[20]In addition to the 2870 bilateral observations for which we can compute the Grubel-Lloyd index, the NBER database reports positive levels of trade at an aggregate level for three more pairs: Kenya/Sierra Leone, Madagascar/Ireland, and Zambia/Guatemala. We have not been able to track down the reason for this inconsistency, but do not expect that it will affect the results significantly.

correlations are positive; in particular, the correlation of the average Grubel-Lloyd index with GDP per capita is 0.78, and the correlation between the average Grubel-Lloyd index and the average number of industry classes traded is 0.93.

In Figure 3, we plot the cumulative distribution of bilateral Grubel-Lloyd indices (there are 1435 distinct indices, corresponding to 1435 country-pairs in our sample). The distribution is very skewed: 44% of all indices are equal to 0, and 78% of all (1120 country-pairs) have a value of 0.05 or less. In the empirical analysis below, we will treat these 1120 country-pairs as trade relations where there is little or no IRS-based trade, whereas the remaining 315 country-pairs with Grubel-Lloyd indices above 0.05 are taken as those where IRS-based trade is present .[21] It is important to note that this separation of the data set into 1120 and 315 country-pairs does not mean that IRS-based trade must account for a small share of international trade; in fact, the 315 country-pairs with Grubel-Lloyd indices above 0.05 account for 87.1 % of all of the imports by the 1435 country-pairs.

We are concerned about heteroskedasticity in our sample. To that end, let the total imports of country i and total exports by country j be denoted by $\bar{M}^i$ and $\bar{X}^j$, respectively. Then, we scale all the import values M^{ij} and the GDP term $\frac{Y^i Y^j}{Y^w}$ by $\psi^{ij} = (\bar{M}^i \times \bar{X}^j)^{0.5}$ and assume that this creates a scalar variance in regressions based on the Gravity Equation prediction. Finally, the gravity prediction gives a relationship between two endogenous variables, so the results might be affected by simultaneity bias (as discussed by Saxonhouse 1989 and Harrigan 1996). However, while in a world with positive levels of production of all goods in all countries[22] factor endowments might be valid instruments for countries' GDPs, here we are especially concerned with perfect product specialization across countries. Because this makes factor endowments much less desirable instruments, we have decided to leave the gravity prediction un-instrumented.

[21]The threshold level of the Grubel-Lloyd index of 0.05 is based on our priors. In Appendix C we show that our main findings do not depend on this choice of threshold level.

[22]This and other conditions need to be satisfied for factor endowments to be linearly related to GDPs with equal coefficients across countries; see, e.g., Harrigan (1996).

4. Empirical Results

4.1. Perfect Specialization Models

4.1.1. Heckscher-Ohlin-Based Product Specialization (Low-GL Sample)

We first consider the bilateral pairs where the computed Grubel-Lloyd (GL^{ij}) indices are 0.05 or less. As discussed above, those are the country-pairs in which we expect only negligible amounts of IRS-based trade to be present. We assume that the expected value of the true share of trade in perfectly specialized goods rises with the measured bilateral factor endowment differences. Specifically, the true share of trade in perfectly specialized goods (denoted $^{*}FDIF^{ij}$) is determined by the observed level of factor proportion differences ($FDIF^{ij}$) plus a confounding variable:

$$FDIF^{ij} = {}^{*}FDIF^{ij} + \theta^{ij}, \forall ij, \tag{4.1}$$

where we assume that $\theta^{ij} \sim N(0, \sigma_\theta^2)$, $\sigma_\theta^2 > 0$. The confounding variable θ^{ij} accounts for two effects: First, it captures measurement error in the underlying capital-labor ratio data. Second, it picks up changes in the share of product-specialized trade which are due to effects other than changes in factor proportions differences; the latter could be due, e.g., to Ricardian technology differences, bilateral distance, demand or relative country size for that pair ij. With $E\left[\theta^{ij}\right] = 0, \forall ij$, we assume that *on average* the effect of these influences, and all other determinants of the share of trade in perfectly specialized goods, is equal to zero.

Essentially, we want to examine how the estimated parameter in the gravity regression (2.3) varies for subsamples of the data as $FDIF$ changes. However, for the reasons given above, $FDIF^{ij}$ is only a noisy signal of the share of perfectly specialized trade in total trade in country-pair ij. Therefore, even if the relationship between the gravity equation estimate and the true index were strong and monotonic, $FDIF^{ij}$ need not vary monotonically with the gravity equation estimate. We are interested to see whether there is a robust relationship between $FDIF^{ij}$ and the gravity equation parameter. The following approach, which combines elements of resampling (see Efron 1982) and non-parametric regression analysis (e.g., Härdle 1990), suits this need.

We first create, using (4.1), artificial distributions for $FDIF$ which are centered on the data we have collected from the Summers and Heston dataset. Second, for a specified number K of equally sized $FDIF$ classes, we collect all the observations which fall into the kth class, $k = 1, ..., K$. Third, from these K sets of observations, we compute K sets of import parameters corresponding to the gravity predictions of the models under consideration.

To implement this, we rank the 2240 country-pair observations under consideration (those with $GL^{ij} < 0.05$) by the log of the factor proportions difference we obtained from the Penn World Tables. Then, for a given Monte-Carlo replication $r, r = 1, ..., R$, we draw a (2240×1) vector of θ^{ij} and add this to the factor proportions differences column. This is our factor proportions differences vector for that replication, $FDIF^{ij}(r)$. In the following step, we sort the data on the column $FDIF^{ij}(r)$. The kth class, $k = 1, .., K$, class of observations of this replication is denoted $M_k^{ij}(r)$ and $\frac{(Y^iY^j)_k}{Y^w}(r)$. We repeat this for the following replications $r+1, r+2, ..., R$, collecting all bilateral relations falling into a given group k across all R replications. Finally, with these sets of data on imports and GDP's, we run the following least squares regression

$$y_k^{ij}(R) = \alpha_k \, x_k^{ij}(R) + \epsilon_k^{ij}, \forall k = 1, ..., K, \tag{4.2}$$

where $y_k^{ij}(R)$ is the set of all $M_k^{ij}(r)$, and $x_k^{ij}(R)$ is the set of all $\frac{(Y^iY^j)_k}{Y^w}(r)$ across all replications which fall into percentile class k, where ϵ_k^{ij} has mean zero.[23]

Recall that the model under consideration here is the multi-cone H-O model where all trade

[23]Note that ϵ_k^{ij} is not independent, because the resampling procedure leads by design to observations being repeated in a given class k, as in standard bootstrapping techniques. The variance on θ^{ij} which we assume determines the degree of smoothing, related to the choice of the bandwidth in non-parametric regression analysis. What is different here is that we apply the smoothing to the identifying variable $FDIF$, not to the variables entering the regression equation.

is in (homogeneous) specialized products. The imports prediction is given by (2.3): $M^{ij} = \frac{Y^i Y^j}{Y^w}$. Therefore, in the multi-cone H-O model, α_k is only one parameter, predicted to equal 1 when all goods are tradable.

Figure 4 shows the result for $K = 5$ classes, with average values of $FDIF$ rising from class $FDIF = 1$ to $FDIF = 5$.[24] The line indicates how the estimated value of α_k varies as $FDIF$ changes. First, note that the estimated values are all around 0.024, which is much smaller than the predicted value of $\alpha = 1$.[25] Recall, however, that we are interested primarily in how α varies across $FDIF$ class, not whether we can reject the hypothesis of $\alpha = 1$. One sees that the value of α_k does not vary systematically with k, the index of the $FDIF$-class. This implies that as there is more product specialization due to factor proportions differences in the sample, the estimated parameter does not move closer to its theoretically predicted value. This is inconsistent with product specialization due to factor proportions differences being an important component in explaining the success of the gravity equation.[26]

We now turn to examining the Pure IRS model.

4.1.2. IRS-Based Product Specialization (High-GL Sample)

Our approach here is analogous to the one used above. We assume that the actual share of IRS-based trade in total trade (denoted $^*GL^{ij}$) in a given country-pair equals the calculated Grubel-Lloyd index

[24] The graphs are based on simulations with $R = 15$, $K = 5$, $\sigma_\theta^2 = 5$. We have experimented with different choices on these parameters, but that did not qualitatively change the results noted in the text. The log of $FDIF$ has a mean of 9.31 and a standard error of 1.21 in this sample of low-GL country-pairs.

[25] The gravity equation regressions conducted here correspond to the bilateral trade volume expressions we derived above; we include therefore none of the other variables frequently found in empirical gravity regressions, such as distance or adjacency dummies. Also note that there is no constant in equation (2.3). Empirical gravity regressions are typically done in logs, estimating the elasticity of imports with respect to the GDP term; from our results, this elasticity ranges between 0.44 and 0.47 for the five $FDIF$ classes, which is somewhat lower than the standard results in the range of 0.7 to 1.0. One of the reasons for our lower elasticity estimate is likely to be that, contrary to most gravity regressions, we constrain the elasticity to be the same for both exporting as well as importing country GDP.

[26] The standard errors of the estimates α_k for the multi-cone H-O model are as follows; FDIF=1: 0.0008, FDIF=2: 0.0008, FDIF=3: 0.0008, FDIF=4: 0.0007, FDIF=5: 0.0007. Using these standard errors and the estimated parameters, therefore, α_5 is not significantly larger than the α_1 in a statistical sense ($\alpha_1 = 0.0242 + 1.96 \times 0.0008$ gives 0.0258, which is larger than the point estimate of α_5, with 0.0245).

(denoted GL^{ij}) plus a confounding variable:

$$GL^{ij} = {}^{*}GL^{ij} + \chi^{ij}, \forall ij,$$

where $\chi^{ij} \sim N(0, \sigma_{\chi}^2)$, with $\sigma_{\chi}^2 > 0$. As noted above, the Grubel-Lloyd index imperfectly captures the share of IRS-based trade in total trade. Therefore, χ captures not only measurement error due to aggregation and the existence of vertically differentiated products, but also the extent of two-way trade that is due to comparative advantage. With $E\left[\chi^{ij}\right] = 0$, though, we assume that, *on average*, the calculated Grubel-Lloyd index determines the share of IRS-based trade in total trade.

In this subsample, there are 630 country-pair observations accounting for 87% of all imports. We classify the 630 observations into $K = 5$ classes, with the Grubel-Lloyd indices rising from class $K = 1$ to class $K = 5$. The model considered here is the pure IRS model, where both X and Z are produced under economies of scale. The model implies a volume of imports prediction of $M^{ij} = \frac{Y^i Y^j}{Y^w}$. The regression is again

$$y_k^{ij}(R) = \alpha_k \, x_k^{ij}(R) + \epsilon_k^{ij}, \forall k = 1, ..., K, \tag{4.3}$$

with $x_k^{ij}(R)$ is equal to the set of all $\frac{\left(Y^i Y^i\right)_k}{Y^w}(r)$ in this percentile class k.

Figure 5 reports the results of this resampling experiment.[27] The line shows for the pure IRS model how the estimated value of α_k varies as the Grubel-Lloyd index increases. The parameter is estimated to be between 0.03 to 0.13, which is higher than the comparable parameter estimates for the multi-cone H-O model. Second, the estimate rises in samples with higher values of the Grubel-Lloyd index; the higher the share of IRS-based trade, the closer is the estimated parameter to the theoretical value of $\alpha_k = 1$. Therefore, this finding is consistent with IRS-based trade theory being the reason why the gravity prediction 'works' empirically. Note, however, that the parameter is never anywhere

[27] This is based on choosing $R = 70$, $K = 5$, and $\sigma_{\chi}^2 = 0.15$; the noted features are not sensitive to this choice. The average (standard deviation) of GL is equal to 0.199 (0.133) for these high$-GL$ country pairs.

close to 1, the theoretically predicted value when all goods are tradable.[28]

Summarizing, from our analysis of the H-O and IRS-perfect specialization models, we find that there is no evidence suggesting that perfect product specialization due to H-O reasons is important in explaining the success of the gravity equation; in contrast, there is some evidence supporting the Pure IRS model's trade volume prediction. However, the trade volume prediction of both of these models holds only if all goods are perfectly specialized. We now turn to estimating the import volume predictions of the two above models which incorporate imperfectly specialized production, the Generalized H-O model with import prediction M_G, and the H-O model with import prediction M_H.

4.2. Imperfect-Specialization Models

4.2.1. Heckscher-Ohlin-Based Product Specialization (Low-GL Sample)

We return to the sample of bilateral pairs where the computed Grubel-Lloyd (GL^{ij}) indices are 0.05 or less, and adopt an analogous approach by estimating gravity equations for samples which differ by their average value of $FDIF$, indicating the extent of product specialization due to factor proportions differences. In the uni-cone H-O model, all trade is in imperfectly specialized products; its import prediction M_H is equal to $M^{ij} = (\gamma^j - \gamma^i)\frac{Y^i Y^j}{Y^w}$ if country i is capital abundant and X is capital-intensive, and $M^{ij} = (\gamma^i - \gamma^j)\frac{Y^i Y^j}{Y^w}$ if country j is capital-abundant. Therefore, we estimate (up to) fifty-eight parameters γ, the share of the labor-intensive good Z in the countries GDP's.

The regression is analogous to (4.2),

$$y_k^{ij}(R) = \alpha_k x_k^{ij}(R) + \epsilon_k^{ij}, \forall k = 1, ..., K,$$

with $x_k^{ij}(R)$ equal to the set of all $(\gamma^j - \gamma^i)\frac{Y^i Y^j}{Y^w}(r)$ (or $(\gamma^i - \gamma^j)\frac{Y^i Y^j}{Y^w}(r)$), depending on which country

[28]The standard errors of the estimates α_k for the Pure IRS model are as follows; GL=1: 0.0006, GL=2: 0.0007, GL=3: 0.0009, GL=4: 0.0012, GL=5: 0.0022. This means that α_5 is also significantly larger than α_1 in a statistical sense.

is capital-abundant and which labor-abundant) in this class k.

Figure 6 shows the average of the estimated differences $\gamma^j - \gamma^i$, that is the difference in the share of the labor-intensive good in the labor-abundant country from that in the capital-abundant country. This average difference is estimated to be in order of 0.039 to 0.051. Furthermore, the average difference increases as the average value of factor proportions differences rises in the sample. This is what the theory (in the $2 \times 2 \times 2$ case) predicts: Larger differences in factor endowment ratios are mapping into larger differences in the shares of GDP derived from a given good.[29]

4.2.2. IRS-Based Product Specialization (High-GL Sample)

In the high-GL subsample, there are 630 country-pair observations. As above when examining the Pure IRS model, we classify the observations into $K = 5$ classes, with the Grubel-Lloyd indices rising from class $K = 1$ to class $K = 5$. The model considered here is the generalized H-O model (where the capital-intensive X sector in IRS, whereas the labor-intensive Z sector is assumed to be labor-intensive.) It implies imports at a level $M^{ij} = (1 - \gamma^i)\frac{Y^i Y^j}{Y^w}$ if country i is capital-abundant relative to j , and $M^{ij} = (1 - \gamma^j)\frac{Y^i Y^j}{Y^w}$ otherwise. The regression is analogous to (4.2), with the set of the $(1 - \gamma^i)\frac{Y^i Y^j}{Y^w}(r)$ (or $(1 - \gamma^j)\frac{Y^i Y^j}{Y^w}(r)$, if country j is relatively capital-abundant) for a given class k as the independent variable.

Figure 7 reports the results of this resampling experiment.[30] The line shows the average estimate of the share of the IRS sector in the capital-abundant country, $(1-\gamma^i)$, across the five different classes. This share is estimated to be between 0.43 and 0.78, and tends to rise as the Grubel-Lloyd index rises. Note that from country-pair to country-pair there are changes as to which of the two, exporter or

[29] The resampling experiment is identical to that for the other model in the Low-GL sample, see footnote 22. The average standard errors of the estimates α_k for the uni-cone H-O model are as follows; FDIF=1: 0.015, FDIF=2: 0.014, FDIF=3: 0.016, FDIF=4: 0.013, FDIF=5: 0.011. Based on the parameter estimates and these standard errors, the true α_5 is only marginally higher than the true α_1 at standard levels of statistical significance. Compared to the multi-cone H-O experiment where one parameter is estimated, here, we estimate 58 parameters; therefore, in order to obtain comparable standard errors, the number of replications should now be higher–leading to lower estimated standard errors–than in the multi-cone H-O case above. The limitations imposed by the computer hardware we use have prevented this estimation.

[30] This is based on choosing the same $R = 70$, $K = 5$, and $\sigma_\chi^2 = 0.15$ as in the above analysis of the Pure IRS model.

importer, is the relatively capital-abundant country. Therefore, a non-structural explanation of the average of the estimated $(1-\gamma^i)$ is that it gives the average size of the share of the differentiated goods sector. According to this interpretation, we estimate a higher GDP share for the IRS-sector in samples with higher values of the Grubel-Lloyd index. This is consistent with the generalized H-O theory, where a higher share of intra-industry trade is driven by IRS and product differentiation.[31]

Summarizing, from our analysis of imperfect-specialization models, we find supportive evidence for factor-proportions driven product specialization both from the Generalized H-O model as well as from the uni-cone H-O model.

4.3. Comparing Perfect and Imperfect-Specialization Models: Model Selection Criteria

We now employ a widely-used model selection criteria to compare the performance of both the multi-cone H-O and uni-cone H-O models for the low-GL sample on the one hand, and the Pure IRS and the Generalized H-O for the high GL-sample on the other. The Amemiya prediction criterion (denoted as APC) is defined as

$$APC = \frac{e'e}{v-w}\left(1+\frac{w}{v}\right),$$

where $e'e$ is the residual sum of squares, v is the number of observations, and w is the number of estimated parameters.[32] The APC criterion penalizes more heavily losses of degrees of freedom than the adjusted R^2 criterion; lower values indicate better model performance.

Figure 8 shows that the uni-cone H-O model is preferred to the multi-cone H-O model across all $FDIF$ classes. This figure also shows the relative performance of the multi-cone to the uni-cone H-O model across FDIF classes, computed as the ratio of the APC criteria of the two models for a given

[31] The average standard errors of the estimates α_k for the Generalized H-O model are as follows; GL=1: 0.054, GL=2: 0.065, GL=3: 0.105, GL=4: 0.227, GL=5: 0.837. With these standard errors, we estimate α_1 to be lower than 0.78, our point estimate for α_5, in a statistical sense. Note, however, that one cannot reject the hypothesis that $\alpha_5 = 0$ at standard significance levels. This is so because the generalized H-O model fits relatively badly for samples with high GL-values, and the pure IRS model fits increasingly better; see below.

[32] We have also considered the Akaike Information Criterion; using that gives very similar results as with the Amemiya criterion, so we report only the latter.

FDIF class (the vertical axis on the right). There is not much change of the relative performance of the multi-cone model across FDIF classes; if anything, the multi-cone H-O model performs relatively best where factor proportions differences are smallest ($FDIF = 1$), and relatively worst where they are largest ($FDIF = 5$), which is the *opposite* of what one would expect if factor proportions differences would be important in driving trade in perfectly specialized products.

The corresponding picture for the Pure IRS and the Generalized H-O model is shown in Figure 9, for different values of GL, indicating a different share of IRS-based trade in total trade. The generalized H-O is preferred to the pure IRS model for all five classes. As in the previous subsection, the model with imperfect specialization performs better. Still, there is an important difference relative to the samples with predominantly H-O-based trade: the relative performance of the generalized H-O model is better in samples with lower Grubel-Lloyd indices. That means that although the model in which all goods are produced with IRS is never as successful as the generalized H-O model, the pure IRS model puts in its best relative performance where it should (namely, in the sample with the highest value of the Grubel-Lloyd index).

Overall, we take these results as evidence which, first of all, confirms that perfect product specialization due to H-O reasons is unlikely to be an empirically important determinant of international trade volumes, whereas IRS appears to be, especially for highly-industrialized countries. Second, perfect specialization of production occurs only for a limited range of goods, and factors leading to imperfect product specialization, such as (limited) factor proportions differences, are important in accounting for the volume of international trade.

4.4. Is the Effect of Distance Driving the Results?

One potential problem with the above analyses is that they do not control for differences across $FDIF$- or GL-classes in the average distance between countries. It is well-established empirically that bilateral trade volumes fall as bilateral distance increases (the second pillar of the gravity equation), raising the possibility that the different estimated parameters are driven by differences in bilateral distance among trade partners.

Even though the import volume predictions presented above do not incorporate the effect of trade costs, at an empirical level it is important to see whether our results depend crucially on the fact that we omit the bilateral distance variable from our gravity regressions. There are two aspects to this question: First, is the average bilateral distance varying *between* $FDIF$- or GL-classes, and if so, how does this affect our findings? Second, does the fact that our analysis ignores the variation of bilateral distance across country-pairs *within* a given $FDIF$- or GL-class explain our findings?

It turns out that the average bilateral distance between trade partners does not vary much between $FDIF$-classes, whereas it does change between GL-classes.[33] However, the analysis in Appendix D shows that controlling for between- and within-class effects related to bilateral distance differences in empirical gravity regressions leaves our results unaffected. While this cannot settle the question of how structural bilateral import relations look in the presence of trade costs, it means that our earlier findings are robust to incorporating distance effects.

5. Conclusions

We have compared the accuracy of the perfect specialization versions of the Heckscher-Ohlin model and the Increasing Returns-based model, both of which predict the gravity equation. In addition, we have demonstrated that, as far as their predictions for bilateral trade volumes are concerned, each of these perfect specialization models is a limiting case of a model with imperfect specialization. Our empirical strategy has exploited two factors to identify which models might explain trade flows in a given sample. First, in the multi-cone H-O model large factor proportions differences are required

[33]For $FDIF$-classes, it hovers around 8800 miles, with a maximum average of 8840 miles for $FDIF$=5 and a minimum average of 8760 miles for $FDIF$=2; for GL-classes, it falls from about 6300 miles for GL=1 to circa 3000 for GL=5. We would like to thank Jon Haveman at Purdue University for making the distance data available at his webpage (http://intrepid.mgmt.purdue.edu/Jon/Data/TradeData.html#Gravity); the data gives the Great Circle distance between capital cities, as the crow flies.

to induce product specialization, whereas product specialization in the pure IRS model occurs for arbitrary differences in factor proportions. Secondly, there is no intra-industry trade in the multi-cone H-O model, whereas there is intra-industry trade, and possibly *only* intra-industry trade, in the Pure IRS model.

There are three major findings: First, little production is perfectly specialized due to factor proportions differences, making the perfect specialization version of the H-O model an unlikely candidate to explain the empirical success of the Gravity Equation. Secondly, increasing returns are important causes for perfect product specialization and the Gravity Equation, especially among industrialized countries. Thirdly, models of imperfect specialization better explain the variation of bilateral trade flows than perfect specialization models. Factor proportion differences are important determinants of trade flows within the context of imperfect specialization models only, whereas increasing returns is a cause of product specialization both in models with imperfect as well as with perfect specialization of production.

In the light of this paper, there is no reason to believe that the results by Hummels and Levinsohn (1995) throw doubt on the empirical relevance of IRS trade theory. Our results suggest that the H-O model with imperfect specialization is likely to account for the regression results that these authors found in their sample where little intra-industry trade was expected. More generally, the data is supportive of several predictions of IRS-based trade theory, and we take our results as indicating that IRS-based trade plays, in conjunction with trade based on factor proportions differences, an important role in determining world trade flows.

Several important determinants of trade flows are missing from our analysis. First of all, we do not allow for preferences to differ across countries, even though other studies have found that countries tend to overproportionately demand home produced goods (or, exhibit a 'home bias'). This appears to be one reason of why standard trade models predict trade volumes which are far larger than what is found in the data (Trefler 1995). Secondly, a related point is that we do not consider explicitly models with transport costs and trade barriers. We know that the latter are empirically relevant and can also lead to a home bias. However, CRS production of the same good in two or more countries in the presence of transport costs is inconsistent with factor price equalization; moreover, as emphasized by the recent economic geography literature (e.g. Davis and Weinstein 1996), CRS and IRS models might behave differently in the presence of transport cost and differences in demand across countries.

Thirdly, we have already noted that part of the product specialization which we attribute to factor proportions differences or IRS might actually be due to technology differences across countries. A fourth point is that we estimate two-country models with data from a multi-country world. The problem of developing multi-lateral trade predictions arises from the well-known indeterminacy of bilateral trade flows when there are imperfectly specialized homogeneous goods produced with CRS. Traditionally, researchers have instead considered the H-O-Vanek equation which predicts net factor service trade of a country with the rest of the world. However, as Trefler (1996) has shown, generally, the H-O-Vanek equation holds only in cases when also the simplest gravity equation with perfect product specialization holds. It is clear, though, and our results confirm that imperfect product specialization is empirically important. Therefore, we hope to make further progress by extending this work in the context of the gravity equation to a multilateral world with trade costs.

References

[1] Armington, P. S. (1969), “A theory of demand for products distinguished by place of production”, *IMF Staff Papers* 16: 159-177.

[2] Anderson, J. (1979) “A Theoretical Foundation for the Gravity Equation,” *American Economic Review*, 69(1): 106-116.

[3] Antweiler, W., and D. Trefler (1997), “Increasing Returns and All That: A View from Trade”, mimeo, University of Toronto, August.

[4] Bergstrand, J. (1990), “The Heckscher-Ohlin-Samuelson Model, the Linder Hypothesis and the Determinants of Bilateral Intra-Industry Trade”, *Economic Journal* 100: 1216-1229.

[5] Chipman, J. S. (1992), “Intra-Industry Trade, Factor Proportions, and Aggregation”, in *Economic Theory and International Trade. Essays in memoriam J. Trout Rader*, W. Neuefeind and R. Riezman (eds.), pp.67-92. Berlin: Springer.

[6] Davis, D. (1997), ”Critical Evidence on Comparative Advantage? North-North Trade in a Multilateral World”, *Journal of Political Economy* 105: 1051-1060.

[7] Davis, D., (1995), “Intra-industry Trade: A Ricardo-Heckscher Ohlin Approach”, *Journal of International Economics*, 39: 201-226.

[8] Davis, D., and D. Weinstein (1997), “Increasing Returns and International Trade: An Empirical Confirmation”, mimeo, Harvard University, October.

[9] Davis, D., and D. Weinstein (1996), “Does Economic Geography Matter for International Specialization?”, NBER Working Paper # 5706, Cambridge, MA.

[10] Davis, D., D. Weinstein, S. Bradford, and K. Shimpo (1997), "Using International and Japanese Regional Data to Determine When the Factor Abundance Theory of Trade Works", *American Economic Review* 87: 421-446.

[11] Deardorff, A. (1998) "Determinants of Bilateral Trade: Does Gravity Work in a Neoclassical World?", chapter 1 in J. A. Frankel (ed.),*The Regionalization of the World Economy*, Chicago: The University of Chicago Press.

[12] Dixit, A., and J. E. Stiglitz (1977), "Monopolistic Competition and Optimum Product Diversity", *American Economic Review*, 67: 297-308.

[13] Eaton, J., and S. Kortum (1997), "Technology and Bilateral Trade", mimeo, Boston University, April.

[14] Efron, B. (1982), *The Jackknife, the Bootstrap and Other Resampling Plans*, Philadelphia: Society for Industrial and Applied Mathematics.

[15] Feenstra, R., R. E. Lipsey, and H. P. Bowen (1997), "World Trade Flows, 1970-1992, with Production and Tariff Data", NBER Working Paper # 5910 and accompanying CD-ROM, Cambridge, MA.

[16] Finger, J. (1975), "A New View of the Product Cycle Theory", *Weltwirtschaftliches Archiv* : 79-99.

[17] Fontagne, L., M. Freudenberg, and N. Peridy (1997), "Intra-Industry Trade and the Single Market: An Empirical Assessment Based on Trade Types", mimeo, Universite de Paris I Pantheon-Sorbonne.

[18] Greenaway, D., R. Hine, and C. Milner (1994), "Country-Specific Factors and the Pattern of Horizontal and Vertical Intra-Industry Trade in the UK", *Weltwirtschaftliches Archiv* 130, Vol. 1.

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

[20] Härdle, W. (1990), *Applied Nonparametric Regression*, Econometric Society Monograph, Cambridge, U.K.: Cambridge University Press.

[21] Harrigan, J. (1996), "Openness to trade in manufactures in the OECD", *Journal of International Economics* 40: 23-39.

[22] Haveman, J., and D. Hummels (1997), "What can we learn from bilateral trade? Gravity and beyond", mimeo, University of Chicago Business School, June.

[23] Helpman, E. (1981), "International Trade in the Presence of Product Differentiation, Economies of Scale, and Monopolistic Competition: A Chamberlin-Heckscher-Ohlin Approach", *Journal of International Economics* 11: 305-340.

[24] Helpman, E. (1987) "Imperfect Competition and International Trade: Evidence from Fourteen Industrial Countries," *Journal of Japanese and International Economics*, 1: 62-81.

[25] Helpman, E. and P. Krugman (1985) *Market Structure and Foreign Trade: Increasing Returns, Imperfect Competition and the International Economy*, MIT Press, Cambridge, MA.

[26] Hummels, D. and J. Levinsohn (1995) "Monopolistic Competition and International Trade: Reconsidering The Evidence," *Quarterly Journal of Economics*, 110(3): 799-836.

[27] Keller, W. (1998), "Gravity Models of Trade without Specialization of Production: A Note", mimeo, University of Wisconsin, May.

[28] Krugman, P. (1994), "Empirical Evidence on the New Trade Theories: The Current State of Play", in *New Trade Theories: A Look at the Empirical Evidence,* CEPR Conference Report, pp. 11-31.

[29] Leamer, E. (1992) "The Interplay of Theory and Data in the Study of International Trade", chapter 9 of M. Nerlove (ed.) *Issues in Contemporary Economics*, Vol. 2, New York University Press, New York, NY.

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

[31] Markusen, J. (1986), "Explaining the Volume of Trade: An Eclectic Approach", *American Economic Review* 76: 1002-11.

[32] Markusen, J., and R. Wigle (1990), "Explaining the Volume of North-South Trade", *Economic Journal* 100: 1206-1215.

[33] Saxonhouse, G. (1989), "Differentiated Products, Economies of Scale, and Access to the Japanese Market", chapter 5 in R. C. Feenstra (ed.), *Trade Policies for International Competitiveness*, Chicago: University of Chicago Press.

[34] Summers, R., and A. Heston (1991), "The Penn World Table (Mark 5): An Expanded Set of International Comparisons, 1950-1988", *Quarterly Journal of Economics* 106: 327-368.

[35] Trefler, D. (1996), "The Structure of Factor Content Predictions", mimeo, University of Toronto, March 1996.

[36] Trefler, D. (1995), "The Case of the Missing Trade and other Mysteries", *American Economic Review* 85: 1029-1046.

A. The Pure IRS Model

(See Helpman and Krugman 1985). Consider two countries, i and j, two goods, X and Z, and two factors, K (capital) and L (labor), the latter being immobile across countries. The goods X and Z come in many differentiated varieties which are identically produced with increasing returns to scale. Assume that countries have identical homothetic preferences, and that consumers value all varieties symmetrically in the Dixit-Stiglitz (1977) CES fashion. In equilibrium, any product variety will only be manufactured by one producer (perfect product specialization). Producers of differentiated products behave monopolistically competitively, and free entry determines the number of varieties (and firms) in equilibrium. Let n_g^c be the number of good f varieties produced in country c, $g = X, Z$, and $c = i, j$, s^c the share of country c in world spending and $x^c(z^c)$ the equilibrium quantity of a type X (type Z) variety. Let Y^c denote a country's GDP, and world GDP is given by $Y^w = Y^i + Y^j$. Finally, choose good Z as numeraire and let p_x denote the relative price of a good X variety.

With balanced trade–so that $s^c = \frac{Y^c}{Y^w}, \forall c$–and zero trade and transport costs, both countries will demand all foreign varieties according to the countries' GDP as a share of World GDP. Given that the varieties produced in country j are available in country i only through imports, country i's imports from j, denoted M^{ij}, will be

$$M^{ij} = s^i \left[p_x n_x^j x^j + n_z^j z^j \right],$$

whereas country j's imports from i, M^{ji}, will be

$$M^{ji} = s^j \left[p_x n_x^i x^i + n_z^i z^i \right].$$

The last two equations are shown as (2.1) and (2.2) in the text.

B. Proof of Proposition 2

See Keller (1998) for the following result and further discussion. Starting from equation (2.8),

$$VT^{ij} = p_x \left(X^i - s^i X^w \right) + \left(Z^j - s^j Z^w \right),$$

country i's imports from j are $M^{ij} = \left(Z^j - s^j Z^w \right)$. With balanced trade, $M^{ij} = M^{ji} = p_x \left(X^i - s^i X^w \right)$; using the definition of γ leads to

$$M^{ij} = \left(1 - \gamma^i \right) Y^i - s^i \left(1 - \gamma^i \right) Y^i - s^i \left(1 - \gamma^j \right) Y^j.$$

Because $s^j = 1 - s^i$, this gives

$$M^{ij} = s^j \left(1 - \gamma^i \right) Y^i - s^i \left(1 - \gamma^j \right) Y^j = s^j s^i Y^w \left(\gamma^j - \gamma^i \right) = \left(\gamma^j - \gamma^i \right) \frac{Y^i Y^j}{Y^w}.$$

The last term on the right hand side is expression (2.9) given in the text. ■

C. Alternative Prior Beliefs on the Presence of IRS-based Trade in Relation to the Grubel-Lloyd Index

Here we briefly review the sensitivity of our results with respect to the assumption that there is little to no IRS-based trade for Grubel-Lloyd values of $GL^{ij} < 0.05$. First, we compare the multi-cone H-O (M1) to the uni-cone H-O model (M4) for two alternative assumptions: (a) No IRS-based trade for $GL^{ij} < 0.075$, and (b) No IRS-based trade for $GL^{ij} < 0.033$. Here, $\emptyset(M1)$ means, e.g., the average import parameter for model 1.[34]

		FDIF=1	FDIF=2	FDIF=3	FDIF=4	FDIF=5
	$\emptyset$(M1)	0.033	0.021	0.019	0.019	0.017
Case (a)	$\emptyset$(M4)	0.062	0.049	0.038	0.038	0.044
	$\frac{APC(M1)}{APC(M4)}$	1.28	1.2	1.13	1.2	1.21
	$\emptyset$(M1)	0.03	0.022	0.022	0.02	0.02
Case (b)	$\emptyset$(M4)	0.045	0.055	0.042	0.048	0.033
	$\frac{APC(M1)}{APC(M4)}$	1.16	1.08	1.17	1.25	1.33

According to the table, one finds irrespective of the GL cut-off point that the import parameter is non-increasing as FDIF rises in the multi-cone model M1. This parallels the findings in the main text. Furthermore, the ratio of the Amemiya Prediction Criterion (APC) does not vary systematically for either case (a) or case (b). Again this is identical to what we find with a $GL = 0.05$ cut-off, and it is inconsistent with perfect product specialization driven by factor proportions differences being a major element in explaining the success of the gravity equation. The average import parameter for the uni-cone H-O model, however, increases monotonically with FDIF for the $GL = 0.05$ cut-off point discussed in the text, whereas now, the relation between the average parameter estimate and FDIF is non-monotonic. In case (a), for instance, it first falls and then rises with FDIF. Note that case (a) includes country-pairs with a higher GL-value than that of 0.05 as discussed in the text. The finding of high values of $\emptyset(M4)$ for FDIF=1 and FDIF=2 is consistent with country-pairs with relatively high GL values (which have relatively low FDIF values) leading to relatively high import parameter

[34] For these experiments, $R = 10$ and $\sigma_\theta^2 = 2$.

estimates. Therefore, this finding need not contradict the earlier result in the text as much as it underlines the importance of the model identification issue. This point cannot explain the pattern of $\emptyset(M4)$ in case (b), though.

Second, comparing the pure IRS (M2) with the generalized H-O model (M3) for the same cases (a) and (b), we find the following: [35]

		GL=1	GL=2	GL=3	GL=4	GL=5
	$\emptyset$(M2)	0.077	0.081	0.087	0.109	0.131
Case (a)	$\emptyset$(M3)	0.287	0.298	0.304	0.323	0.358
	$\frac{APC(M2)}{APC(M3)}$	1.41	1.28	1.28	1.22	1.24
	$\emptyset$(M2)	0.026	0.031	0.042	0.075	0.145
Case (b)	$\emptyset$(M3)	0.323	0.33	0.332	0.351	0.377
	$\frac{APC(M2)}{APC(M3)}$	2.43	2.0	1.88	1.48	1.24

The table shows that for comparing the pure IRS with the generalized H-O model, we obtain the same results with a cut-off values of $GL = 0.075$ (case (a)) and $GL = 0.033$ (case (b)) as with the value of $GL = 0.05$ as employed in the text: First, the import parameter estimate of the pure IRS model is rising as GL is rising. Secondly, also the average import parameter of the generalized H-O model is rising as GL is rising. Thirdly, the generalized H-O model is less preferred for higher values of GL.

Summarizing, five out of six inferences made in the text continue to hold even if different reasonable prior beliefs on the presence of IRS-based trade in relation to the GL index are adopted. With different prior beliefs, the importance of factor proportions differences in determining trade volumes is evidenced

[35] The following results are based on $R = 10$ and $\sigma_\chi^2 = 0.1$.

primarily by the our findings in support of the generalized H-O model M3.

D. The Effect of Distance

D.1. Between-Class Effects

For the between-class effects, we focus on the High-GL sample ($GL^{ij} > 0.05$), as the average bilateral distance across $FDIF$-classes varies little. We conduct a two-dimensional resampling analysis analogous to the earlier ones, smoothing both the Grubel-Lloyd and distance variables:

$$GL^{ij} = {}^{*}GL^{ij} + \chi^{ij},$$

$$DIST^{ij} = {}^{*}DIST^{ij} + \pi^{ij}, \forall ij,$$

where $\pi \sim N(0, \sigma_{\pi}^2), \sigma_{\pi}^2 > 0$, and $DIST^{ij}$ denotes the bilateral distance between i and j.[36]

In Figure A1, we have plotted the average estimated $(1 - \gamma^i)$, denoted α_{km}, for $K \times M$ classes of GL ($K = 5$ classes) and $DIST$ ($M = 5$ classes), where the latter denotes the average bilateral distance in a sample. It is clear from Figure A1 that we estimate a lower value of α_{km}, the higher is the average bilateral distance. The correlation between GL and α_{km} remains positive. However, this still does not determine whether there is an independent effect associated with GL, because the average distance for $(GL = 1, DIST = 1)$ is much higher than for $(GL = 5, DIST = 1)$, i.e., GL and $DIST$ co-vary. Therefore, we have taken the 25 values of α_{km}, GL_{km}, and $DIST_{km}$, and run the following least-squares regression:

$$\alpha_{km} = \beta_0 + \beta_1 GL_{km} + \beta_2 dist_{km} + \mu_{km}, \tag{D.1}$$

where μ_{km} is assumed to be mean zero and constant variance σ_{μ}^2, and $dist$ denotes the log of the

[36] Here, $R = 50$, $\sigma_{\chi}^2 = 0.3$, and $\sigma_{\pi}^2 = 1.5$.

variable $DIST$. We find the following parameters (standard errors) for β_1 and β_2 : 0.174 (0.064), and -0.084 (0.011), respectively. In sum, higher values of the Grubel-Lloyd index are associated with higher import parameter estimates even when differences in the average distance between trade partners across subsamples are controlled for.

Figure A2 shows the analogous picture for the Pure IRS model. The import parameter estimate rises both with the Grubel-Lloyd index rising and the average bilateral distance falling; for maximum values of GL and minimum values of $DIST$ (the combination $GL = 5$, $DIST = 1$), a value of $\alpha_{51} = 0.26$ is estimated, whereas for the minimum values of GL and maximum values of $DIST$, we estimate $\alpha_{15} = 0.03$. Because the resampling experiments are identically designed, the same co-variance of distance and the Grubel-Lloyd index is present as above for the Generalized H-O model analysis. We run again the least squares regression (D.1), this time with the data generated by the pure IRS model underlying Figure A2. The estimated coefficients (standard errors) are: $\beta_1 = 0.08$ (0.02) and $\beta_2 = -0.02$ (0.003). Also here, the estimated import parameter is higher as the Grubel-Lloyd index rises, even when the effects of distance are controlled for. This suggests that our earlier interpretation of the results is robust.

D.2. Within-Class Effects

To see whether the omission of the distance variable is important in obtaining our results, we use the same resampling approach as in the text, and then run, instead of (4.2) or the analogous regressions, the following gravity regression:

$$M^{ij} = \alpha \left(\frac{Y^i \times Y^j}{Y^w \times dist^{ij}} \right) + \epsilon^{ij}, \tag{D.2}$$

where α is the import parameter of the respective model (a constant for the multi-cone H-O and the pure IRS models, $(1-\gamma^i)$ for the generalized H-O model if country i is capital-abundant, and $(\gamma^j - \gamma^i)$

for the uni-cone H-O model if i is capital-abundant relative to country j). Equation (D.2) corresponds in levels to the gravity equation which is often run in logs, with a common-parameter restrictions imposed. The results of (D.2) are given in Table A1:

Table A1: The Effect of Including a Distance Variable

Low-GL Sample			High-GL Sample		
	Multi-cone H-O	Uni-cone H-O		Pure IRS	Gener. H-O
$FDIF = 1$	0.281	0.302	$GL = 1$	0.324	3.386
	(0.014)	(0.292)		(0.010)	(0.873)
$FDIF = 2$	0.239	0.418	$GL = 2$	0.403	3.502
	(0.016)	(0.318)		(0.012)	(1.075)
$FDIF = 3$	0.206	0.427	$GL = 3$	0.544	3.601
	(0.009)	(0.194)		(0.015)	(1.489)
$FDIF = 4$	0.202	0.456	$GL = 4$	0.655	3.782
	(0.009)	(0.196)		(0.017)	(2.127)
$FDIF = 5$	0.199	0.670	$GL = 5$	1.004	5.057
	(0.010)	(0.173)		(0.023)	(3.584)

The entries give average import parameter estimates for the four models (in brackets are the estimated standard errors). It is clear that the estimated parameters average are, with the inclusion of the distance variable, different from those in the text. The *patterns* of how the estimates change across classes, however, are the same in both cases: the estimate of α is non-increasing for the multi-cone H-O model as FDIF increases, whereas the average α rises with class index for the other three models.

Building on the same resampling approach, we have also estimated an equation where the distance variable enters additively,

$$M^{ij} = \alpha \left(\frac{Y^i \times Y^j}{Y^w} \right) + \beta DIST^{ij} + \epsilon^{ij}, \tag{D.3}$$

as well as an equation which includes an overall constant:

$$M^{ij} = \alpha \left(\frac{Y^i \times Y^j}{Y^w} \right) + \beta DIST^{ij} + \beta_0 + \epsilon^{ij}. \tag{D.4}$$

For both equation (D.3) and equation (D.4), we obtain the same patterns for the average import parameter as given in Table A1 and in the main text. Overall, this suggests that the findings are robust to including the distance variable into the analysis as an empirical specification.