# 4

# The Gravity Model of Bilateral Trade

Our plan is to examine data on bilateral trade between pairs of countries in order to sort out the influence of geographical proximity versus preferential trading policies in creating regional concentration in trade. The natural framework with which to attack this question is the gravity model of bilateral trade.

The gravity model has long been something of an ugly duckling of international economics: obscure and allegedly lacking respectable theoretical foundations. It has recently enjoyed a swan-like revival, however. There are at least three reasons for that revival: (1) its empirical success at predicting bilateral trade flows, (2) its improved theoretical foundations arising mostly from modern theories of trade in imperfect substitutes, and (3) a new interest among economists in the subject of geography and trade, which seeks to treat countries or regions as physically placed at particular locations rather than as disembodied constructs.

It is not easy to decide whom to anoint as the inventor of the gravity model. The concept is so ''natural'' that it seems always to have been used to describe economic links between pairs of geographical units, either with or without the word ''gravity.'' Perhaps the most classic and extensive early application of the model to international trade was by Linnemann (1966), who continued work first reported in Tinbergen (1962), who in turn was contemporaneous with Poyhonen (1963). Specialists in other fields, however, had used versions of the gravity model before international economists did. Regional economists and urban sociologists used it by name as far back as 1946 (Isard 1990; Stewart 1948; Zipf 1946). No doubt there are other early references. It seems safest to cite Isaac Newton as the original progenitor of the gravity model!

Newton's gravitational model says that the attraction between two heavenly bodies is proportional to the product of their masses and inversely related to the distance between them. The gravity model of bilateral trade, in its most basic form, says that trade between country *i* and country *j* is proportional to the product of $GDP_i$ and $GDP_j$ and inversely related to the distance between them. Other explanatory variables that are often added are other measures of size—namely, populations (or per capita GDPs) and land areas—and dummy variables representing other measures of geographical or cultural proximity—namely, landlockedness, common borders, common languages, and common membership in regional trading arrangements. We elaborate on these factors below.

## The Technique of OLS Regression

One fits an equation such as the gravity model to the data by means of ordinary least-squares (OLS) regression analysis. The wonderful thing about OLS is that the technique holds constant for various factors in order to ascertain the effect of another factor. Those unfamiliar with econometrics must rest assured that the technique is not invalidated when the explanatory variables are correlated with each other. To take an example, there is a strong correlation between proximity, as measured by the distance between a pair of countries, and whether they share a common boundary. Yet OLS regression can estimate the independent effect of each factor, so long as we have correctly specified the additive form of the equation. Trade between France and the United Kingdom will be high due to their proximity, but trade between France and Germany will be further boosted by the effect of their common border in addition to their proximity.

The estimates of the effects will be subject to a margin of error, as always. But the estimates will be the best they can be, given the data, so long as the model is correctly specified. Moreover, the standard errors reported for the coefficient estimates will be the correct ones. Thus we will be able to judge whether the estimates are reliable, or whether, to the contrary, the data set is too small to give us the information we want. (The same point holds with respect to whether the explanatory variables are too highly correlated to give us the information we want—what is known as the problem of high multicollinearity.) Fortunately, there is such a huge amount of information in bilateral trade data that we can in fact obtain relatively reliable estimates of the effects of country size, proximity, common borders, and the other variables in the gravity model.

When we have finished thinking of all the other variables that should be expected to determine bilateral trade, we add dummy variables to represent the bloc effects. The dummy variable is equal to 1 when both countries in a given pair belong to the same regional group and 0 otherwise. The estimated coefficient will then tell us how much of the trade within each region can be attributed to a special regional effect. Again, the decision to form a free trade area (FTA) is correlated with geographical proximity. That is what makes regional trading arrangements ''regional.''[1] Despite this correlation, the regression analysis can still separate out the independent effects that each has on trade (so long as we have not omitted any correlated factors from the list of explanatory variables).

Answers depend on what questions one asks. If one is aware of the existence of the 15-country European Union and tests for it while also testing for the existence of an operational FTA among a subset of European countries, such as the original EEC-6, one will generally get a different answer than if one tests for the smaller subset alone. This is as it should be. The EEC-6 variable has a different interpretation when the EU variable is simultaneously included than when it appears alone. If the regression technique tells us that some arbitrary subset is an apparently significant trade bloc, and the conclusion is erroneous in that it spuriously reflects the effect of the more comprehensive group, then the error is ours in omitting the proper group dummy, not the regression technique's or the computer's. It will be important to keep this in mind when we interpret the bloc effects in the next chapter.

Where hypotheses of possible interest include groups that do not yet formally operate, such as the Asia Pacific Economic Cooperation forum (APEC) or the continentwide groups, then it is a matter of judgment whether one is interested in the regression results that include these groups. In other words, those readers who do not consider interesting the hypothesis that there is currently an intraregional bias to trade in APEC, the Free Trade Area of the Americas (FTAA), or East Asia are welcome to skip the results for those cases. They can turn directly to the results for bloc effects among the European Union, the North American Free Trade Agreement (NAFTA), and the other formal FTAs if they wish.

One could argue that dummy variables for *all* possible groups should be tested, so that the data can decide what questions are important: if the true effect of TACBLF (the Trading Area of Countries Beginning with the Letter F) is zero, then the regression technique can be expected to give an estimate of zero for the coefficient on that dummy variable.[2] There is the problem, however, that even a data set of 1,953 country pairs has only so much information to give. If one tires it out by asking a lot

1. Nonregional preferential agreements exist, of course, such as the British Commonwealth system of preferences or the US-Israel Free Trade Agreement. Indeed, Krugman (1991b) has given them a name, ''unnatural trading blocs,'' for reasons that are explained in chapter 8.

2. More technically, the OLS estimate is an *unbiased, efficient, and consistent* estimate of the true parameter, which in this case is zero.

of silly questions, the answers will become increasingly haphazard—including answers to the questions we really care about.[3] In the extreme, if we were to try to estimate more coefficients than there are data points, we would not get an answer at all. The set of all possible groups among 63 countries is in fact very large (2 to the 63rd power, minus 65). Thus we must use some discretion in choosing what groups to test for.

Less silly than fabricating nonexistent groups would be to allow each country to have its own dummy variable or constant term. This would reflect the possibility that some countries are more open than others to all partners, regardless of whether they share membership in a regional trading arrangement. We think that most of the variation in openness may be captured by the effects of per capita incomes (richer countries are more open), dummy variables for Hong Kong and Singapore (which function as entrepots), and a dimension of openness that is shared by other members of the groups we test (e.g., East Asian countries are on average more open than would be predicted). Since we have tried including each of these factors, we omit adding separate dummy variables for openness of each individual country in the central results reported here.[4] Having generally discussed the OLS regression technique, we now turn to examination of the gravity equation itself.

## The Gravity Equation

### Theoretical Foundations

Although aspersions have been cast on the respectability of the gravity model of trade in the past, by now its theoretical pedigree has been proven. Earlier work surveyed by Deardorff (1984, 503-06) provided a partial foundation for the approach. Leamer and Stern (1970), for example,

3. Technically, estimating many needless parameters uses up ''degrees of freedom.'' Then the parameter estimates, even if unbiased in small samples and accurate in large samples, will be needlessly wide of the mark in *our* sample.

4. We have tried also giving each country its own intercept, and we report some of the results below. But the results appeared unreliable when testing country effects and bloc effects at the same time. Thus our standard equation does not allow each country to have its own intercept.

At least one study even allows the coefficients on distance and other variables to vary from country to country (Dhar and Panagariya 1995). If one believes that distance has a bigger effect (e.g., shipping costs are higher per kilometer, even after holding constant for per capita income and the other variables) for some countries' trade than others', then this is the right thing to do. If one is unpersuaded of the importance of allowing for such variation, however, then one might suspect that the resultant loss of degrees of freedom for any surprising results that may come out for estimates of some of the parameters about which we care most.

noted that bilateral trade is indeterminate in the absence of transport costs (i.e., there is nothing to determine whether Japan imports apparel from China or from Morocco) and so assumed that countries essentially draw their trading partners out of a hat, according to various probabilities. More formal approaches relied on product differentiation. The assumption that products are differentiated symmetrically by country of origin has become associated with Armington (1969). Anderson (1979) adopted a linear expenditure system in which the preferences for a country's goods are assumed to be homothetic and uniform across importing countries. Bergstrand (1985) assumed a more flexible utility function that allowed him to find evidence that imports were closer substitutes for each other than for domestic goods. He called his equation a generalized gravity model because it also included price terms.

The best-known theoretical rationale for the idea that bilateral trade depends on the product of GDPs comes from work by Helpman (1987) and Helpman and Krugman (1985, section 1.5). Their approach is the one we build on when we turn to a theoretical analysis of the effects of preferential trading arrangements on the volume of trade and on economic welfare (chapters 7 and 8). In this theory, consumers seek variety in the products they consume, products are differentiated by firm, not just by country, and firms are monopolistically competitive. The authors argued that the classical Heckscher-Ohlin theory of comparative advantage does not have the property that bilateral trade depends on the product of incomes, as it does in the gravity model. Deardorff (1984, 500-04) concurred. Since the data *do* have the property that bilateral trade depends on the product of incomes, as we shall see, this seemed to be a point in favor of the differentiated products model of trade and a vote against the classical models.

More recently, Deardorff (1997) has changed his mind, having discovered how to derive the gravity model from Heckscher-Ohlin theory almost as easily as from the theory of imperfect substitutes. His main purpose is to show that the empirical success of the gravity equation does not necessarily support the imperfect-substitutes model relative to the Heckscher-Ohlin model. For our purposes, the main point is that it seems possible to derive the gravity model from a variety of leading theories. The equation has thus apparently gone from an embarrassing poverty of theoretical foundations to an embarrassment of riches!

To most readers who have not studied enough trade theory to have lost sight of the obvious, the assumption that trade between countries depends positively on their size and inversely on distance may seem self-evident. Those trade theorists who previously questioned the foundations of the gravity model did not have an alternative model of bilateral trade to offer. It was just that economists had not tried very hard to model bilateral trade. Deardorff (1997) concludes:

> All that the gravity equation says, after all, aside from its particular functional form, is that bilateral trade should be positively related to the two countries' incomes and negatively related to the distance between them. Transport costs would surely yield the latter in just about any sensible model. And the dependence on incomes would also be hard to avoid.

While the derivation of a proportionate relationship between trade flows and country size is an important foundation, the theories of Helpman (1987) and most of the other authors cited do not include a role for distance and thus cannot properly be called foundations of the full gravity model. The few exceptions include Bergstrand's (1985) version of the imperfect-substitutes theory, which incorporated a role for shipping costs, proxied in practice by distance. Distance is also included in the second of the two Heckscher-Ohlin-based models developed by Deardorff (1997). The proportionality between bilateral trade and the product of incomes, as well as the inverse dependence on distance, are also properties of our theoretical model, introduced in chapter 7. We assume that transportation costs raise the price of a good in the importing country and that distance has a positive effect on transportation costs.

Once one has incorporated a role for distance in raising the cost of trade, it is a small step to think of similar roles for dummy variables indicating whether the pair of countries shares a common border or common language. Each of these links helps reduce the cost of doing business abroad, just as proximity does. Near the border, consumers can cross over to shop in the other country and firms can source intermediate inputs in the other country, much more readily than would be possible if the countries did not share a common border. Linguistic links and other historical and cultural links are particularly important at reducing what in the preceding chapter we called the cost of unfamiliarity in international trade, what Linnemann (1966) called psychic costs, and Garnaut (1994) subjective resistance.

The variable that calls for more serious modeling than anyone has yet done is per capita income. These three variables—common border, common language, and per capita income—are discussed below, as we go through the estimation results. We speak of these three variables as constituting part of a standard ''full'' gravity model, while income and distance constitute the ''basic'' gravity model. In recent years, there have been many other extensions of the model, some pursued as part of our research, some by other authors. These extensions are also discussed below.

## Our Data Sample

The dependent variable in most of our tests is total merchandise trade (exports plus imports), in log form, between pairs of countries in a given year. We have also tried distinguishing between imports and exports, as do many studies; these results are described as one of the later extensions in chapter 6.

We have run tests at five-year intervals and more frequently at the end of the sample. Thus we have estimates for 1965, 1970, 1975, 1980, 1985, 1987, 1990, 1992, and some for 1994. There are enough data in the cross-section dimension that we can estimate each year separately. Our data source is the UN trade matrix, supplemented in the later years by the International Monetary Fund's *Direction of Trade Statistics*. The UN data cover 63 countries in our sample so that there are 1,953 data points (63 × 62/2) for a given year. This trade constituted 88.7 percent of world trade reported to the United Nations in 1992. Expanding the set to a larger number of countries would have allowed testing for more and smaller regional arrangements, but it would have cut short the time span for which sufficient data are available to allow useful estimation. The 63 countries are listed in table 4.1 by geographic group.

Many of the early gravity tests concentrated on data for trade among industrialized countries. There are three possible reasons for this. First, data were more available for these countries. Second, where the motivation was to learn about the effect of regional trading arrangements, it was usually the European Community and the European Free Trade Association (EFTA) that investigators had in mind. Third, when the modern theories of trade in imperfect substitutes were introduced as the justification for the model, they were thought to apply only to the industrialized countries, not to the developing countries. Theories based on abundant endowments of unskilled labor or natural resources seemed more applicable to the developing countries.

Each of these reasons for limiting the analysis to industrialized countries is no longer convincing, even if they once were. First, enough data exist to include many developing countries. Second, regional arrangements in Latin America and elsewhere in the developing world certainly deserve serious attention now. Third, even if the goal is only to assess FTAs in Europe, one should use as broad a set of data as possible to estimate normal patterns of trade so that one can ascertain how European links differ from what is normal. Fourth, some of the countries that have previously been omitted from the list of industrialized countries (Singapore and Hong Kong) are now richer than most of the ones that are included. Fifth, we believe that the gravity model is in fact applicable to developing countries, not just to industrialized countries.

Hummels and Levinsohn (1995) extended the test of Helpman (1987) to a data set including developing countries, taking it for granted that the monopolistic competition model would only do a good job of explaining trade among OECD countries. They were surprised to find that it works equally well for the larger set of countries. Product varieties from

**Table 4.1 Countries included in the gravity equation by regional group and main city[a]**

| Americas (WH.13) | European Community (EC.11) | European Free Trade Area (EFTA.6) | East Asia (EAEG.10)[b] | Other countries (23) |
|---|---|---|---|---|
| Canada, Ottawa | West Germany, Bonn | Austria, Vienna | Japan, Tokyo | South Africa, Pretoria |
| United States, Chicago | France, Paris | Finland, Helsinki | Indonesia, Jakarta | Turkey, Ankara |
| Argentina, Buenos Aires | Italy, Rome | Norway, Oslo | Taiwan, Taipei | Yugoslavia, Belgrade |
| Brazil, Sao Paulo | United Kingdom, London | Sweden, Stockholm | Hong Kong, Hong Kong | Israel, Jerusalem |
| Chile, Santiago | Belgium, Brussels | Switzerland, Geneva | South Korea, Seoul | Algeria, Algiers |
| Colombia, Bogota | Denmark, Copenhagen | Iceland, Reykjavik | Malaysia, Kuala Lumpur | Libya, Tripoli |
| Ecuador, Quito | Netherlands, Amsterdam | | Philippines, Manila | Nigeria, Lagos |
| Mexico, Mexico City | Greece, Athens | | Singapore, Singapore | Egypt, Cairo |
| Peru, Lima | Ireland, Dublin | | Thailand, Bangkok | Morocco, Casablanca |
| Venezuela, Caracas | Portugal, Lisbon | | China, Shanghai | Tunisia, Tunis |
| Bolivia, La Paz | Spain, Madrid | | | Sudan, Khartoum |
| Paraguay, Asuncion | | | | Ghana, Accra |
| Uruguay, Montevideo | | | | Kenya, Nairobi |
| | | | | Ethiopia, Addis Adaba |
| | | | | Iran, Tehran |
| | | | | Kuwait, Kuwait |
| | | | | Saudi Arabia, Riyadh |
| | | | | India, New Delhi |
| | | | | Pakistan, Karachi |
| | | | | Hungary, Budapest |
| | | | | Poland, Warsaw |
| | | | | Australia, Sydney |
| | | | | New Zealand, Wellington |

a. The distance between countries was computed as the great circle distance between the relevant pair of cities.

b. APEC consists of East Asia, plus Australia, New Zealand, Canada, and the United States. Latecomers Mexico and Chile are not counted in APEC.

developing countries are often quite imperfect substitutes, for each other as well as for goods from rich countries. This applies both to agricultural products (say, Chilean wine versus Romanian wine versus French wine) or manufactures (Chinese jeans versus Mexican jeans versus Italian jeans).

Furthermore, as we have already noted, Deardorff has derived the gravity model for a version of Heckscher-Ohlin trade, which has always been considered well-suited to North-South trade. This is not to deny that trade patterns for countries at earlier stages of development are different from those of countries at later stages. We shall be taking into account countries' incomes per capita. The point is that developing countries should be included in the data set.

## Size, Income Per Capita, and Competing Models of Trade

There are two standard ways of measuring the size of countries in the gravity model: GNP (output) or population.[5] When holding constant for GNP, the coefficient on population is generally negative. This captures the well-known phenomenon that larger countries tend to be relatively less open to trade as a percentage of GNP. A Singapore or a Luxembourg is highly dependent on trade, in part because it lacks many natural endowments and because it lacks room to exploit economies of scale in the domestic market. A Japan or United States, while engaging in far more trade in absolute terms, will engage in less trade as a percentage of GNP because it can find more within its own borders. There is an additional reason for this pattern. Interstate trade in the United States is considered domestic; interstate trade within the European Union is considered international trade. The accidents of political history alone will give the result that the ratio of trade to income falls with the size of the unit.

In some studies these size variables are supplemented by a measure of land area. This is generally a way of getting at natural resources. A country with a large land area, holding constant for the other measures of size, is relatively more self-sufficient and less dependent on trade. In one gravity equation, we find that, for every 1 percent increase in land area, trade falls by about 0.2 percent.[6] But we shall focus here on population as the measure of size and self-sufficiency.

Mathematically, it is precisely equivalent, whether we express the explanatory variables as GNP and per capita GNP, or as GNP and population. To see this, consider the basic gravity equation, with some typical parameter values substituted in for concreteness:

---

5. Our standard tests use GNP rather than GDP (gross domestic product) because of greater data availability. We have also tried the latter, however. The difference does not seem important in this context (Linnemann 1966).

6. The estimate is highly significant statistically (Frankel and Romer 1996, table 1).

$$\log T_{ij} = .7 \log(GNP_i GNP_j) + .3 \log[(GNP/pop_i)(GNP/pop_j)] - .7 \log(Dist._{ij}), \quad (4.1)$$

where $T_{ij}$ is trade between country $i$ and country $j$. This equation is precisely the same as an equation that expresses the income and population terms as explanatory variables:

$$\log T_{ij} = 1.0 \log(GNP_i GNP_j) - .3 \log(pop_i pop_j) - .7 \log(Dist._{ij}) \quad (4.2)$$

In equation (4.1), the tendency of trade to rise less than proportionately with size is reflected in a GNP coefficient less than one; in equation (4.2), it is reflected in a population coefficient less than zero. We will choose to use the first formulation. Although the estimation is precisely equivalent either way, mathematically, the reader of the results may be led to different interpretations.

Observing equation (4.1), one is usefully led to think about how a country's trade depends on its stage of development. Why should trade depend on development? We do not get much guidance from the standard gravity model foundation—theories of trade based on imperfect substitutes. In models of the Krugman-Helpman type, there is no role for GNP per capita. The models are simply expressed in terms of economic size, without distinguishing between the roles of output and population. It is clear empirically that the two effects are in fact independent. So this is an area that bears further research.

One possible explanation for the independent effect of income per capita is that exotic foreign varieties are superior goods in consumption. Low-income countries are dominated by subsistence farming. Other possibilities come out of the literature on endogenous growth.[7] For example, the process of development may be led by the innovation or invention of new products that are then demanded as exports by other countries. It has also been suggested that the more developed countries have more advanced transportation infrastructures, including seaports and airports, which facilitate trade.

Perhaps the most important reason industrialized countries trade more than less developed countries is that countries tend to liberalize as they develop. One reason for this pattern is that governments of very poor countries depend on tariff revenue for a large share of their budget, while more advanced countries can apply other forms of direct and indirect taxes to the domestic economy. If we had good direct measures of trade policy that we could use in the equation, such as a country's overall level of tariffs, then we would not expect any of this effect to show up in the

7. Of the many relevant works, some of the more important are Grossman and Helpman (1989, 1991a) and Rivera-Batiz and Romer (1991). For further references on the connections between trade and growth, see Frankel, Romer, and Cyrus (1995).

term for per capita income. But given the absence of good measures of this kind and the importance of nontariff barriers, the most important reason to expect low per capita GNP to inhibit trade may be the correlation with protectionist trade policies.

To take an example, China and Japan have roughly the same aggregate outputs. Yet China trades less with its partners than does Japan. This is what the equation would lead us to expect, because China's large output derives primarily from its large population, while Japan's derives from its high level of *GNP per capita*.

Perhaps the easiest way of keeping in mind all the factors that may be involved in this relationship is to think in terms of a third way of expressing the same equation:

$$\log T_{ij} = 1.0 \log[(GNP/pop_i)(GNP/pop_j)] + .7 \log(pop)_i(pop)_j - .7 \log(Dist._{ij}) \quad (4.3)$$

Now, to see the effect of growth on trade, we first ask what is the source of growth. If an increase in GNP (relative to other countries) takes the form of an increase in GNP per capita, then the effect on trade is proportionate, as called for in the trade theories based on imperfect substitutes. If an increase in GNP instead is entirely accounted for by an increase in population, then the effect on trade is somewhat less than proportionate (0.7 instead of 1.0), as economies of scale make the country proportionately less dependent on trade. The rapid growth in incomes of most East Asian countries in recent years has been based primarily on productivity growth rather than population growth—call it 8 percent a year. Then the equation predicts that their trade with slow-growing countries will increase by 8 percent a year and their trade with each other will increase by 16 percent a year (8 + 8).[8]

It is also instructive to focus explicitly on the product of per capita GNPs as a determinant of trade, as in equation (4.1). The prediction that the *product* of per capita incomes enters the equation positively contradicts the prediction of traditional Heckscher-Ohlin theories of trade. If the two factors of production are capital and labor, then these theories predict that countries with dissimilar levels of output per capita will trade more than countries with similar levels. (Or, more precisely, dissimilar capital/labor ratios.) But the standard gravity model predicts that *countries with similar levels of output per capita will trade more than countries with dissimilar*

8. Assume that the country in question is small and gross world product can be held constant. Then we can jump to the time-series context, even though these estimates are from a pure cross-section. Otherwise, we need to express growth rates as *relative to the worldwide average*: If East Asian countries grow 6 percent per year faster than the world average, then their trade with other countries also grows 6 percent faster, and their trade with each other grows 12 percent faster.

*levels*.[9] Seldom do competing theories have such directly contradictory empirical implications.

Before there was Krugman (and others), there was Linder. The Linder (1961) hypothesis says that countries with similar levels of per capita income will have similar preferences and similar but differentiated products, and thus will trade more with each other. It is often viewed as similar to the Krugman-Helpman hypothesis in its empirical predictions, if somewhat different in derivation. There is in fact a crucial difference in the empirical implications, however. The Krugman-Helpman theory predicts that the sum of the logs of $(GNP/pop_i)$ and $(GNP/pop_j)$ will have a positive effect on the log of trade. The Linder hypothesis is usually described as predicting that the *absolute value of the difference* of the two variables will have a negative effect. (There is no theory to predict whether they enter in log form or simply as ratios.) Heckscher-Ohlin is then seen as the third case: The absolute value of the difference of the two variables will have a *positive* effect. The diametric opposite of Heckscher-Ohlin is thus the Linder theory, not the Helpman-Krugman theory. The Linder hypothesis is properly viewed as a distinct third school of thought, not just a forerunner of Helpman-Krugman.

Deardorff (1997, 15) has argued that it can be viewed as having a certain kinship to Heckscher-Ohlin. He opines—as Markusen (1986) has already shown—that if high-income consumers tend to consume larger budget shares of capital-intensive goods, which Heckscher-Ohlin tells us are produced by capital-rich countries, then it follows that (1) capital-rich countries will trade more with other capital-rich countries than with capital-poor countries, and (2) capital-poor countries likewise will trade more with their own kind. These predictions are the same as those of the Linder hypothesis, derived in a different way. The first of them is borne out by the common finding in gravity equations that the product of per capita GNPs has a positive effect on trade. However, the second prediction is contradicted by the common gravity equation finding, which says instead that a poor country will trade more with a rich country than with another poor country. In other words, trade stems from economic development, not from *similarity* of the stage of development.

To distinguish among these influences—Heckscher-Ohlin-style factor-endowment differences, Linder-style taste differences, and the effect of

9. The Helpman-Krugman sort of theory and its child, the standard gravity equation, also predict that if the distribution of national incomes across countries becomes more equal over time, the volume of trade should increase. The United States declined from almost half of gross world product among market economies in the 1950s to 32 percent in 1970 and then to 26 percent in 1993. Thus a more uniform size distribution among economies is indeed one explanation for the increase in global trade (Helpman 1987; Krugman 1995, 341). As noted, Hummels and Levinsohn (1995) find that the same pattern that Helpman found for OECD countries holds for less developed countries.

development on trade—one must try to capture the distinctive features of each. One of the experiments we try below, within the standard gravity model formulation of bilateral trade, is adding a term for the difference in per capita GNPs. A negative sign on this term would support the Linder hypothesis, while a positive sign would support the Heckscher-Ohlin hypothesis.

One must note that our specification of these other terms remains more ad hoc than the basic gravity specification that we attribute to Helpman-Krugman, Bergstrand (1985), and our own chapters 7 and 8. This is in part for the reason that, despite the progress made by Deardorff (1997) and the earlier authors in deriving theories of bilateral trade, geography has seldom been introduced into the classical models.

Gruber and Vernon (1970) and Thursby and Thursby (1987) have already added absolute differences in per capita incomes to the gravity equation in an attempt to capture the Linder effect. Bergstrand (1989) generalized his gravity model foundations further to include both a role for factor-endowment differences in the spirit of Heckscher-Ohlin and a role for taste variables in the spirit of Linder. The resulting equation, however, uses only per capita income variables. Specific commodities are revealed to be capital-intensive in production if the per capita income of the exporting country is estimated to have a positive coefficient and revealed to be labor-intensive if it has a negative coefficient.[10] Commodities are revealed to be luxuries in consumption if per capita income of the importing country is estimated to have a positive coefficient and necessities if it has a negative coefficient.

One can also try to measure factor endowments directly to get at the Heckscher-Ohlin model more appropriately. Leamer (1974) added factor-endowment variables to a gravity-type equation and found that they performed less well than the standard income and population variables. We have also tried adding terms for differences in factor endowments as a more direct test of the Heckscher-Ohlin hypothesis. The results are described in chapter 6.

## Our Results for the Basic Gravity Variables

The equation to be estimated, in its basic form, is:

$$\log(T_{ij}) = \alpha + \beta_1 \log(GNP_i GNP_j) + \beta_2 \log(GNP/pop_i GNP/pop_j) + \beta_3 \log(Dist_{ij}) + \beta_4 (ADJ_{ij}) + \beta_5 (Lang_{ij}) + \gamma_1 (WE_{ij}) + \gamma_2 (WH_{ij}) + \gamma_3 (EA_{ij}) + u_{ij}. \quad (4.4)$$

10. More precisely, the technical difficulties in generalizing two-commodity two-factor Heckscher-Ohlin theory to a world with more than two commodities require that we say that specific commodities *tend* in an average sense to be capital-intensive or labor-intensive (Bergstrand 1989; Deardorff 1984).

## Table 4.2 Gravity model estimations of explicit regional trading arrangements,[a] 1965–92 (with trade between country pairs as dependent variable)

| | 1965 | | 1970 | | 1975 | | 1980 | | 1985 | | 1990 | | 1992 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Intercept | −7.910** | −9.632** | −9.157** | −10.763** | −9.326** | −10.820** | −12.006** | −13.564** | −10.956** | −12.146** | −9.599** | −10.523** | −12.146** | −13.521** |
| | (0.532) | (0.619) | (0.591) | (0.664) | (0.544) | (0.619) | (0.530) | (0.635) | (0.492) | (0.576) | (0.464) | (0.509) | (0.469) | (0.530) |
| GNP | 0.637** | 0.685** | 0.646** | 0.702** | 0.744* | 0.786** | 0.775** | 0.804** | 0.797** | 0.834** | 0.796** | 0.832** | 0.930** | 0.963** |
| | (0.018) | (0.019) | (0.019) | (0.021) | (0.018) | (0.019) | (0.016) | (0.017) | (0.016) | (0.017) | (0.016) | (0.016) | (0.018) | (0.018) |
| Per capita GNP | 0.235** | 0.284** | 0.337** | 0.403** | 0.255** | 0.294** | 0.283** | 0.323** | 0.247** | 0.264** | 0.080** | 0.128** | 0.128** | 0.153** |
| | (0.026) | (0.028) | (0.026) | (0.028) | (0.023) | (0.025) | (0.022) | (0.025) | (0.022) | (0.024) | (0.017) | (0.018) | (0.019) | (0.020) |
| Distance | −0.483** | −0.447** | −0.562** | −0.594** | −0.698** | −0.683** | −0.588** | −0.555** | −0.715** | −0.707** | −0.572** | −0.656** | −0.770** | −0.733** |
| | (0.044) | (0.052) | (0.042) | (0.049) | (0.042) | (0.048) | (0.039) | (0.048) | (0.039) | (0.047) | (0.037) | (0.043) | (0.038) | (0.044) |
| Adjacency | 0.433** | 0.482** | 0.458** | 0.394* | −0.398* | 0.400* | 0.571** | 0.602** | 0.658** | 0.626** | 0.751** | 0.609** | 0.445** | 0.506** |
| | (0.161) | (0.162) | (0.165) | (0.170) | (0.160) | (0.166) | (0.174) | (0.182) | (0.165) | (0.171) | (0.189) | (0.189) | (0.157) | (0.170) |
| Language | 0.550** | 0.586** | 0.348** | 0.410** | 0.368** | 0.446** | 0.675** | 0.754** | 0.474** | 0.571** | 0.572** | 0.635** | 0.768** | 0.823** |
| | (0.095) | (0.096) | (0.094) | (0.096) | (0.094) | (0.099) | (0.093) | (0.098) | (0.093) | (0.097) | (0.090) | (0.088) | (0.090) | (0.090) |
| EU15 bloc | 0.218# | 0.143 | 0.061 | −0.078 | −0.140 | −0.229* | −0.021 | −0.076 | 0.227* | 0.134 | 0.267** | 0.158 | −0.083 | −0.135 |
| | (0.116) | (0.114) | (0.110) | (0.107) | (0.104) | (0.104) | (0.103) | (0.103) | (0.099) | (0.100) | (0.102) | (0.103) | (0.097) | (0.099) |
| NAFTA bloc | 0.020 | 0.178 | −0.227 | −0.050 | −0.313 | −0.028 | 0.098 | 0.379 | −0.264 | 0.185 | 0.152 | 0.367 | −0.226 | 0.201 |
| | (0.311) | (0.263) | (0.333) | (0.275) | (0.298) | (0.269) | (0.274) | (0.290) | (0.268) | (0.289) | (0.292) | (0.339) | (0.294) | (0.333) |
| Mercosur bloc | −0.343 | −0.051 | 0.311 | 0.451 | 0.277 | 0.427 | 0.561* | 0.746** | 0.808* | 0.686# | 1.918** | 1.324** | 0.690* | 0.934* |
| | (0.444) | (0.444) | (0.331) | (0.358) | (0.326) | (0.351) | (0.236) | (0.253) | (0.356) | (0.379) | (0.235) | (0.264) | (0.340) | (0.364) |
| Andean bloc | −1.310** | −1.198** | −0.307 | −0.283 | 0.311 | 0.351 | 0.082 | 0.103 | −0.103 | 0.046 | −0.104 | 0.204 | 0.965** | 1.187** |
| | (0.446) | (0.467) | (0.253) | (0.275) | (0.321) | (0.342) | (0.248) | (0.263) | (0.466) | (0.479) | (0.467) | (0.481) | (0.238) | (0.256) |
| ASEAN bloc | 1.621** | 1.274** | 2.045** | 1.570** | 1.824** | 1.512** | 2.272** | 1.925** | 1.704** | 1.487** | 1.757** | 1.196** | 1.766** | 1.126** |
| | (0.487) | (0.503) | (0.379) | (0.393) | (0.315) | (0.324) | (0.393) | (0.403) | (0.370) | (0.378) | (0.335) | (0.316) | (0.281) | (0.286) |
| AUS-NZ bloc | | | | | | | 1.263** | 1.448** | 1.399** | 1.380** | 1.732** | 1.768** | 1.716** | 1.688** |
| | | | | | | | (0.106) | (0.159) | (0.106) | (0.158) | (0.097) | (0.151) | (0.095) | (0.140) |
| EU15 openness | | −0.120 | | −0.159* | | −0.096 | | −0.036 | | 0.077 | | −0.186** | | −0.132* |
| | | (0.074) | | (0.072) | | (0.074) | | (0.070) | | (0.067) | | (0.063) | | (0.064) |

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| NAFTA openness | | −0.600** | | −0.664** | | −0.630** | | −0.491** | | −0.616** | | −0.434** | | −0.751** |
| | | (0.114) | | (0.112) | | (0.126) | | (0.115) | | (0.112) | | (0.094) | | (0.096) |
| Mercosur openness | | −0.289** | | −0.091 | | −0.136 | | −0.132 | | 0.252** | | 0.818** | | −0.295** |
| | | (0.096) | | (0.093) | | (0.097) | | (0.092) | | (0.098) | | (0.101) | | (0.098) |
| Andean openness | | −0.147 | | 0.030 | | −0.032 | | 0.058 | | −0.087 | | −0.106 | | −0.190# |
| | | (0.110) | | (0.092) | | (0.101) | | (0.101) | | (0.101) | | (0.103) | | (0.108) |
| ASEAN openness | | 0.451 | | 0.620** | | 0.392** | | 0.469** | | 0.312** | | 0.640** | | 0.610** |
| | | (0.112) | | (0.110) | | (0.103) | | (0.106) | | (0.105) | | (0.085) | | (0.094) |
| AUS-NZ openness | | | | | | | | −0.331* | | −0.072 | | −0.272* | | −0.154 |
| | | | | | | | | (0.145) | | (0.132) | | (0.136) | | (0.128) |
| Number of observations | 1,194 | 1,194 | 1,274 | 1,274 | 1,453 | 1,453 | 1,708 | 1,708 | 1,647 | 1,647 | 1,573 | 1,573 | 1,546 | 1,546 |
| Adjusted $R^2$ | 0.660 | 0.674 | 0.684 | 0.701 | 0.703 | 0.713 | 0.694 | 0.703 | 0.721 | 0.730 | 0.750 | 0.776 | 0.798 | 0.816 |
| Standard error of regression | 1.096 | 1.072 | 1.126 | 1.094 | 1.200 | 1.180 | 1.242 | 1.223 | 1.204 | 1.185 | 1.115 | 1.057 | 1.135 | 1.085 |

**, *, # denotes significant at the 99%, 95%, and 90% levels, respectively.
a. All variables except dummy variables are in logs.

The last five explanatory factors are dummy variables. $ADJ_{ij}$, short for Adjacency, is equal to 1 when countries *i* and *j* share a common border and 0 when they do not. $Lang_{ij}$ is equal to 1 when the countries share a common language or past colonial links and 0 otherwise. *WE*, *WH*, and *EA* are three examples of the dummy variables we use when testing the effects of membership in a common regional groups, standing in this case for Western Europe, Western Hemisphere, and East Asia.

Table 4.2 reports the results of OLS estimation of the equation, where the trading blocs tested for are six groups that currently have formal status as regional trading arrangements: the European Union, NAFTA, Mercosur, the Andean Community, ASEAN, and the Australia-New Zealand Closer Economic Relations (CER) pact. Table 4.3 reports results when, as in equation (4.4), we test for broader, less formal, blocs that currently exist only as proposals or hypotheses: Western Europe, the Western Hemisphere, East Asia, APEC, and TAFTA. In both cases, the tests are run separately at five-year intervals from 1965 to 1990 and are also run on more recent data available for 1992. In these tables, for comparability across time, we elect to test the effects of country groups defined to have the same membership in every year, notwithstanding that in the 1970s, for example, the European Economic Community contained nine countries, rather than the 15 now in the European Union, or that NAFTA did not exist at all. If we find no significant effects in the early years, we will know the likely reason why.

We find all five standard gravity variables to be highly significant statistically (i.e., significant at greater than the 99 percent confidence level). We discuss them, before turning to the bloc effects in the next chapter.

## Coefficients on Size and Per Capita Income

The log of the product of the two countries' GNPs is always highly significant statistically. It is also generally significantly less than 1. Thus the results indicate that, though trade increases with a country's size, it increases less than proportionately (holding GNP per capita constant). This confirms, as expected, the familiar pattern that small economies tend to be more open to international trade than larger, more diversified, economies. In tables 4.2 and 4.3, the coefficient shows an upward trend, from about .6 in 1965 to more than .9 in 1992. (Earlier research usually did not show a trend like this in the income coefficient.)

The estimated coefficient on GNP per capita is also highly significant statistically, indicating that richer countries do indeed trade more than poor ones. It shows a moderate downward trend, declining from about .3 early in the sample period to about .1 toward the end of the sample period.[11] In most cases, a test would fail to reject the constraint that the sum of the coefficients on GNP and per capita GNP is 1.0. Holding constant for population, trade between two countries is simply proportionate to the product of their GNPs.

The reported results measure GNPs and per capita GNPs at current exchange rates. An alternative is to measure them at purchasing power parity (PPP) rates (Boisso and Ferrantino 1993, 1996).[12] In theory, the PPP rates are probably to be preferred. Otherwise, large temporary swings in the nominal exchange rate can create large swings in the real exchange rate and distort the comparison of incomes. (It is possible that our gravity estimates for 1985, the year when the dollar had reached its peak real appreciation, are distorted.) The disadvantage of using the PPP rates is that they are subject to large measurement errors, as Srinivasan (1995) has pointed out. Most of our reported results thus measure incomes at current exchange rates. We have, however, also tried the Summers-Heston PPP-rate measure. One effect was to restore to a statistically significant positive value the coefficient on per capita incomes in 1991, when the conventional exchange rate measure showed a coefficient of the wrong sign.[13] The results for the other variables were little affected. When exchange-rate-based incomes and PPP-based incomes are entered into the equation at the same time, the data seem to prefer the former, though the multicollinearity is too high for a clear and consistent verdict.

## Distance: Measurement and Connection to the Question of Subnational Provinces

The calculation of the distance variable requires some elaboration. The proximity measure that we use in most of our tests is the log of distance between the two major cities of the respective countries. The cities are usually the capitals of the two countries, but in a few cases we substitute for the capital a major city that seems closer to the country's economic center of gravity (Chicago for the United States rather than Washington, DC, and Shanghai for China rather than Beijing). Table 4.1 lists the cities that we use for each of the 63 countries.[14]

---

11. Linnemann (1966) obtains similar estimates for this parameter (in the range of .21 to .27) for the year 1959, as do others. Eichengreen and Irwin (1995), however, obtain much higher estimates for the interwar period: .59 to .85.

12. Linnemann (1966) tried both and found that it made little difference for the results, except for the coefficient on income itself.

13. The data are not as complete as for earlier years. Appendix table B5.4a shows the results for preliminary 1991 data, with the standard exchange-rate-evaluated GNPs, and appendix table B5.4b shows the same regression with the PPP-evaluated numbers.

14. Boisso and Ferrantino (1996) find very little difference in gravity equation results whether distance is measured between the most populous cities or the geographical centers.

**Table 4.3 Gravity model estimations of prospective trade blocs,[a] 1965–92**
(with trade between country pairs as dependent variable)

| | 1965 | | 1970 | | 1975 | | 1980 | | 1985 | | 1990 | | 1992 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Intercept | −9.135** | −11.671** | −10.040** | −11.473** | −9.521** | −10.538** | −12.261** | −12.377** | −10.780** | −11.167** | −9.558** | −9.841** | −11.776** | −12.043** |
| | (0.524) | (0.771) | (0.544) | (0.777) | (0.562) | (0.840) | (0.527) | (0.753) | (0.496) | (0.758) | (0.461) | (0.652) | (0.472) | (0.643) |
| GNP | 0.643** | 0.671** | 0.647** | 0.661** | 0.730* | 0.733** | 0.752** | 0.752** | 0.771** | 0.774** | 0.767** | 0.770** | 0.901** | 0.883** |
| | (0.018) | (0.018) | (0.018) | (0.019) | (0.018) | (0.019) | (0.016) | (0.017) | (0.016) | (0.017) | (0.016) | (0.016) | (0.018) | (0.018) |
| Per capita GNP | 0.262** | 0.357** | 0.369** | 0.425** | 0.284** | 0.324** | 0.300** | 0.306** | 0.247** | 0.248** | 0.087** | 0.134** | 0.139** | 0.163** |
| | (0.028) | (0.035) | (0.027) | (0.034) | (0.024) | (0.031) | (0.022) | (0.028) | (0.023) | (0.028) | (0.018) | (0.022) | (0.020) | (0.022) |
| Distance | −0.401** | −0.277** | −0.519** | −0.456** | −0.695** | −0.649** | −0.541** | −0.567** | −0.677** | −0.664** | −0.525** | −0.597** | −0.762** | −0.732** |
| | (0.049) | (0.067) | (0.045) | (0.063) | (0.048) | (0.070) | (0.041) | (0.059) | (0.042) | (0.062) | (0.040) | (0.054) | (0.041) | (0.050) |
| Adjacency | 0.568** | 0.660** | 0.621** | 0.640** | 0.497** | 0.525** | 0.668** | 0.634** | 0.742** | 0.740** | 0.781** | 0.662** | 0.561** | 0.574** |
| | (0.160) | (0.163) | (0.156) | (0.161) | (0.154) | (0.159) | (0.158) | (0.162) | (0.148) | (0.154) | (0.164) | (0.166) | (0.144) | (0.143) |
| Language | 0.505** | 0.561** | 0.333** | 0.393** | 0.326** | 0.376** | 0.520** | 0.611** | 0.338** | 0.438** | 0.359** | 0.450** | 0.620** | 0.654** |
| | (0.093) | (0.092) | (0.091) | (0.092) | (0.094) | (0.097) | (0.090) | (0.091) | (0.092) | (0.092) | (0.084) | (0.084) | (0.088) | (0.085) |
| Western Europe bloc | 0.336** | 0.095 | 0.196# | 0.073 | −0.006 | −0.132 | 0.251* | 0.151 | 0.344** | 0.270* | 0.411** | 0.234* | 0.095 | −0.075 |
| | (0.118) | (0.128) | (0.115) | (0.125) | (0.118) | (0.127) | (0.110) | (0.117) | (0.110) | (0.118) | (0.107) | (0.117) | (0.105) | (0.110) |
| Western Hemisphere bloc | −0.321* | 0.004 | −0.199# | 0.048 | 0.067 | 0.357* | 0.338* | 0.687** | 0.241# | 0.522** | 0.821** | 0.943** | 0.362** | 0.824** |
| | (0.138) | (0.160) | (0.116) | (0.134) | (0.136) | (0.156) | (0.133) | (0.151) | (0.141) | (0.150) | (0.157) | (0.173) | (0.116) | (0.132) |
| East Asia bloc | 1.626** | 1.791** | 1.822** | 1.838** | 0.949** | 0.900** | 0.940** | 0.784** | 0.666** | 0.615** | 0.762** | 0.716** | 0.458* | 0.266 |
| | (0.264) | (0.273) | (0.260) | (0.273) | (0.229) | (0.238) | (0.183) | (0.189) | (0.181) | (0.193) | (0.173) | (0.178) | (0.199) | (0.192) |
| APEC bloc | 0.364# | 0.088 | 0.626** | 0.510** | 0.828** | 0.688** | 1.273** | 1.152** | 1.113** | 1.045** | 1.183** | 0.833** | 1.159** | 0.882** |
| | (0.187) | (0.206) | (0.160) | (0.186) | (0.167) | (0.188) | (0.118) | (0.134) | (0.116) | (0.137) | (0.108) | (0.130) | (0.114) | (0.124) |
| TAFTA bloc | −0.113 | 0.133 | −0.236* | −0.090 | −0.326** | −0.165 | −0.238** | −0.139 | −0.105 | −0.011 | −0.096 | 0.050 | −0.320** | −0.008 |
| | (0.090) | (0.096) | (0.094) | (0.102) | (0.092) | (0.103) | (0.090) | (0.100) | (0.096) | (0.105) | (0.085) | (0.093) | (0.086) | (0.084) |
| Western Europe openness | | −0.219# | | −0.006 | | 0.002 | | 0.314** | | 0.253* | | −0.051 | | −0.030 |
| | | (0.124) | | (0.117) | | (0.130) | | (0.115) | | (0.119) | | (0.114) | | (0.117) |

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Western Hemisphere openness | | −0.465** | | −0.223* | | −0.243* | | −0.061 | | −0.009 | | −0.150# | | −0.375** |
| | | (0.101) | | (0.100) | | (0.101) | | (0.087) | | (0.091) | | (0.084) | | (0.091) |
| East Asia openness | | 0.626** | | 0.430** | | 0.465** | | 0.672** | | 0.598** | | 0.936** | | 0.823** |
| | | (0.146) | | (0.133) | | (0.145) | | (0.114) | | (0.111) | | (0.110) | | (0.104) |
| APEC openness | | −0.498** | | −0.300* | | −0.221 | | −0.179 | | −0.312* | | −0.546** | | −0.275* |
| | | (0.142) | | (0.130) | | (0.151) | | (0.125) | | (0.132) | | (0.126) | | (0.113) |
| TAFTA openness | | 0.049 | | −0.105 | | −0.036 | | −0.070 | | −0.036 | | −0.054 | | 0.059 |
| | | (0.090) | | (0.084) | | (0.095) | | (0.083) | | (0.085) | | (0.080) | | (0.079) |
| Number of observations | 1,194 | 1,194 | 1,274 | 1,274 | 1,453 | 1,453 | 1,708 | 1,708 | 1,647 | 1,647 | 1,573 | 1,573 | 1,546 | 1,546 |
| Adjusted $R^2$ | 0.686 | 0.698 | 0.713 | 0.718 | 0.717 | 0.723 | 0.719 | 0.730 | 0.738 | 0.745 | 0.773 | 0.785 | 0.812 | 0.829 |
| Standard error of regression | 1.053 | 1.033 | 1.073 | 1.064 | 1.172 | 1.160 | 1.190 | 1.165 | 1.167 | 1.153 | 1.064 | 1.035 | 1.097 | 1.046 |

**, *, # denotes significant at the 99%, 95%, and 90% levels, respectively.

a. All variables except dummy variables are in logs.

b. TAFTA variables are defined as EU countries plus NAFTA countries.

Let us note in passing that, in an ideal world, we would have data on bilateral trade among provinces, or even among smaller geographical units. Such a data set would have at least three major advantages. The first is that we could then be more precise about the distances rather than being forced to assume, in effect, that the entire economic activity of a large country is concentrated at a single point of mass. This first advantage turns out apparently to be the least important of the three empirically. Second, we would have an even larger number of observations with which to work. The more information in the data set, the more reliably we can answer the questions in which we are interested.

The third and most important advantage of having data at the provincial level is that we could ascertain how trade between two geographical entities is affected by their common membership in a political union. One might infer the intranational bias to trade in other ways—for example, by pondering ratios of trade to output that are low in most countries when judged by the standard of the supposedly borderless world. (This home-country bias to trade is further analyzed in chapter 6 below.)

When one sees that such links as sharing a common language or common membership in an FTA have big effects on trade between geographical units, extrapolation then suggests that common membership in a political union should have effects even larger than those. It would be very instructive to estimate these effects econometrically. It would help us predict declines in the volume of trade among the former constituents of the Soviet Union, or between the Czech and Slovak Republics. It would also help us predict increases in the volume of trade between the old East and West Germany, or among the members of the European Union, if and when hopes for full political union are realized in the 21st century.

Unfortunately, data on trade among such subnational units as provinces or states are seldom available. A rare exception, data on trade undertaken by Canadian provinces, is discussed in chapter 6.[15] The data analyzed in most of this book, however, are all at the national level.

One can measure the distance between two points on the globe in a number of ways. Most of our econometric results use distance measured ''as the crow flies''—what is technically called the great-circle distance between the two latitude-longitude combinations.[16] Attempts to distinguish between land and sea distances and to measure distances along the

---

15. One can obtain reasonably recent data on the international trade of the 50 individual American states but not their trade with each other, which would be crucial. Richardson and Smith (1995) analyze 1987-89 exports. The only known source of data that includes interstate trade pertains to 1963 (even in this case, states' international trade is not broken down by country). Greytak, Richardson, and Smith (1995) plan to analyze the 1963 data, which might shed light on the effect of political federation on trade between states.

16. Similar straight-line measures of distance are used in gravity equations by Linnemann (1966), Eichengreen and Irwin (1995), and many others.

**Table 4.4 US trade by mode of transportation, 1970-93**

| | Fraction of value going by air (percent) | Value (billions of dollars) | | | Shipping weight (billions of kilograms) | |
|---|---|---|---|---|---|---|
| | | All methods[a] | Vessel | Air | Vessel | Air |
| US imports | | | | | | |
| 1970 | 8.5 | 40.0 | 24.8 | 3.4 | 271.4 | 0.3 |
| 1980 | 11.6 | 240.8 | 165.1 | 28.0 | 443.1 | 0.6 |
| 1985 | 14.9 | 345.3 | 208.4 | 51.3 | 361.5 | 1.3 |
| 1990 | 18.4 | 495.3 | 283.4 | 90.9 | 496.3 | 1.7 |
| 1991 | 19.4 | 487.1 | 272.6 | 94.3 | 449.0 | 1.6 |
| 1992 | 19.8 | 532.4 | 293.1 | 105.3 | 476.0 | 1.7 |
| 1993 | 20.6 | 580.5 | 310.4 | 119.7 | 531.0 | 1.9 |
| 1994 | 21.6 | 663.4 | 339.4 | 143.0 | 586.9 | 2.2 |
| 1995 | 23.4 | 743.5 | 356.6 | 174.2 | 562.5 | 2.3 |
| US exports | | | | | | |
| 1970 | 14.1 | 43.2 | 24.6 | 6.1 | 218.0 | 0.4 |
| 1980 | 20.9 | 220.7 | 120.9 | 46.1 | 363.7 | 1.0 |
| 1985 | 24.5 | 213.1 | 91.7 | 52.3 | 317.7 | 0.8 |
| 1990 | 28.1 | 393.0 | 150.8 | 110.5 | 372.4 | 1.5 |
| 1991 | 27.3 | 421.7 | 162.4 | 115.1 | 385.5 | 1.6 |
| 1992 | 27.0 | 447.5 | 169.8 | 121.0 | 379.3 | 1.7 |
| 1993 | 29.1 | 464.9 | 166.6 | 135.1 | 349.5 | 1.7 |
| 1995 | 29.3 | 512.4 | 177.6 | 150.3 | 334.5 | 2.0 |
| 1995 | 31.1 | 583.0 | 215.9 | 181.1 | 401.1 | 2.3 |

a. Includes types other than vessel and air (i.e., land) and revisions that are not distributed by method of transport.

*Source:* US Bureau of the Census, Statistical Abstract of the United States (CD-ROM), ''Highlights of U.S. Export and Import Trade, through 1988;'' thereafter, ''U.S. Merchandise Trade, Selected Highlights.''

shipping routes actually followed, described below, do not turn out to shed a great deal of additional light.

This seems to be a good place in the book to observe that much trade goes by neither land nor sea these days, but by air: 23 percent of US imports in 1995, up from 8 percent in 1970, and 31 percent of US exports, up from 14 percent (table 4.4). This represents a large increase in the use of air transport at the expense of sea transport. (The estimated share by land is up a little in the case of exports and down in the case of imports.) As recently as 1970, the ratio of value of US imports shipped by vessel to shipments by air was 7.3; by 1995 it had fallen to 2.0. For US exports, the ratio of value shipped by vessel to shipments by air has fallen from 4.0 to 1.2. Extrapolating logarithmically, more American goods, by value, will be shipped abroad by air in 1999 than by sea. Air routes, whether used for shipping goods or human travel, would be the most convenient justification for using the straight-line or great-circle measure of distance.

The ultimate justification, however, is that it seems to be a reasonable way of averaging across different modes of transportation and works well in practice.

We noted when analyzing the c.i.f. measure of transport costs in chapter 3 that it does not seem to rise linearly with distance, but rather less rapidly than that. This is as one would expect: there is a large cost associated with loading cargo onto a vessel and then a relatively small marginal cost per mile of distance traveled. The logarithm has the property that average cost diminishes with distance (assuming the coefficient comes out less than 1). We usually specify distance, like the other variables, as entering the equation in log form. We have also tried other functional forms, however, as explained below.

It must be admitted that transport costs will not always and everywhere be monotonically increasing in distance, let alone in a convenient logarithmic form. Lipsey and Weiss (1974) showed that the distance a product is shipped is positively associated with the level of transport costs, but with more unexplained variation than one would expect. Other factors found to be important determinants of shipping costs were bulk (cubic feet per ton), small consignments (under one ton), the unit value of exports (which they attribute to price discrimination on the part of oligopolistic shippers), and the commodity in question (grains can be shipped in more open, competitive markets, specifically by tanker rather than liner).

One can imagine a variety of ways to measure transport costs more directly than simply using distance. The c.i.f.-f.o.b. differential has drawbacks, which have already been mentioned in chapter 3. A gravity model by Geraci and Prewo (1977) supplements distance data (air routes) with numbers on the ratio of c.i.f. to f.o.b. values (evidently bilaterally), while acknowledging serious measurement error, for 18 OECD countries. This approach has the virtue of allowing rough estimates of the effects of shipping costs per se on trade, expressed as a percentage of value, as opposed to the indirect effects of distance. Their average estimate for transport costs within the OECD is 12.8 percent—higher for the non-European countries and lower (averaging 5.2 percent) for the European countries. The elasticity of transport cost with respect to distance is approximately 0.53. In the tests described in chapter 3, we found somewhat higher elasticities of c.i.f. margins with respect to distance, though the estimates are very imprecise.

## Estimated Effects of Distance and Adjacency

When the adjacency variable is not included in the gravity equation, the estimated coefficient on the log of distance is about −.75 (Frankel 1993). This means that when the distance between two countries is increased by 1.0 percent, trade between them falls by about three-quarters of a percent. The adjacency variable should be included, however. One has only to think of the Mexican *maquiladora* strip along the US border, or the large amount of intermediate products and consumer goods that go back and forth across the Canadian border, to see the relevance of adjacency, beyond distance. The Netherlands is close to France and Korea to Japan, but without the common border the effect is not the same.

When we hold constant for common borders, the estimated coefficient on the distance variable is diminished in magnitude, to the range of about $-0.5$ to $-0.7$, as seen in tables 4.2 and 4.3. In pooled time-series estimates (to be discussed later), the distance coefficients cluster around $-0.6$. Such an estimate implies that when the distance is increased by 1 percent, trade falls by about .6 percent.

The coefficient on the dummy variable for a common border (adjacency) itself is estimated at around .6. Because trade is specified in logarithmic form, the way to interpret the coefficient on a dummy variable is to take the exponent. Two countries that share a common border are estimated to engage in 82 percent more trade than two otherwise-similar countries [$\exp(.6) = 1.82$].

We have tested for possible interactive effects of the common-border variable. For example, one might imagine that small adjacent countries are more highly integrated than would be predicted by the simple sum of their size and common-border effects. But we found no significant interactive effects of adjacency with any of the other variables (reported in Frankel and Romer 1996).

One can also include an effect for landlockedness, which adds to transportation costs. We have estimated the coefficient on this dummy variable at $-0.36$. This means that the lack of ocean ports reduces trade by about one-third, holding constant distance, population, and land area.[17]

We checked for possible nonlinearity in the log-distance term, as it could conceivably be the cause of any apparent bias toward intraregional trade that is left after controlling linearly for distance. The log of distance appears to be sufficient; the level and square of distance add little.[18] We have also tried a more sophisticated measure of shipping distances. Wang (1992) enters measures of such sea distances and land distances separately in a gravity model. She finds a small, though statistically significant, difference in coefficients between the two. Her measure of sea distance

17. The estimate is quite significant statistically. Both its magnitude and significance fall, however, when the regular GNP measure of size, in addition to population and land, is included in the equation, and estimated by instrumental variables to account for the possible endogeneity of income with respect to trade (Frankel and Romer 1996, table 1).

18. A significant positive coefficient on the quadratic term confirms the property that "trade resistance" increases less than linearly with distance, but the log is able to capture this property. Quadratic and cubic terms in the log were not at all significant when tried alongside the log (Frankel 1993).

**Table 4.5 Estimates of coefficients on log distance**

| Source | Estimate |
|---|---|
| Without controlling for adjacency | |
| Brada and Mendez (1983) | −0.76 |
| Bikker (1987) | −0.9 to −1.1 |
| Boisso and Farrantino (1995) | −1.0 to −1.5 |
| Linnemann (1966) | −.77 |
| Oguledo and MacPhee (1994) | −.76 |
| Mansfield and Bronson (1997) | −.51 to −.69[a] |
| Controlling for adjacency | |
| Aitken (1973) | −.35 to −.51 |
| Biessen (1991) | −.74 to −.86 |
| Bergstrand (1985) | −.75 to −.78 |
| Tinbergen (1962) | −.56 |

a. For 1950-90.

takes into account the lengthier trips involved in sea voyages around obstacles such as the Cape of Good Hope and Cape Horn, and she also adds the land distance from the center of the country to the major port. We tried these data, generously supplied by Winters and Wang (1994), in place of own simpler distance measure. The change had relatively little effect on most of the results (Frankel, Wei, and Stein 1995).[19] In short, the precise method of measuring distance appears to be less of an issue than one might have thought.

Our estimate for the effect of distance on bilateral trade is similar to that estimated by many others (table 4.5). Those controlling for adjacency tend to get lower coefficients on the log of distance—as we do and as one would expect.[20]

Bikker (1987), who measures distance by sea routes, tries a clever way of isolating the role of physical shipping costs from the other costs of doing business at a distance. For those years when the Suez Canal was closed by blockade, 1967-75, he adds a variable for the additional sea distance that had to be covered between the country pair in question, divided by the normal distance. The Suez variable is statistically significant. Its estimated coefficient is low, however (0.2)—less than one-fifth of the effect of the regular distance variable. This leads him to conclude that physical shipping costs are less important than conventionally assumed and that other sorts of costs to doing business at a distance are

19. The coefficient of distance varies a bit over the course of the observations but with no clear trend.

20. Leamer (1993) obtains a similar elasticity, .68, for West German trade. He is struck by the importance of distance and concludes that, under NAFTA, Southern California will experience the greatest increase in trade with Mexico.

correspondingly more important. While air transport may explain a bit of the difference, the result suggests that shipping costs do not constitute the majority of costs to doing business at a distance.

We have tried disaggregating trade into three categories. The results show higher distance effects for manufactures than for agricultural products or other raw materials, even though the latter are bulkier. In 1980, for example, the coefficient on distance (−0.58 if estimated on trade in aggregate form) is only −0.30 on agricultural goods, but −0.53 on manufactures (and −0.57 on other raw materials).[21] In other years, the coefficient on manufactures is the highest in absolute value of all three categories: 1965, 1970, 1975, and 1985 (the last year for which we have this disaggregated data).[22] These findings confirm that physical transport costs are not necessarily the most important component of costs associated with distance. The costs associated with transport time and cultural unfamiliarity may be greater, and these costs are more important for manufactured goods than for agriculture.

There is no observable tendency for the effect of distance to fall over time during the course of our sample: 1965-92. If anything, the trend seems to be upward. The same is true for the distance coefficients in the gravity model estimated by others, most extensively by Boisso and Ferrantino (1996) in their year-by-year study between 1950 and 1988.[23]

Indeed, our distance coefficient seems to be quite similar to what other gravity-equation researchers estimated in the 1920s, 1930s, and 1950s. De Menil and Maurel (1994) found a coefficient of −.70 during 1924-26, despite what must have been much higher shipping costs at that time. In analyzing 34 countries (561 bilateral flows) later in the interwar period, Eichengreen and Irwin (1995, 14) also get coefficients similar to ours: −.71 in 1928, −.55 in 1935, and −.51 in 1938. Related estimates in Eichengreen and Frankel (1995) show a distance coefficient of −.48 in 1928, −.37 in

21. The standard errors are all 0.06 to 0.07. The agriculture estimates are significantly less than the others.

22. This was reported in Frankel, Stein, and Wei (1994, table 2) but omitted in the published version to save space. See also Frankel, Wei, and Stein (1995, tables 4a and 4b). Smeets (1994) gets similar results when studying Germany's bilateral trade in 1978: the negative effect of distance is least evident when estimated for crude materials, is more evident for food and live animals (followed by beverages and tobacco), and is consistently very highly significant for mineral fuels, manufactures (the bulk of Germany's trade), and other categories.

23. Earlier, when Boisso and Ferrantino (1993) estimated the gravity model without allowing a role for FTAs, they found that the apparent coefficient on distance did decline steadily over the period. But the coefficient must have been appropriating some of the effect of the regional trading arrangements. This illustrates the need to allow for both bloc effects and proximity effects in explaining existing regional concentration, neither omitting the former (as in the Krugman-Summers school) nor the latter (as in the Bhagwati-Panagariya school).

1935, and −.34 in 1938. The coefficient thus tends to be lower in magnitude than estimates from the postwar period.

This trend goes against what one might expect from declining transportation costs. The authors attribute it to missing variables. But we suspect that the coefficient of distance is not reliably linked to physical shipping costs. Eichengreen and Irwin (1997) confirm those estimates for the interwar period and find similar effects for distance in 1949, 1954, and 1964, as well. Compounding the apparent failure of the distance effect to decline over time, Flandreau (1995) finds coefficients for the 19th century that are even lower in absolute value: −.48 for 1860, −.49 for 1870, and −.27 for 1880.[24] In short, we have more than a century of gravity estimates, and nowhere is there evidence of a decline over time in the distance coefficient.

This trend in the distance coefficient, or lack thereof, is at first thought surprising. At second thought, it looks like a consequence of the proposition that physical shipping is only one of several sorts of costs associated with doing business at a distance. Neither reaction is correct. The other costs to doing business at a distance, such as unfamiliarity with foreign cultures, should be declining over time right along with physical shipping costs, as a consequence of increased ease of travel and communication.

The most likely of the possible explanations is rather that, even though the average cost of shipping per kilometer has undoubtedly declined over time, there is no reason to think that the *marginal cost per percentage increase in distance* has declined over time. But this is what the coefficient in the gravity equation measures. The reverse could even be the case, if technology has for some reason reduced transportation costs at relatively short distances more rapidly than it has reduced costs at long distances. Trucking, for example, may have reduced shipping costs at short distances by more than innovations in air and sea shipping have reduced costs at long distances. Imagine that technological progress reduced shipping costs *at all distances* by some fixed percentage of their previous level. Then there would be no reason for the coefficient on log distance to fall.[25]

## Linguistic Links

Next, we added a dummy variable to represent when both countries of a pair speak a common language or had colonial links earlier in the century. We allowed for English, Spanish, Chinese, Arabic, French, German, Japanese, Dutch, and Portuguese.[26] We allow countries to speak

24. Eichengreen-Frankel, Eichengreen-Irwin, and Flandreau all hold constant for adjacency.

25. I am most indebted to J. David Richardson for this point.

26. Havrylyshyn and Pritchett (1991) find that three languages are significant in the gravity model—Portuguese, Spanish, and English, in decreasing order of magnitude. In a study of poor countries, Foroutan and Pritchett (1993) find that French, Spanish, and English are statistically significant.

more than one language. Switzerland, for example, is counted as having linguistic links with both France and Germany.

The results show a highly significant effect when all languages are constrained to have the same coefficient. The estimate fluctuates over time between .33 and .77. Pooled time-series estimates of the coefficient cluster around 0.44. The implication is that two countries sharing linguistic/colonial links tend to trade roughly 55 percent more than they would otherwise [exp(.44) = 1.55]. In other words, the effect of sharing a common language, even for far-removed countries, is very similar in magnitude to the effect of sharing a common border.

We tested whether some of the major languages were more important than others. When we tried supplementing the general language term and allowing each of the five major languages to have an independent extra coefficient, the language effects lost statistical significance for half the years, due to multicollinearity (Frankel and Wei 1995c, table 2).[27] Nonetheless, two languages, English and Chinese, appear to qualify as especially important. Two Chinese-speaking countries appear to trade four times as much as other countries. There is a problem here, however. Direct Taiwan-China trade is not reported in the official statistics because it was officially nonexistent during our sample period. Much of it goes through Hong Kong and is thus counted twice. An attempt to correct for this factor turns out to eliminate the extra effect of the Chinese language term. The difference between the Chinese effect and that for other languages is not, after all, highly significant (Frankel and Wei 1993a, table 3).

Somewhat surprisingly, the inclusion of the linguistic/colonial terms has little effect on the other coefficients. The regional bloc effects, which we will turn to in the next chapter, are not much lower when linguistic links are included than when they are not.[28]

Boisso and Ferrantino (1996) construct a new measure of linguistic distance that is a continuous scalar rather than a discrete dummy variable, thereby taking into account linguistic diversity within countries. They generate the percentages of the population in a given country speaking each language, and then construct an index of linguistic dissimilarity for each pair of countries. The coefficient on their index exhibits a trend whereby trade in the postwar period has taken place among countries that are increasingly similar linguistically. They interpret this as possible evidence of increased cultural barriers to trade. That interpretation is consistent with the upward trend in our estimated language coefficients. Their result, however, does not hold constant for the other variables in

27. When manufactured goods are considered alone and the five individual major languages are estimated independently, the language coefficients lose all statistical significance (Frankel, Stein, and Wei 1994, table 5).

28. Estimates of the equation without the language effect were reported in Frankel (1993).

the gravity model. It could just be a reflection of the effect of regional trading arrangements that themselves tend to be drawn along linguistic lines. When the gravity model takes into account the increasing importance of FTAs, the significance and trend in the coefficient on the Boisso-Ferrantino linguistic dissimilarity variable are less clear.

## The Panel Approach—Pooling Time Series and Cross-Section

To bring the most information possible to bear at once, one can pool the data across the cross-section and time-series dimensions. Estimates will be reported in the next chapter. We flag one immediate benefit of pooling. We can now conduct meaningful tests for such FTAs as NAFTA and the Australia-New Zealand Closer Economic Relationship that could not be reliably tested in the cross-section dimension because of the very small number of observations (three for NAFTA and one for Australia-New Zealand).

To pool observations from different years without either allowing each to have a separate constant term or else converting the dollar figures to suitable real terms would let inflation distort the estimates.[29] Furthermore, our theoretical rationale for the gravity model (chapter 7) says that, in a time-series dimension, we should normalize for growth in real gross world product. We have adopted the approach of computing year-specific intercept terms, which absorb the effects of global inflation and growth.[30]

29. Bikker (1987, note 3) points out that this is a problem with Aitken (1973).

30. Appendix C discusses some of the details of the interpretation of the intercept. Brada and Mendez (1988) also allow yearly constant terms.