

Jeff Frankel, Senior Fellow; Jan. 2, 1996; rev. 6/26

For REGIONAL TRADING BLOCS,
forthcoming, Institute for International Economics, Washington DC, 1997

CHAPTERS 4, 5 and 6: THE GRAVITY MODEL

CHAPTER 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: its empirical success at predicting bilateral trade flows, its improved theoretical foundations arising mostly from modern theories of trade in imperfect substitutes, and a new interest among economists in the subject of geography and trade, which seeks to treat countries or regions as physically existing at particular locations in space rather than as disembodied constructs.

It is not easy to decide whom to anoint the original 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 (1967), who continued work first reported in Tinbergen (1962), who in turn was contemporaneous with Pöyhönen (1963). Specialists in other fields had, however, used versions of the gravity model before

the international economists. Regional economists and urban sociologists used it by name as far back as 1946, for example.¹ No doubt there are other early references. It seems safest to cite Newton (1687) as the original progenitor of the gravity model!

The gravitational model of Newton 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 GNP_i and GNP_j , and inversely related to the distance between them. Other explanatory variables that are often added include populations (or per capita GNPs), land areas, and dummy variables representing 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 like the gravity model to the data by means of OLS (Ordinary Least Squares) 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 obvious 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

¹ Isard (1960), Steward (1948), and Zinf (1946).

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, 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 grouping, 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 an FTA is correlated with geographical proximity. That is what makes regional trading arrangements "regional."² Despite this correlation, the regression analysis is still capable of separating out the independent effects that each has on trade (so long as we have not omitted any correlated factors from the list of explanatory

² Non-regional preferential agreements exist, of course, like 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.

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 at the same time as testing for the existence of an operational FTA among a subset of countries, such as the EEC Six, one will generally get a different answer than if one tests for the smaller subset alone. This is as it should be. The EEC Six variable has a different interpretation when the EU variable is included at the same time 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 grouping, then the error is ours, in omitting the proper grouping 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 possible hypotheses of interest include groupings that do not yet formally operate, such as APEC or the continent-wide groupings, then it is a matter of judgment whether one is interested in the regression results that include these groupings. In other words, those readers who do not consider of interest the hypothesis that there is currently an intra-regional bias to trade in APEC, 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 EU, NAFTA, and the other formal FTAs, if those are the questions they find of most interest.

One could argue that dummy variables for *all* possible groupings should be tested, so that the data can decide for themselves what questions they believe 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.³ There is the

³ More technically, the OLS estimate is an *unbiased, efficient, and consistent* estimate

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 of silly questions, the answers to the questions we really care about will become increasingly haphazard.⁴ In the extreme, if we were to try to estimate more coefficients than there are data points, one would not get an answer at all. The set of all possible groupings among 63 countries is in fact very large (2 to the 63rd power, minus 65). Thus we must use some discretion in choosing what groupings to test for.

of the true parameter, which in this case is zero.

⁴ 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.

Less silly than fabricating non-existent groupings 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 whether or not 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 GNPs (richer countries are more open), dummy variables for Hong Kong and Singapore (which function as entrepôts), and a dimension of openness that is shared by other members of the groupings 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 results reported here. We have tried also giving each country its own intercept, and we report some of the results below.⁵ But the results appeared unreliable when testing country effects and bloc effects at the same time.⁶ Thus our standard equation does not allow each country to have its own intercept.

The Gravity Equation

We now turn to the origins of the equation itself.

⁵ Appendix Table A10.1, described at the end of Chapter 9.

⁶ 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 is responsible when surprising results come out for estimates of some of the parameters about which we care most.

Theoretical foundations of the basic gravity model

The gravity model is anxious to prove its theoretical pedigree, and is somewhat sensitive to aspersions that have been cast on its respectability. Earlier work surveyed by Deardorff (1984, pp.503-6) provided a partial foundation for the approach. Leamer and Stern (1970), for example, noted that bilateral trade is indeterminate in the absence of transport costs, 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 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 GNPs 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 [previously 8]). This is the theory where consumers love 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. Deardorff (1984, 500-504) 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 will 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 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. [p.4]

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 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 [previously 8]. We assume, as do they, that transportation costs raise the price of a good in the importing country, and that distance has a monotonic effect on transportation costs.

Once one has rationalized a role for distance as an element that raises 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 analogously to the way that 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 for anyone 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 called *psychic costs* and Garnaut (1993) *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 as well, some pursued as part of the present project, 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-yearly intervals, and more frequently at the end of the sample. Thus we have estimates for 1965, 1970, 1975, 1980, 1985, 1987, 1990 and 1992 [and some for 1994]. There is enough data in the cross-section dimension that we can estimate each year separately. Our data source is the UN trade matrix [eventually supplemented by the IMF's Direction of Trade statistics], which covers 63 countries in our data set, so that there are 1,953 data points ($=63 \times 62/2$) for a given year. This trade constituted 88.7 per cent 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, smaller, regional arrangements, but it would have cut short the time span for which sufficient data is available to allow useful estimation. The countries are listed in Table 4.1 [previously 5.1], by geographic grouping.

[Table 4.1 (previously 5.1) about here]

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 EC and 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 various natural resources seem 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. 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 some 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 should do a good job of explaining only trade among OECD countries. They were surprised to find that it works equally well for the larger set of countries. Varieties from 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 vs. Romanian wine vs. French wine) or manufactures (Chinese jeans vs. 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 than for 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/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.⁷ 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 percent 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. Inter-state 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 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, 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.⁸ 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

⁷ 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).

⁸ The estimate is highly significant statistically. Frankel and Romer (1995), Table 1.

equation, with some typical parameter values substituted in for concreteness:

$$(1) \log T_{ij} = .7 \log(GNP_i GNP_j) + .3 \log[(GNP/pop_i)(GNP/pop_j)] - .7 \log(Dist_{ij}),$$

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:

$$(2) \log T_{ij} = 1.0 \log(GNP_i GNP_j) - .3 \log(pop_i pop_j) - .7 \log(Dist_{ij}).$$

In equation (1), the tendency of trade to rise less-than-proportionately with size is reflected in a GNP coefficient less than 1; in equation (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 (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 there are in fact two independent effects. So this is an area that bears further research. One possibility is that exotic foreign varieties are superior goods in consumption. Low-income countries may be dominated by subsistence farming. Other possibilities come out of the literature on endogenous growth.⁹ 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 more developed countries have more

⁹ 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).

advanced transportation infrastructures, including seaports and airports, that help facilitate trade.

The most important reason why industrialized countries trade more than less developed countries is probably the pattern that countries tend to liberalize as they develop. One reason for this pattern is that very poor countries depend on tariff revenue for a large share of their budget, while more advanced countries are able to 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 term for per capita income. But given the absence of good measures of this kind and the importance of non-tariff barriers, the most important reason to expect per capita GNP to have a negative effect on trade may be the correlation with protectionist trade policies.

To take an example, China and Japan have roughly the same aggregate incomes. Yet China trades less with its various 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/population*.

Perhaps the easiest way of keeping in mind the variety of factors that are potentially involved in this relationship is to think in terms of a third way of expressing the same equation:

$$(3) \log T_i = 1.0 \log[(GNP/pop_i)(GNP/pop_i)] + .7 \log(pop_i)(pop_i) - .7 \log(Dist_{.i}).$$

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 takes the form entirely of 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 per cent a year. Then the equation predicts that their trade with slow-growing countries will increase by 8 per cent a year, and their trade with each other will increase by 16 percent a year (= 8+8).¹⁰

¹⁰ 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 per cent per year faster than the world average, then their trade with other countries also grows 6 per cent faster, and their trade with each other grows 12 per cent faster.

It is also instructive to focus explicitly on the product of per capita GNPs as a determinant of trade, as in equation (1). The prediction that per capita incomes enter the equation in multiplicative form is contrary to the prediction of traditional Heckscher-Ohlin theories of trade. If the two factors of production are capital and labor, then these theories have the prediction that *countries with dissimilar levels of output/population will trade more than countries with similar levels*. (Or, more precisely, dissimilar capital/labor ratios.) The standard gravity model makes the prediction that *countries with similar levels of output/population will trade more than countries with dissimilar levels*.¹¹ 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 being 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) and (GNP/pop) will have a negative 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

¹¹ The Helpman-Krugman sort of theory, and its child the standard gravity equation, also makes the prediction 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 per cent in 1970, and then to 26 per cent in 1993. Thus a more uniform size distribution among economies is indeed one explanation for the increase in global trade. (Helpman, 1987; Krugman, 1995, p.341.) As noted, Hummels and Levinsohn (1995) find that the same pattern holds for LDCs that Helpman found for OECD countries.

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 precise 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. Indeed Deardorff (1997 [p.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 analogously will trade more with others of their own kind. These are the same as the predictions 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 various influences -- Heckscher-Ohlin-style factor-endowment differences, Linder-style taste differences, and the effect of 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 7-8. 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) 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 at the same time 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 a negative coefficient.¹² 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 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 tried adding terms for differences in factor endowments, as a more direct test

¹² 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, and Deardorff, 1982.)

of the Heckscher-Ohlin hypothesis. The results are described below, in Chapter 6.

Our Results for the Basic Gravity Variables

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

$$\log(T_{ij}) = \alpha + \beta_1 \log(GNP_i GNP_j) + \beta_2 \log(GNP_i / pop_i GNP_j / pop_j)$$

(4)

$$+ \beta_3 \log(DIST_{ij}) + \beta_4 (ADJ_{ij}) + \beta_5 (LANG_{ij}) + \gamma_1 (WE_{ij}) + \gamma_2 (WH_{ij}) + \gamma_3 (EA_{ij}) + u_{ij}.$$

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. $Lang_{ij}$ is equal to one when they share a common language or past colonial links. WE , WH , and EA are three examples of the dummy variables we use when testing the effects of membership in a common regional groupings, standing in this case for Western Europe, Western Hemisphere, and East Asia.

Table 4.2 [previously 5.2] reports the results of OLS estimation of the equation, where the trading blocs tested for are six groupings that currently have formal status as regional trading arrangements: the EU, NAFTA, Mercosur, the Andean Pact, ASEAN, and the Australia-New Zealand CER. Table 4.3 [previously 5.3] reports results when, as in equation (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 the most recent data available for 1992. In these tables, for comparability across time, we elect to test the effects of country groupings 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 EU, or that NAFTA

did not exist at all. If we find no significant effects in the early years, we will know a likely reason why.

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

[Table 4.2 (*formerly 5.2*) about here]

[Table 4.3 (*formerly 5.3*) about here]

The 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 [*previously 5.2 and 5.3*], 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.¹³

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

¹³ Linnemann (1966) obtains similar estimates for this parameter (in the range .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.

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 PPP rates.¹⁴ The PPP rates are probably to be preferred in theory. 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, tried the Summers-Heston PPP-rate measure. One effect was to restore to a statistically significant positive value the coefficient on per capita GDPs in a year (1991) when the conventional exchange-rate measure showed a coefficient of the wrong sign.¹⁵ 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 clear and consistent verdict.

Distance: its measurement, and connection to the question of sub-national 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

¹⁴ Boisso and Farrantino (1993, 1995). Linnemann tried both, and found that it made little difference for the results, except for the coefficient on income itself.

¹⁵ The data are not as complete as for earlier years. Appendix table 5.4a shows the results for preliminary 1991 data, with the standard exchange-rate-evaluated GNPs, and Appendix table 5.4b shows the same regression with the PPP-evaluated numbers.

countries. The cities are usually the capitals of the two countries, but in a few cases we substitute for the capital a top city that seems closer to the country's economic center of gravity (Chicago for the United States rather than Washington, D.C., and Shanghai for China rather than Beijing). Table 4.1 [previously 5.1] lists the cities that we use as the location points for each of the 63 countries.¹⁶

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 would be able to ascertain how trade between two geographical entities is affected by their common membership in a political union. One might infer the intra-national bias to trade in other ways, e.g., 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.) Alternatively, 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 even larger effects. It would be very instructive to estimate these effects econometrically. It would help us predict declines in the

¹⁶ Boisso and Farrantino (1995) find very little difference in gravity equation results whether measuring distance between most populous cities or geographical centers.

volume of trade among the former constituents of the Soviet Union, or 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 sub-national units as provinces or states are seldom available. A rare exception, data on trade undertaken by Canadian provinces, is discussed in Chapter 6.¹⁷ The data analyzed in this book, however, are all at the national level.

¹⁷ 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 inter-state trade pertains to 1963 (though states' international trade is not broken down by country). Greytak, Richardson and Smith (1995) plan to analyze the 1963 data, which are capable of estimating the effect of political federation on trade between states.

One can measure the distance between two points on the globe in a number of different 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.¹⁸ Attempts to distinguish between land and sea distances, and to measure distances along the trade routes actually followed, described below, do not turn out to shed a great deal of light.

This may be a good place in the book to observe that much trade goes neither by land nor sea these days, but by air: 21 per cent of U.S. imports in 1993, up from 8 percent in 1970, and 29 per cent of U.S. exports, up from 14 percent. 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 U.S. imports shipped by vessel to shipments by air was 7.3; by 1993 it had fallen to 2.6. For U.S. exports, the ratio of value shipped by vessel to shipments by air has fallen from 4.0 to 1.4. The numbers are reported in [new] Table 4.4.¹⁹ Extrapolating (logarithmically), more American goods, by value, will be shipped abroad by air in the year 2000 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 a variety of different modes of transportation, and works well in practice.

[new Table 4.4 about here]

We noted when analyzing the c.i.f. measure of transport costs in chapter 3 that it does not

¹⁸ Similar straight-line measures of distance are used in gravity equations by Linnemann (1966), Pritchett (), Eichengreen and Irwin (1995), and many others.

¹⁹ The source is the U.S. Bureau of the Census, Highlights of U.S. Export and Import Trade, through 1988, FT 990, and thereafter, U.S. Merchandise Trade: Selected Highlights, FT-920; as reported in Statistical Abstract of the United States.

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 one). 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 the unit value of exports (which they attribute to price discrimination on the part of oligopolistic shippers), bulk (cubic feet per ton), small consignments (under one ton), 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 of attempting 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. A gravity model by Geraci and Prewo (1985) 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 percent of value, as opposed to the indirect effects of distance. Their average estimate for transport costs within the OECD is 12.8 per cent – higher for the non-European countries and lower (averaging 5.2 per cent) for the European countries. The elasticity of transport cost with respect to distance is approximately 0.53.

In tests described in Chapter 3, we found somewhat higher elasticities of c.i.f. margins with respect to distance.

Estimated effects of distance and adjacency in our gravity equation

When the adjacency variable is not included in the gravity equation, the estimated coefficient on the log of distance is about $-.75$.²⁰ This means that when the distance between two countries is increased by 1.0 per cent, trade between them falls by about 3/4 of a percent. The adjacency variable should be included, however. One has only to think of the Mexican *maquilladora* strip along the U.S. 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, to the range of about -0.5 to -0.7 , as seen in the tables. 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.

[Table 5.4 about here? No. Old table DROPPED]

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 per cent more trade than two otherwise-similar countries [$\exp(.6)=1.82$].

²⁰ Frankel (1993).

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 the size and common border effects. But we found no significant interactive effects of adjacency with any of the other variables.²¹

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.²²

²¹ Reported in Frankel and Romer (1995).

²² 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 (1995, Table 1).

We checked for possible non-linearity in the log-distance term, as it could conceivably be the cause of any apparent bias toward intra-regional 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.²³ 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. Her measure of sea distance takes into account the lengthier trips involved in sea voyages around obstacles like 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 (1992), in place of our simpler distance measure. It had relatively little effect on most of the results.²⁴ 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. The table displays some of their results. Those controlling for adjacency tend to get lower coefficients on the log of distance, as we do and as one would expect.²⁵

²³ 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).

²⁴ Frankel, Wei, and Stein (1995). The coefficient of distance varies a bit over the course of the observations, but with no clear trend.

²⁵ 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.]

[Table 4.5 [formerly 5.5] about here]

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, 1967-1975, 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, but its low estimated coefficient [0.2], less than one fifth of the effect of the regular distance variable, 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 correspondingly more important. While air transport may explain a bit of the difference, this result is a good reminder for us 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, which 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).²⁶ In other years, the coefficient on manufactures is the highest of all three categories: 1965, 1970, 1975, and 1985 (the last year for which we have this disaggregated data).²⁷ These findings confirm to us that

²⁶ The standard errors are all 0.06 to 0.07. The agriculture estimates are significantly less than the others.

²⁷ Reported as Table 2 in Frankel, Stein and Wei (1994), but omitted in the published version to save space. (Henceforth we often abbreviate Frankel, Stein, and Wei by last initials.) Also

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 greater for manufactured goods than for agriculture.

Tables 4a and 4b of F, W and S (1995). 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.

There is no observable tendency for the effect of distance to fall over time during the course of our sample, 1965-1992. 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 Farrantino (1995) in their year-by-year study between 1950 and 1988.²⁸

Indeed, our coefficient seems to be quite similar to what other gravity-equation researcher estimated in the 1920s, 30s, and 50s. De Menil and Maurel (1994) found a coefficient of -.70 during the period 1924-26, despite what must have been much higher shipping costs at that time. Eichengreen and Irwin (1995, p.14), analyzing a set of 34 countries (561 bilateral flows) later in the inter-war period, also get coefficients similar to ours: -.71 in 1928, -.55 in 1935, and -.51 in 1938. Related estimates in Eichengreen and Frankel (1994) show a distance coefficient of -.48 in 1928, -.37 in 1935, and -.34 in 1938. The coefficient thus tends to be lower than estimates from the post-war 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 coefficient to decline over time, Flandreau (1993) finds coefficients for the 19th century that are even lower: -.48 for 1860, -.49 for 1870, and -.27 for

²⁸ Earlier, when Boisso and Farrantino (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 other (as in the Bhagwati-Panagariya school).

1880.²⁹ In short, we have from different authors 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.

²⁹ Eichengreen-Frankel, Eichengreen-Irwin, and Flandreau all hold constant for adjacency.

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.³⁰

Linguistic links

³⁰ I am most indebted to J. David Richardson for this point.

Next, we added a dummy variable to represent when both countries of a pair spoke a common language or had colonial links earlier in the century. We allowed for English, Spanish, Chinese, Arabic, French, German, Japanese, Dutch, and Portuguese.³¹ (We allow countries to speak 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 [to 0.57 (e.g., Tables 6.1-6.3 **DROPS.4**)]. Two countries sharing linguistic/colonial links tend to trade roughly 55 per cent more than they would otherwise [$\exp(.44)=1.55$]. In other words, the effect of sharing a common language, even for countries far-removed, 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 the 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.³² Nonetheless, two languages, English and Chinese, seem to qualify. Two Chinese-speaking countries appear from the official statistics to trade four times as

³¹ 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 (1992) find that French, Spanish and English are statistically significant.

³² Table 6 in Frankel and Wei, 1995a. [When manufactured goods are considered alone and the five individual major languages are estimated independently, the language coefficients lose all statistical significance (Table 5 in F, S and W, 1994).]

much as other countries. There is a problem here, however. Direct Taiwan-China trade is not reported in the statistics, because it was officially non-existent 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.³³

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.³⁴

Boisso and Farrantino (1995) 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 over time in our estimated language coefficients. Their result, however, does not hold constant for the other variables in 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-Farrantino linguistic dissimilarity variable is less clear.

³³ Table 3 of F & W (1993a).

³⁴ Estimates of the equation without the language effect were reported in F (1993).

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. Tables 5.1 and 5.2 will report estimates that pool the four years 1970, 1980, 1990, and 1992. One table tests the groupings that qualify as formal regional trading arrangements, and the other tests the broader continental groupings. Table 5.3 [formerly 6.3] will allow for a linear trend in the bloc coefficients over this period. (See also Appendix Tables A5.2 and A5.3.)

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, which could not be meaningfully 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 be to allow inflation to distort the estimates.³⁵ 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.³⁶

³⁵ Bikker (1987, fn.3) points out that this is a problem with Aitken (1973). Brada and Mendez allow yearly constant terms.

³⁶ A brief Appendix to Chapter 4 discusses some of the details of the interpretation of the intercept.

CHAPTER 5: ESTIMATED EFFECTS OF TRADING BLOCS

If there were nothing to the notion of trading blocs, then the five basic variables in the gravity equation – size, per capita income, bilateral distance, common borders and common languages – might soak up all the explanatory power. There would be little left to attribute to a dummy variable representing whether two trading partners are both located in the same region or preferential trading area. In this case the observed tendency for trade to be intra-regionally concentrated would be due solely to the proximity of the countries. The observed tendency for this concentration to increase, for example, in Asia, would be due solely to its rapid rate of overall economic growth. In this chapter, however, we find that dummy variables for intra-regional trade are highly significant statistically.

As we have already mentioned, there is a choice whether to define a bloc as consisting of only those countries who were formal members of the PTA (when there was a PTA) as of the year in question. The disadvantage of doing this is that it impairs the comparability of results across years. Do we want the reported EC effect to go down from one year to the next if three new members join and, even though rapidly integrated into the club, they have a somewhat lower degree of trade concentration than the average of the members who have long been in the club? We have generally opted for comparability over legal niceties. One reason for doing so is that it would not always be clear precisely in what year specific countries joined. Typically the year that an agreement is negotiated is different from the year it is ratified, which is in turn different from the year it goes into effect, which is in turn different from the year that the transition period of trade liberalization is completed. There is a tendency for trade flows to be affected in advance of the date when the agreement goes into effect, as businesses in anticipation position themselves for future markets. Another reason for defining the membership of blocs to be uniform over time is that some of the

groupings in which we are interested are possible continent-wide informal trading areas that do not (yet) constitute formal PTAs. To impose a legal date of inauguration on them would be to exclude them from consideration entirely. Nevertheless, in the case of explicit regional arrangements, we will also report the results of some estimates that carefully distinguish the effects before and after the formal establishment of an FTA or the expansion in its membership.

We begin by looking at the apparent effects of formal regional trading arrangements defined by their membership in 1996. Given that these institutions either did not exist in 1965, the beginning of our sample period, or did not consist of their current membership, we do not necessarily expect to see effects early in the sample period. The aim, rather, is to see when their effects seem to take hold.

The reader is also warned from the beginning that the estimated bloc effects seem to bounce around, more so than one would expect from the variation over time in trade policies and their effects. In some cases, the results, even if surprising, are probably telling us something that we need to know. But in other cases, the fluctuations seem too great to be plausible. We take this to be the result of estimation error. This is why we also report the results of some tests that pool the time series and cross-section observations together, which forcibly smooths out some of the variation (Tables 5.1, 5.2 and 5.3). If, on the other hand, one thinks that the variation over time in estimated coefficients contains meaningful information, and that the impact of formal regional trading arrangements should be identified with a particular year of implementation, then one should look for a change in the coefficient over the interval that includes that year. This is why we also report estimates of the effect of bloc formation on the change in trade.¹ Finally, we report some results for

¹ In taking the first difference of the equation, such variables as distance and linguistic links drop out. Probably as a result, the estimates are less precise and the effects are usually insignificant

trade disaggregated by major sector.

[Tables 5.1, 5.2 and 5.3 (formerly 6.1-6.3) about here]

5.1: Trading Blocs in Europe

We begin with Europe, the continent where regional trading arrangements appear to be the furthest advanced, in terms of formal agreements as well as in terms of the level of intra-regional trade.

The European Union

statistically when estimated in this way.

The relevant grouping for western Europe has changed frequently. We start, in Table 4.2, at the level of the European Union, counting all fifteen current members (as of 1996). The dummy variable represents when both members of the country pair are among the EU 15. Its coefficient is statistically insignificant in the years 1965 through 1980. Only in 1985 does it attain a statistically significant level of .2, rising to .3 in 1990. The estimate suggests that in 1990, two members of the EU Fifteen traded only 35 percent more, after holding constant for GNP, proximity, etc., than two otherwise similar countries [$\exp(.3)=1.35$]. The somewhat weak effect could either be due to the fact that the European Union did not exist and many of the 15 were not in the precursor EEC, or to the progressively more serious steps toward integration that were taken over time by the EC.²

The EC

² The EU bloc effect becomes even weaker when we introduce terms into the equation to allow for the general level of openness of the EU and of the other groupings. These are dummy variables that are equal to 1 when at least one country out of the pair in question is a member of the grouping (as opposed to the bloc dummy, which is equal to 1 when *both* countries of the pair is located in the EU). The results are reported in Table 5.2, in the second column under each year. Introducing these terms deprives the EU bloc term of its statistical significance in 1965, 1985 and 1990. The implications are discussed in the next sub-section.

In other tests, we have tested for the membership of the EC and EFTA as they stood at the end of the 1980s. There is an upward trend in the coefficient for the EC. As recently as 1980, the EC bloc effect was not statistically significant in most of our results. Only in 1985, 1990, and 1991, does the EC bloc effect generally become highly significant.³ In some specifications, the EC effect becomes significant as early as 1980, but rarely earlier than that.⁴ A typical coefficient estimate for 1990, a year when the EC actually comprised all 12 of the members that we attribute to it, is .5. This point estimate suggests that two EC members trade an extra 65 percent, relative to two otherwise-similar countries.⁵ If the data from four years -- 1970, 1980, 1990, and 1992 -- are pooled together, the estimated coefficient on the EC is a smaller .15, implying a 16 per cent effect.⁶

³ F (1993, 1994) and F and W (1994a, 1995a, 1995b). The EC bloc coefficient, highly significant in 1985 and 1990, can also be seen in Table A6.1b.

⁴ F & W (1994b, 1995c), and F, W & S (1995).

⁵ $\exp(.5)=1.65$. This estimate can be obtained either from F & W (1994a, Table 12.13; or 1995a, Table 3), which do not allow for the effect of linguistic links, or F & W (1995b, Table 1), which does have a dummy variable for common language. If one allows the English language to have a different effect from Spanish, Chinese, Arabic and French, the difference is positive and significant. In that case, the estimated EC bloc effect is lower in 1990, but higher in 1985. F & W (1995c), Table 2.

⁶ Table 6.1; taken from F & W (1997, Table 1).

When we allow a term to capture groupings' general levels of openness, we find that the members of the EC 12 are more open than one would expect from the gravity determinants. [In 1980 and 1985, for example, a typical estimate of the openness coefficient is a highly significant .4, indicating that an EC member traded about 50 percent more with all partners.⁷] When one takes into account this tendency for EC members to trade with all partners, the tendency to trade with fellow EC-members is slightly less pronounced. Indeed, in the test where the years 1970-1992 are pooled together, thus constraining the EC bloc effect to be constant throughout, the introduction of the openness term changes a significantly positive bloc effect into a negative one.⁸ Still the EC bloc effect remains usually significant in 1980 and thereafter.⁹

Clearly the EC 12 effect is not constant over the sample period. Accordingly, as promised, we now depart from our usual practice of keeping the definitions of the grouping unchanged over time, to study the effects of the expansion of the EEC 6 into the EEC 9 in 1973 (the accession of the United Kingdom, Ireland, and Denmark) and the expansion into the EC 12 during the period 1981-86 (the accession of Greece, Spain and Portugal). We are looking here for an effect expressed in terms of *changes* in bilateral trade flows. Therefore we take the first difference of the gravity equation. Distance, and the border and language dummies, drop out, as they don't change over time, and we are left with the GNP variables.¹⁰ The results show that both expansions of EC membership

⁷ The openness effect is particularly clear for the case of manufactured products. F & W (1995c), Tables 2 and 3; or F, W & S (1995), Tables 2, 4 and 4a.

⁸ Appendix Tables A5.2; also F & W (1996, 1997).

⁹ Results are reported, for example, in Appendix Table A7.2 for 1980; Table 2 in F & W (1995c) for 1985; and Appendix Tables A5.4a and A5.4b, with incomes measured either as GNPs on an exchange rate basis or as GDPs on a PPP basis, respectively, for 1991.

¹⁰ The results are reported in W & F (1995), Table 1.

increased trade between the existing members and the new members. In each case, the point estimate is .3: trade between the existing EC members and the newcomers expanded about 30 per cent over the appropriate five-year interval. The effects of the first expansion are not at all significant statistically, however. The coefficient on the second expansion, estimated at .3, is significant only at the 80 percent confidence level.¹¹

To summarize the results on EC trade so far, it appears that, despite the high level of intra-EC trade in the 1960s and 1970s that shows up in measures such as the trade share in Table 2.1 and Figure 2.1, most of this trade can be explained by the size, level of development, proximity, common borders and common languages that the EC members share. There is little intra-EC trade left over, after correcting for these influences, to be attributed to the effect of the regional trading arrangement itself, until the 1980s. One possibility, which we will consider in the next chapter, is that the stabilization of bilateral exchange rates in the 1980s, under the Exchange Rate Mechanism of the European Monetary System, might be in part responsible for the promotion of intra-European trade after 1979, but not before. The more likely explanation, however, is simply that half of the Twelve were not members in 1965 and 1970.

¹¹ This is also the technique used by Bayoumi and Eichengreen (1995). They find significant effects from both the EC and EFTA. [However, they do not include developing countries in their sample.]

Most of our statistical results are based on data for all merchandise trade. (We have not analyzed trade in services, because comprehensive bilateral data are not available.) We have, however, tried some disaggregation, estimating separate equations for trade in manufactures, trade in agricultural products, and trade in other raw materials (for the years 1965-1985). In the case of Europe, this level of disaggregation appears to sharpen the tests. Whereas the EC bloc effect did not turn positive until 1980 and did not become significant until 1985, when estimated on aggregate merchandise data, it in fact is positive throughout¹², and turns statistically highly significant in 1980, when estimated on manufactures alone.¹³ The estimated effect for 1980 is about .5: two EC members engaged in about 65 percent more manufactures trade with each other than two otherwise-similar countries.

¹² In 1965, the estimate is .6 and even significant at the 99 percent confidence level, when not holding constant for the general openness of EC countries. The estimate is .3 and significant at the 90 percent confidence level, when holding constant for openness. These results are reported in F, S and W (1994), Table 3b; and F, W & S (1995, Table 4a), respectively. They omit the common languages term in the equations.

¹³ The EC loses a little of its bloc effect in manufactures when holding constant for linguistic links (in particular, the highly significant effect of the English language), but is still positive more consistently than when estimated for aggregate merchandise trade. This comparison holds especially if one is not attempting to hold constant for significant EC openness, but also holds even if one does so. The results with linguistic links are reported in F, S & W (1994, Table 5) and F & W (1995c, Table 3), respectively.

The EC bloc effect is even stronger when estimated on agricultural products alone: it is highly significant in every year, with an upward trend. The point estimate in 1985 is 1.4, indicating that two members of the EC 12 trade in agricultural products four times as much as two otherwise similar countries.¹⁴ (The data were not available for 1990 or 1992.) Needless to say, this strong result is the outcome of the interventionist Common Agricultural Policy, which one might not wish to describe as the operations of a free market, even with the usual caveats about trade diversion from outsiders.

The bloc effect for non-agricultural raw materials turns positive in 1975, and significantly so in 1980 and thereafter.¹⁵

The decision to adopt a high level of aggregation in most of this book is deliberate. There exist many other thorough statistical studies of the effects of integration in Europe, and elsewhere, carried out on a disaggregated basis. We wish in this book to be able, in the end, to think in terms of an overall degree of regionalization, without having to choose how to aggregate over different degrees of regionalization estimated within hundreds of individual sectors. Nevertheless, the important point for the purpose of the moment is that the high level of aggregation is capable of

¹⁴ F, W & S (1995), Table 4b, and F, S & W (1994), Table 2. Smeets (1994) had a similar finding in his study of Germany's bilateral trade: agriculture was one of only two sectors (along with chemicals), out of ten, that showed a significant EC bloc effect as far back as 1978. There was no significant EC effect in manufactures, which constituted 67 percent of Germany's trade.

¹⁵ F, S & W (1994), Table 2. Even if the effect is statistically significant within each of the three component categories, it does not necessarily follow that it will appear statistically significant when estimated in aggregate form. The technical econometric problem is *aggregation bias*: if the equation is seriously misspecified when estimated in aggregate form, because in reality the coefficients differ by commodity, then (technically) all bets are off. On the one hand, the fact that most of the coefficient estimates are similar for the three components of trade, should help the aggregation. On the other hand, the estimated coefficients for GNP per capita and distance tend to be lower for the gravity equation estimated on agricultural products, as compared to the case of manufactures. This heterogeneity alone is capable of invalidating the aggregation.

leading us astray. In particular, the regionalization of trade in manufactures among the EC 12 had become highly significant statistically by 1980, even though the tests run on aggregate merchandise trade data make it look like this did not happen until 1985.

European Free Trade Area (EFTA)

We have seldom found EFTA to be a statistically significant bloc.¹⁶ The sole exception, where EFTA shows a coefficient of .2 that is statistically significant at the 95 per cent level, comes in the case where data are pooled from 1970, 1980, 1990, and 1992, and one allows at the same time for the tendency of EFTA countries to trade less with outsiders.¹⁷ The finding that EFTA countries are less open than predicted by the equation appears repeatedly, but in most cases the presence of the openness terms does not make the EFTA bloc effect statistically significant.

Boisso and Farrantino (1995) find a moderate bloc effect for EFTA, but only for the period around 1966-72. They, like we, show no significant effect in the 1980s. (Where our tests look only every five years during the period 1965-1990, theirs are year-by-year.) Hamilton and Winters (1992) also find EFTA insignificant for the one period that they estimate (1984-1986).

Here again, some disaggregation of trade by sector may be helpful. EFTA applied primarily to manufactured goods. When the equation is estimated on trade in manufactures, the EFTA bloc coefficient is usually positive, is of borderline significance in 1975, and is highly significant in 1980. That is only one or two years out of five, however. The same is true for trade in agricultural products and other raw materials, except that here there are clear monotonic upward trends in the coefficients

¹⁶ F, S & W (1995), F & W (1993b, 1995a, 1995c, 1996), F, W & S (1995), and Eichengreen and Frankel (1995, Table 2).

¹⁷ F & W (1997).

over time. By 1985, the bloc effect in farm products reaches a highly significant 1.1: the estimated effect on intra-EFTA trade in this sector is 2.9 times.¹⁸

Western Europe

¹⁸ F & W (1995c), Table 3; F, S & W (1994), Table 2; F, W and S (1995), Tables 4a and 4b.

Next we consider Western Europe as a whole.¹⁹ We find that either the EC or Western Europe is significant when considered by itself. But the true test comes when both are included simultaneously. Here it is clear that the EC is the correct grouping: the Western Europe bloc is insignificant alongside it.

Still, if we are looking toward a potential future world in which all of Western Europe might become one trading bloc, rather than looking at the past, continental aggregate estimates such as those in Table 4.3 might be of interest. Here we see that by 1990, the Western Europe bloc coefficient had reached a highly significant .4, indicating that two Western European countries trade 49 per cent more than two otherwise-similar countries. Unfortunately, this coefficient bounces around more than seems reasonable. In Table 5.2 we pool the four observation years 1970-1992 together and impose the constraint that the Western European bloc effect is constant throughout; the estimated coefficient is .2.²⁰ We allow for a time trend in the coefficient, but the coefficient in the table shows little trend.

When we introduce the openness term for Western Europe, in Table 4.3, it shows up significant and positive in 1980 and 1985, though insignificant in the other years. In the pooled estimates for 1970-92, the openness is significant and positive. Evidently the openness of the EC

¹⁹ Grant, Papadakis, and Richardson (1993, p.53) observe that intra-regional trade is greater in Western Europe than in the EC. Based on this and similar statistics for other groupings, they conclude that "trade indeed follows the flag," i.e., that political affinities may be more important than formal regional trading arrangements. Our finding (from F & W, 1995a) suggests that this is an example of how one can be misled by relying on intra-regional trade shares, unadjusted for country size.

²⁰ Similarly with 1970-1990 pooled, in F, S & W (1993).

Twelve dominates the closedness of the EFTA Five.²¹ Taking into account the general openness of Western European countries reduces the estimate of the tendency for these countries to concentrate their trade with each other.

European bloc effects estimated by others

²¹ F & W (1995a), Tables 1 and 2; also F & W (1995d).

Perhaps the most common use of the gravity model by other authors in the 1970s and 1980s was precisely to evaluate the effects on trade of the establishment of the European Economic Community and its subsequent expansion. Their results often show significant evidence of the hypothesized effects of the EEC and EFTA, such as the study of 1967 trade by Aitken (1973). In Bergstrand's (1985) results, the EC and EFTA were estimated to raise intra-group trade during the period 1965-1976, but only the EEC effect was consistently significant in the preferred general form of the equation.²²

Those studies used data on trade among European or industrialized countries *only*, excluding trade among most non-European countries. Even when one is interested only in the question of European trading arrangements, it helps to have data on other trade as a standard of comparison. Trade among 20 European countries gives only 190 data points ($=20 \times 19/2$). If one expands the number of countries even by a factor of 2, the number of observations goes up with the square, to 780 ($=40 \times 39/2$). Brada and Mendez (1983) extend the sample to include some developing countries, for a total number of 46 countries, during the period 1954-77. They find that the EC raised internal trade by a factor of 5, and EFTA a factor of 4. Smeets (1994) found for 1978 a significant EC effect in most sectors, when considering Germany's trade with the original EC Six alone, but did not when the three newcomers were included in the definition of the EC. Hamilton and Winters (1992) find, as we do, the EC 12 significant for the mid-1980s (1984-1986).

These studies did not hold constant for common languages or, in some case, for common borders, so it is possible that a bit of these effects were picked up by the FTA variables.

²² Hewett (1976) estimated the EFTA effect at 6, and the EC at 4. Bayoumi and Eichengreen (1995) found significant effects of the expansions of the EC in 1972, 1981 and 1986.

Some readers have found surprising our result that intra-European trade can be mostly explained by the various natural factors, with little role for the EC, until the 1980s (with the exception of agricultural trade, where we find the EC effect to be highly significant throughout). As noted, the most likely explanation is that the membership of the EC Twelve was not complete until the mid-1980s. But it is worth noting that some others have gotten negative results even with narrower definitions of the membership. Bergstrand (1985) finds, as we do, the EC effect insignificant in 1976. Bikker (1987, p.330) gets the surprising result that intra-EEC trade in 1974 was less than would be predicted by the gravity equation, though the intra-bloc term had increased relative to 1959.²³ Similarly, Boisso and Farrantino (1995) find that trade among EEC members was actually significantly less than would be predicted in the 1950s, 60s and 70s. The negative term steadily erodes, so that by 1986 the EEC has attained neutrality. They attribute to the formation and expansion of the EEC the trend whereby the European countries have progressively overcome what was originally an apparent reluctance to trade with each other. To the extent that one's impression of a large effect before the 1980s is based on calculations of simple intra-regional shares that do not adjust for these other factors,²⁴ the answers from the gravity approach are probably more reliable.

EC preferences

We have not studied patterns in trade between Western European countries and LDC beneficiaries of EC preferences or Eastern European countries. But we can report here some findings by others. In a study of 1976 data, Oguledo and MacPhee (1994) find a significant positive

²³ Nitsch (1993) obtains a negative EC bloc effect as recently as 1990.

²⁴ E.g., Wonnacott and Lutz (1989 p.76) show increases in intra-regional trade that they attribute to the expansion of the EC and EFTA.

effect of the Lome preferences on trade with the EC [and ten other industrialized countries].²⁵ Hamilton and Winters (1992) found significant effects for the Lome preferences [as for the GSP].

Eastern Europe

Needless to say, trade among Eastern European countries was until very recently artificially diverted away from Western partners, and toward trade among themselves and the Soviet Union (e.g., Holzman, 1985). In a gravity equation, Hewett (1976) estimated a coefficient on intra-CMEA trade of 2.6. Biessen (1991) finds in a study of bilateral trade among seven eastern European countries and fifteen western European countries that trade between the CMEA and the West has been depressed far below what would be normal. Once this factor is taken into account, the independent effects of trade undertaken by centrally planned economies, trade within the CMEA, and even trade within the EC or EFTA, are not statistically significant.

²⁵ And a negative effect that is apparently very highly significant for GSP preferences. These results are probably tainted by failure to allow for the fact that poor countries trade less, even after allowing for tariff levels as they do.

One of the most important uses of the gravity model is to project the natural pattern of trade that would hold if artificial political barriers were removed. Havrylyshyn and Pritchett (1991), Wang and Winters (1991), Hamilton and Winters (1992), Winters and Wang (1994), and Baldwin (1994), all have used the gravity model to predict the normal pattern of trade to which the Eastern European countries could be expected to return after the end of the Cold War.²⁶ We have used the gravity equation to estimate significant 1980 coefficients of -.97 for Hungary, -.98 for Poland, and -.41 for Yugoslavia. These coefficients can be interpreted as the trade-inhibiting effects of Soviet-style economics, which should now be disappearing. Apparently, such predictions are rapidly coming true. Even by 1990, Hungary, Poland and Yugoslavia, show improved openness coefficients in our estimates. Schumacher (1996) finds, in 1992 data, that Germany's trade with the Central and Eastern European countries and the former Soviet Union has already reached the levels implied by the gravity model. (In other words, the negative openness terms have disappeared). Vittas and Mauro (1996) find that, unlike Germany, five other western European countries still have a ways to go (a 64 per cent gap, on average) before experiencing the full predicted share of trade with Central and Eastern European countries.

TAFTA

Even less in operation during the sample period than the EU is the TAFTA, which was only a gleam in the eye of some western leaders as of 1995. There has, however, been sufficient interest in the TAFTA proposal recently to call for a test of the effect of this grouping. We defined a potential TAFTA bloc as including all 15 members of the EU plus all three members of NAFTA.

²⁶ Collins and Rodrik (1991) and Brown, Deardorff, Djankov and Stern (1996) also forecast trade between these countries and the west.

The estimates are reported in Table 4.3. The coefficient is always negative. It even appears to be statistically significant in many years. In 1992, the coefficient is a significant $-.3$, indicating that any two North Atlantic countries trade 35 percent less than one would expect given their GNPs, etc.

For boosters of TAFTA, of which the author does not count himself one, this finding need not be interpreted pessimistically. The negative trans-Atlantic effect means that trade among these countries is less than its true potential. Perhaps the U.S. has been overly distracted by APEC or Western Hemisphere relations, and the EC by its own regional trading arrangements. In any case, current trans-Atlantic trade falls significantly short of the high "natural" level of trade that is justified by incomes and geography. So, one could argue, a TAFTA is needed to correct the situation.

A dummy variable for general openness of TAFTA members reveals no effect. (The term is collinear with the term for openness of Western Europe.) The introduction of openness terms for all the regions, nevertheless, eliminates the statistical significance of the negative trans-Atlantic bloc effect. The bottom line is that there is no evidence of a TAFTA bloc.

5.2: Trading Blocs in the Western Hemisphere

As we turn to the Americas, let us begin with the existing formal FTAs. There are three large ones in the Western Hemisphere, comprising respectively North America, the Eastern half of South America, and the Western half of South America. In no other part of the world has there been such a dramatic shift in the direction of successful preferential trading arrangements.

We begin by considering dummy variables for the three important sub-regions: NAFTA, Mercosur, and the Andean Pact. Tight standard errors and significant coefficients are not to be expected, in light of the small number of observations: 3 ($=3 \times 2/2$) for NAFTA, 6 ($=4 \times 3/2$) for

MERCOSUR, and 10 ($=5 \times 4 / 2$) for the Andean Pact. (Pooling across time should help a bit, to the extent one is willing to view the structure as constant and the observations as independent.) But the point estimates are in any case of interest, as these are the groupings with explicit trade preferences.

We have also tried dropping the individual regional FTAs in Latin America, instead adding one dummy variable to indicate whenever a pair of countries belongs to the same PTA or FTA, regardless which one it is, and another to indicate whenever the pair belongs to the same Customs Union or Common Market.²⁷ This constrains all PTA/FTAs to have the same effect on bilateral trade, and the same for all Customs Unions. The PTA/FTA variable is often statistically significant, particularly when the tests are run on manufacturing products alone. The CU/CM variable is not. It is probably better not to constrain such disparate agreements to have the same effects, and so we do not explore this specification any further here.

NAFTA and the Canada-U.S. FTA

²⁷ Tables 4 and 4a in F, S & W (1994). Recall the distinction that under Customs Unions, external tariffs are made uniform, while under FTAs countries retain their individual external tariffs.

In Table 4.2, the coefficient for a North America bloc is almost never significant, not even in 1992, when NAFTA was actually negotiated. The lack of significance could be due in part to the small number of observations: there are only three pairs of countries in NAFTA. The only exception is when data are pooled over 1970-1992 (and one allows for a significant negative openness term, which we may call a trade-diversion effect). Then the estimated coefficient on the NAFTA bloc is a significant .36, implying a 43 percent effect on intra-NAFTA trade.²⁸

A test for a specific Canadian-U.S. effect yields an insignificant result.²⁹ The outcome is the same when the test is run on the change in trade between 1985 and 1990. (Recall that the Canada-U.S. FTA was signed in 1988.)

We found in the previous section that estimates could in some cases be sharpened by disaggregating merchandise trade, into manufactures, agricultural products, and other raw materials. In the case of North America, this disaggregation does not make a great deal of difference.³⁰ The small number of observations is probably the greatest problem with the estimates. A bilateral relationship, after all, offers only one observation per year. This does not reflect adversely on the efficacy of the Canadian-U.S. Free Trade Agreement or NAFTA. The problem may simply be that we do not have enough data to address the question. This is certainly true in the case of NAFTA,

²⁸ [1.43 = exp(.36).] F & W (1997, Table 1). In Table 4 of F, W & S (1995), however, a negative openness effect for North America (often significant) does not produce a positive NAFTA bloc effect.

²⁹ F & W (1996).

³⁰ In the case of agricultural trade, there is an upward trend in the point estimates of the NAFTA bloc effects, but only in 1980 is it statistically significant. Given that the important formal agreements took place in 1988 and 1992, one should perhaps in any case not expect anything out of a data set that ends in the mid-1980s. [The results are reported in Tables 4a and 4b of F, W & S (1995).]

where the time elapsed since the 1994 implementation is insufficient to allow the drawing of reliable conclusions, regardless of the statistical methodology.

Clausing (1995), however, does have enough data, in the case of the Canada-U.S. FTA. She, as we, finds no statistically significant impact on trade between the two neighbors after 1988, when using a gravity model on aggregate merchandise trade. This is despite that all her tests use data that go up to 1993. But then she focuses on a data set of thousands of goods, disaggregated into a six-digit level of classification. (The use of this disaggregated data to study international trade is made possible by a Harmonized Classification system adopted by the U.S. and Canadian authorities. This would not be possible for other pairs of countries, as different countries have different classification schemes.) At this level of disaggregation, one can use direct information on how high tariffs were before the agreement, and how quickly they are being phased out under the agreement.

The tariff reductions had highly significant effects on Canada-U.S. trade during the first five years of the agreement. The boost to trade was significantly greater in commodities that were subject to high tariffs before 1988, than those subject to low tariffs. The effects are about twice as great in the long run as in the short run. She estimates that each one percent cumulative reduction in Canadian tariffs is associated with imports from the United States in 1993 that are 6 percent higher than they would otherwise be. Each one percent cumulative reduction in U.S. tariffs is associated with imports from Canada in 1993 that are 12 percent higher than they would otherwise be. Using five-year elasticities, and multiplying by the actual (trade-weighted) reduction in tariffs in the two countries under the agreement, she obtains estimates for the overall impact of the FTA on trade. The implied boost to northbound trade was 18 percent, as of 1993, and to southbound trade 16 percent. In both cases, this was substantially less than the actual increase in trade over this period. Overall, the estimated increase in trade that can be attributed directly to the FTA was a little more than half of the

actual increase in trade. Much of the rest is presumably due to the increases in incomes in the two countries over the period (and to other elements of the gradual worldwide increase in trade).

One moral of this exercise is that one cannot simply look at the actual increase in trade over the period, and attribute it to the effects of an agreement. This lesson is another illustration of the importance of adjusting for other determinants, such as income, and supports our approach. Another moral, however, is that some effects of regional trading arrangements are lost in tests like ours on highly aggregated data. The problem is particularly severe in a context, such as the Canadian-U.S. FTA, where the small number of observations in the case of aggregated data makes sheer number of observations a constraint. But Clausing's finding that the effects differ widely, depending in particular on the initial level of tariffs and the speed of phase-out, also illustrates the deeper perils of aggregation.

Mercosur

There is a clear upward trend in the bloc effect for Mercosur. The coefficient is not significant during 1965-1975. It becomes higher and statistically significant thereafter, particularly in 1990. In that year it reaches a highly significant 1.9. This is the strongest bloc effect estimated anywhere. It implies that these countries -- Argentina, Brazil, Paraguay and Uruguay -- trade among themselves almost seven times as much as otherwise-similar countries. This is another good example of how the gravity equation can give very different answers than the simple intra-regional trade shares in Table 2.1. Intra-regional trade among the Mercosur countries looked low and slowly-growing, by that measure. But when one takes account of the fact that the countries constitute a small fraction of Gross World Product, and that they did not grow very rapidly in the 1980s, their intra-regional trade looks far more impressive.

A variable for trade between Mercosur and non-member countries shows a monotonically increasing coefficient. It starts out with a highly significant coefficient in 1965 of -.2 or -.3, and rises to a highly significant coefficient of *plus* .8 or .9 in 1990.³¹ This clearly reflects the shift in thinking over time, from import-substitution philosophy to the gospel of the marketplace, in the geographical epicenter of that shift. The recognition of a greater propensity to trade in 1985 and 1990 knocks down the estimated Mercosur bloc effect a little, but it remains statistically significant.

³¹ In addition to Table 4.2 here, see Table 4 of F, W & S (1995).

When the 1970-1992 data are pooled, the Mercosur coefficient is a highly significant .93 (or .71, if one allows for the significant openness effect). This is an effect on trade of 2.5 times.³² Mercosur is another case where a bit of disaggregation seems to help. The bloc effects are even stronger in 1980 and 1985 when the equation is estimated on manufactures alone, and stronger still when estimated on agricultural products alone, than when estimated on all merchandise trade in the aggregate.³³

Our results for Mercosur are consistent with those of Abreu and Bevilaqua (1995, p.26), who find a very highly significant increase in intra-Mercosur trade in 1991-1994.³⁴

The Andean Group

Perhaps the most dramatic turnaround occurs in the countries of the Andes. (The membership of this group excludes Chile, and includes Venezuela). Despite the establishment of the Andean Pact in the 1960s, our estimated coefficient for 1965 is a highly significant -1.3, meaning that these countries actually traded with each other only one-quarter as much as they should have [$\exp(-1.3) = .27$]. The coefficient becomes insignificant, then turns positive, and in 1992 -- after the re-invigoration of the Andean Pact -- it suddenly attains a highly significant 1.0: the Andes countries trade 2.7 times as much as otherwise-similar countries.

³² $2.5 = \exp(.93)$. F & W (1997, Table 1). Given the evolution in South American attitudes toward trade over this period, however, one should probably not take this pooling too seriously, as it imposes an unchanging structure of trade.

³³ F, W & S (1995), Tables 4, 4a and 4b.

³⁴ They estimate a doubling attributable to Mercosur after 1990 [$= \exp(.72)$]. Amjadi, Winters and Yeats (1995, p.17) also find, using intensity coefficients, a very rapid 1991-1993 increase in Argentina and Brazil's trade within Mercosur.

An Andes openness term shows a fluctuating and insignificant coefficient in Table 4.2. In other estimates it has exhibited an upward tendency suggestive of the aforementioned spread of the liberalization gospel in South America toward the end of the 1980s.³⁵ The openness terms have little effect on the Andean bloc effect, which remains large, significant, and negative at the beginning of the sample period, 1965, and large, significant, and positive at the end of the sample period, 1992. Disaggregation helps a little. The bloc coefficient in 1985, for example, becomes statistically significant when tested on manufactures alone.³⁶

When the 1970-1992 data are pooled, the bloc effect is insignificant. Clearly, one does not want to mix the 1960s and 1970s, when the effect was if anything negative, with the later years. A test for the effect of the reinvigoration of the Andean Pact in 1990 on the change in trade among the members produces a point estimate of .1.

Other authors tend to find results weaker than ours. Brada and Mendez (1985) find that the Andean Pact misses being significant, with a point estimate implying a bloc effect of 1.3. Boisso and Farrantino (1995) find neutrality for the Andean Pact overall; the coefficient is insignificantly positive, except for a significant bloc effect in 1971-76. Hamilton and Winters (1992) also find the Andean Pact insignificant for the period tested (1985-86).

Central American Common Market and Caribbean FTA

We do not have Central American or Caribbean countries in our sample. Some relevant gravity equation results are available from others, however. Brada and Mendez (1985) find a

³⁵ F & W (1996, 1997) and Table 4 of F, W & S (1995).

³⁶ F, W & S (1995), Table 4a.

significant effect of the CACM [a factor of 4]. Primo Braga, Safadi, and Yeats (1994) also find that CACM is statistically significant and positive. Boisso and Farrantino (1995) find a strong positive bloc effect for Central America going as far back as 1961, the first year tested. The bloc effect began a steady decline in 1969, however, until it abruptly lost all significance in 1988. The authors attribute this negative trend to the political conflicts of the 1970s and 1980s in Central America, and to an ineffective CACM.

Boisso and Farrantino find a strong positive bloc effect for the Caribbean FTA, going as far back as 1973, the first year tested. It has no tendency to increase thereafter, however, from which they infer that the Caribbean FTA agreement itself has had no effects at all.

The Free Trade Area of the Americas

Next we consider a potential bloc at the level of the entire Western Hemisphere. There does exist a proto-regional trading arrangement, the Free Trade Area of the Americas, that the region's leaders agreed to form in Miami in December 1994. Thus we could call this bloc an FTAA effect. But even in 1996, a hemispheric FTA is a long way from fruition. It certainly would be stretching things to argue that the businesspeople of the Americas could have foreseen the FTAA in the 1970s and 1980s and acted in anticipation of it during our sample period (except, conceivably in 1992, our last year of data; President Bush had proposed the Enterprise for the Americas Initiative two years earlier, and NAFTA had been negotiated by then.)

Thus we consider the Western Hemisphere effect on a par with the Western Europe and East Asia effects. They are of interest when we consider the question whether there is a trend toward a world of three major continental trading blocs, centered on the United States, Germany, and Japan, respectively. For this question, one need not have in mind only formal preferential trading

arrangements. Such institutions as the Monroe Doctrine, the Organization of American States, and the Alliance for Progress are tangible evidence, if any is needed, that the United States has long exercised a degree of hegemony over its part of the world. More abstractly, one of the questions we will consider in Chapters 8 and 9 is the hypothetical welfare effects if the world broke up into three symmetric blocs. We will need statistical estimates to help pin down some parameter values. Clearly we want the Americas to be one of those three blocs.

In Table 4.3 we see a steady upward trend in the Americas bloc coefficient. It actually starts out at a statistically significant minus 0.3 in 1965. This reflects a general lack of openness on the part of the Latin American countries during the hey-day of the import substitution philosophy. (The negative bloc effect loses its statistical significance when one allows a dummy variable to reflect the lack of openness of the countries in the region.³⁷) But the bloc coefficient turns positive in 1975, and significantly so in 1980. By 1990 the coefficient had become a highly significant .8, indicating a boost to intra-regional trade by a factor of 2.2. The pattern here -- intra-regional biases that are small or even negative in the 1960s, becoming positive and significant in the 1980s -- mirrors the results that we got from Mercosur and the Andean Group.³⁸ When one allows for the low level of openness, the

³⁷ Surprisingly, the lack of openness appears in the Table to continue right up to 1992. The finding of a negative openness effect in the Western Hemisphere, or even one that worsens over time, is repeated in F & W (1997), and Table 2 of F, W & S (1995). Evidently the North American countries, which exhibit less openness than predicted by the gravity variables, dominate the South American countries, which toward the end of the sample period exhibit more openness than would be predicted. In Table 2 of F & W (1995c), on the other hand, one gets a result that is more consistent with what one expects: the Western Hemisphere term has a strong upward trend, from a highly significant -.3 in 1965 to a significant +.2 in 1990. In those estimates, the Americas bloc effect remains significant in four out of six observations.

³⁸ Other estimates show significant coefficients as high as 1.2 in 1990. (Table 2 in F & W, 1995c, which is the one that also shows a strong effect for the English language and a Western Hemisphere openness term that becomes significantly positive in 1985 and 1990.)

bloc coefficient registers .7 in 1991 {and .8 in 1992}.³⁹

When the data from 1970 to 1992 are pooled, in Table 5.1, the Americas bloc effect is a highly significant .4, an effect of 49 percent. When we allow for a below-average level of openness in the region, the coefficient rises to .6. When we allow for a linear trend in the bloc coefficient, it comes out positive and highly significant, as we would expect from looking at the year-by-year results.⁴⁰

³⁹ Appendix Table A4.4a.

⁴⁰ Similar levels and trends in the Western Hemisphere bloc coefficient are evident in F (1993, 1994), F & W (1994b, 1995b) and F, S & W (1993, 1996).

Disaggregation of trade by sector reveals that the bloc effect is stronger and more consistent in manufactures trade than in farm products or other raw materials. The bloc effect in manufactures is particularly strong when one allows for the fact that Western Hemisphere countries have been in general less open than the average, but does not allow for common languages.⁴¹

Other authors' gravity results that are most relevant here pertain to the two historical attempts at comprehensive trading arrangements in the Latin American region. Brada and Mendez (1985) find that pairs of LAFTA countries in fact had significantly less trade than other pairs of countries. Hamilton and Winters (1992) find LAIA significant (1985-86), while Primo Braga, Safadi, and Yeats (1994) find LAFTA is not statistically significant. Boisso and Farrantino (1995) find a negative bloc effect for LAFTA that has turned positive over time. Thus earlier results tend to match our finding that the intra-American bloc effect was zero or negative in the 1960s and 1970s, and turned positive in the 1980s.

5.3: Trading Blocs in East Asia and the Pacific

As we have noted before, formal regional trading arrangements are less developed in the Asia-Pacific region than in Europe or the Americas. Only the FTA between Australia and New Zealand is truly well-advanced, though ASEAN is beginning to get serious as well.

Australia-New Zealand Closer Economic Relationship

⁴¹ F, S & W (1994), Tables 2 and 5; and F & W (1995c), Table 3, respectively. Allowing for linguistic links reduces a bit the Americas bloc effect in manufactures, although the effect for Spanish is no higher than for other languages.

Table 5.2 shows the bilateral CER as highly significant statistically, in every year tested. It has a slight upward trend, reaching 1.7 in 1990 and 1992. This estimate says that the antipodean pair trade 5.5 times as much as an otherwise-similar pair. The openness terms shows that trade by Australia and New Zealand with other partners is consistently low, and significantly so in 1980 and 1990. The openness term has no discernible impact on the bloc coefficient, however.

When the data years 1970, 1980, 1990 and 1992 are pooled, the coefficient is 1.6.⁴² We have dated the cementing of the ANZCER as having taken place in 1983. The test of the effect on the *change* in Australia-New Zealand trade between 1980 and 1985 shows a point estimate that is close to zero. As with the other FTAs tested, it appears that too much information is lost when distance and the other unchanging variables are dropped out. As we have noted, the significance problem is especially compounded in the case of a bilateral agreement, where there is only one data point available.

ASEAN

Many studies of ASEAN have reported that trade creation is small, but these conclusions generally seem not to take into account the incomes of the ASEAN countries.⁴³ Table 5.2 shows ASEAN, alone among the six contemporary FTAs tested, as having a statistically significant apparent intra-regional bias in every year tested, 1965 through 1992. The coefficient estimate in 1992 is 1.8, which also happens to be close to the mean, median, and mode of the yearly estimates. Two ASEAN countries trade six times more than two otherwise-similar countries.

⁴² F & W (1997, Table 1).

⁴³ E.g., DeRosa (1995, 25-28, 34). F & W (1996) gives other references, and extends the gravity analysis to focus on Southeast Asian trade patterns.

We know that Singapore plays an entrepot role: its imports and exports are more than 100 per cent of GNP. It is possible that the apparent intra-ASEAN bias is partly or wholly a reflection of the extreme openness of Singapore. To examine this, we have elsewhere tried adding a Singapore dummy to the regression, representing any bilateral trade involving the city-state. The Singapore dummy does indeed have a positive and very significant coefficient [1.51]. The coefficient on the ASEAN dummy is reduced to 1.40 but remains quantitatively large and statistically significant. This suggests that Singapore's extreme openness does not explain all of the apparent inward bias among the ASEAN countries.⁴⁴

The effect in each year is reduced, a little more if one allows for the fact that the entire group of ASEAN countries are more open than are typical countries at their stage of development, not just Singapore. But the coefficient is still in every year highly significant statistically, equaling 1.1 in 1992. These findings -- that ASEAN countries are significantly more open than predicted by the gravity determinants, but that allowing for this openness only reduces the strong estimated bloc effect by a little -- are confirmed in other tests as well.⁴⁵

When the data from 1970 to 1992 are pooled together, the ASEAN coefficient is 2.0, or 1.3 when allowing for ASEAN openness.⁴⁶ Allowing for a trend in the coefficient shows no evidence of one.⁴⁷ To test the effect of ASEAN on the *change* in trade, there is no one clear date on which to focus. ASEAN negotiated a preferential trading arrangement within its membership in 1977, but

⁴⁴ F & W (1995d).

⁴⁵ F & W (1996, 1997).

⁴⁶ F & W (1997, Table 1).

⁴⁷ Appendix Table A4.3.

serious progress in removal of barriers did not get underway until 1987. As recently as 1989, the fraction of goods eligible for regional preferences was only on the order of 3 per cent. It was not until January 1992 that the members proclaimed plans for an ASEAN Free Trade Area to be implemented by reduction of tariffs and non-tariff barriers in phases from 1993 to 2008.⁴⁸ Thus we choose 1992 as the key date. A test of the *change* in intra-ASEAN trade between 1990 and 1992 shows an insignificant point estimate of .2.⁴⁹

A question like "what is the effect of ASEAN on trade among its members" can change radically, depending what other bloc effects are being tested at the same time. When we test for an East Asian bloc effect simultaneously with an ASEAN effect, the latter disappears completely. We will explore further layers of potential blocs below. If one is interested solely in formal regional arrangements, then one can accept at face value the strong bloc effects for ASEAN reported here. If one considers the larger less formal blocs to be on equal footing a priori, then one will want to accept the verdict of the data that ASEAN has no independent effect: Southeast Asian countries trade a lot with each other simply as an example of the phenomenon that Asian countries trade a lot with each other, not out of any special ASEAN effect.

Wang (1992), Wang and Winters (1991), and Winters and Wang (1994) in gravity tests found the ASEAN dummy to reflect one of the most significant trading areas in the world. They did not include a broader dummy variable for intra-Asian trade.

East Asia Economic Caucus

⁴⁸ References include DeRosa (1993a,b, 1995), Jackson (1991), Panagariya (1994), and Jaggi (1995).

⁴⁹ W & F (1995), Table 1.

We define East Asia to be the membership of the East Asian Economic Caucus, which also coincides generally with the boundaries of the hypothesized yen bloc. As can be read from Table 4.5, the East Asia bloc effect follows a very different pattern than any of the other regional groupings. In the first place, it starts out at a high level, by far the highest of any in the 1960s. The estimate in 1965 is 1.6, implying an effect on intra-Asian trade of 5-fold. In the second place, the coefficient gradually diminishes, though remaining statistically significant throughout. By 1992, it has fallen to .5, for an effect of 65 percent.

As with the other groupings, we want to allow for the possibility that most East Asian countries are open to trade of all sorts. The coefficient on East Asian openness is positive and significant. It appears to rule out any "trade-diversion" effects arising from the existence of the East Asian bloc: these countries trade an estimated 22 per cent more with all parts of the world, other things equal, than do average countries [$\exp(.20)=1.22$]. The addition of the openness dummy often reduces a bit the level and significance of the East Asian bloc dummy. The coefficient is usually still significant statistically. The exceptions are 1991 and 1992, when making allowance for the high level of openness in East Asia deprives the bloc coefficient of its statistical significance. For the other years, 1965-1990, the finding appears to be robust that allowing for a high level of openness among East Asian countries does not eliminate the significance of the extra tendency for them to trade among themselves.⁵⁰

Disaggregation reveals that the bloc effect occurs largely in manufactures trade and (to a lesser extent) non-agricultural raw materials, rather than agricultural products.⁵¹ The East Asia bloc

⁵⁰ Table 5.3, F & W (1995c, 1995d, 1997) and F, W & S (1995).

⁵¹ F, W and S (1995), Tables 4a and 4b; and F, S and W (1994), Table 2.

effect in manufactures remains almost as strong when one adds a term for linguistic links, of which the Chinese language is the strongest, as when one does not.⁵²

When the four 1970-1992 years are pooled, in Table 5.1, the East Asia bloc coefficient is estimated at .9 (.8 with the allowance for openness). When one allows for a trend, it is negative and statistically significant, as one would expect from the yearly results.⁵³

This downward trend is not only a departure from the pattern in Western Europe and the Americas. It is, more importantly, a departure from the popular impression that Japan is rapidly turning East Asia into a Tokyo-centered yen bloc. Recall that this impression is supported by an appeal to the simple statistics on intra-regional trade, which show a rapid increase in East Asia. But we have now learned that all of the increase in intra-Asian trade can be explained by the rapid growth of those economies. (GNP and GNP per capita are the only variables on the gravity list that change over time.) There is nothing left over to be attributed to an intensifying bloc. A typical East Asian country i trades far more with another East Asian country j than it did twenty years ago for much the same reason that countries everywhere in the world trade more with East Asian country j than they did twenty years ago: as a result of rapid growth, it looms much larger in the world economy now.

⁵² Compare Tables 3b and 5, in F, S & W (1994).

⁵³ There are similar results in F (1993, 1994), F, S & W (1996), and F & W (1993a, 1993b, 1994a, 1995a, 1995b, 1995c, 1997). Also Petri (1993) and Saxonhouse (1993).

What about bilateral trade between Asian/Pacific countries and Japan in particular? Like intra-regional trade overall, trade with Japan increased rapidly in the second half of the 1980s. Most of this increase merely reversed a decline in the first half of the 1980s however.⁵⁴ More importantly, the recent trend in bilateral trade between Japan and its neighbors can be readily explained as the natural outcome of the growth in Japanese trade overall and the growth in trade levels attained by other Asian countries overall. Lawrence (1991b) has calculated that, out of the 28 percentage point increase in the market share of Pacific Asian developing countries in Japanese imports from 1985 to 1988, 11 percentage points was attributable to improved competitiveness (as reflected in increased exports from Pacific Asia to worldwide markets), and 18 percentage points was attributable to the commodity mix of these countries' exports. There is no residual to be attributed to Japan's development of special trading relations with other countries in its region.⁵⁵

We confirmed this finding by adding to our gravity model a separate dummy variable for bilateral Asian trade with Japan in particular. It was not even remotely statistically significant in any year, and indeed the point estimate was a small negative number.⁵⁶ Thus there was no evidence that Japan has established or come to dominate a trading bloc in Asia.

It is perhaps surprising that the estimated level of the intra-regional trade bias was higher in

⁵⁴ Petri (1991, 1993).

⁵⁵ The empirical literature on whether Japan is an outlier in its trading patterns, particularly with respect to imports of manufactures, includes Saxonhouse (1989), Noland (1991) and Lawrence (1991a), among others. Eaton and Tamura (1996) find that Japan, along with other East Asian countries, is more open to U.S. exports than would be predicted by the gravity model (but less open to direct investment).

⁵⁶ F (1993, 1994).

East Asia as of the beginning of our sample period than in the other two regions. One possible explanation is that there has historically been a sort of "trading culture" in Asia. To the extent that such a culture exists and can be identified with a particular nation or ethnic group, we find the overseas Chinese to be a more plausible factor than the Japanese. But there are other possible regional effects that may be showing up spuriously as an East Asian bloc, to be considered below.

Before we anoint East Asia a trading bloc we must consider the entrepot effect. We have allowed for the level of openness of East Asian countries in general. But we should also allow for the extraordinarily high level of openness in two East Asian countries: Hong Kong and Singapore. Hong Kong, like Singapore, has levels of imports and exports well in excess of its GNP. They have long functioned as entrepots, transshipping goods on the way to other destinations, and they also import and export a lot of intermediate products. We have tried two kinds of dummy variables, one to represent the trade of these two island city-states with other East Asian countries, to test the hypothesis that they are the true hubs of the East Asian bloc, and another to represent the general openness of Hong Kong and Singapore, that is, their trade with all partners regardless of location. The latter variable is clearly the appropriate one: When the two dummy variables are included simultaneously, it is only the openness variable that is significant. Its highly significant coefficient is about 1, indicating that Hong Kong and Singapore are each more open than other countries by a factor of about 2.7. When the Hong Kong and Singapore variable is used at the same time as the dummy variable for openness of all East Asian countries, the latter remains significant as before. Its point estimate is .25, which for 1980 and 1990 is lower than the East Asian openness coefficient when estimated without the presence of the Hong-Kong-Singapore dummy. The introduction of the dummy does, however, knock down the strength of the East Asia bloc effect in the three years tested, 1980, 1985, and 1990. It is now in the range of .3 to .5, and at best of borderline statistical

significance.⁵⁷

⁵⁷ The years were 1980, 1985, and 1990. F (1993, Table 2.5; 1994).

Continuing the process that began with ASEAN, we consider a sequence of nested candidates for trading blocs in the Pacific.⁵⁸ The significance of a given bloc effect turns out to depend on what other blocs are tested at the same time. One way to draw the boundaries is to include all the countries with eastern coasts on the Pacific, which includes Australia and New Zealand along with East Asia. We call this grouping "Asian Pacific." Its coefficient and significance level are both higher than the East Asia dummy. When we broaden the bloc-search wider and test for an effect of APEC, which includes the United States and Canada in with the others, it is highly significant. The significance of the Asian Pacific dummy completely disappears. The East Asia dummy remains significant, though at a lower level than the initial results that did not consider any wider Pacific groupings.

APEC

APEC appears to be the correct place to draw the boundary. When we test for an even broader definition of a Pacific Rim bloc, including Mexico and the South American countries that border the Pacific (Colombia, Ecuador, Peru and Chile), it is not at all significant, and the other coefficients do not change.⁵⁹ It remains true that the intra-regional biases in the EC and Western Hemisphere blocs each roughly doubled from 1980 to 1990, while intra-regional biases in the Asia and Pacific areas did not increase at all.

⁵⁸ These results are reported in F (1993, Table 2.2-2.4; 1994).

⁵⁹ We count only the core 15 in APEC, excluding Mexico, Chile and Papua New Guinea which were not added until 1993. [CHK]

The surprising finding is the magnitude of the APEC effect. Table 4.3 shows APEC highly significant statistically in most years (when estimated along with the East Asia bloc, which still seems to belong in, but without the other nested groupings, which do not). There is a slight upward trend, reaching 1.2 in 1992: two APEC countries trade 3.3 times as much as two otherwise-similar countries. Introducing a term for openness of members (which, surprisingly, is negative), does little to change the APEC bloc estimate. The strong APEC bloc effect shows up under various permutations of the specification.⁶⁰

When the years 1970, 1980, 1990, and 1992 are pooled, in Table 5.1, the coefficient on the APEC bloc effect is again a highly significant 1.1. When we allow for a trend, it turns out to be positive and significant.⁶¹ The introduction of dummy variables for openness of the various groupings reduces the estimated APEC bloc effect only slightly, though it eliminates the significant upward trend in that coefficient over the 1970-1992 period. The introduction of a dummy variable to represent the extra openness of Hong Kong and Singapore does not reduce the strength of the APEC bloc effect.⁶²

Others have emphasized the high volume of trans-Pacific trade. But it has been difficult to evaluate such statistics when no account is taken of these countries' collective size. As noted in Chapter 3, a higher percentage of economic activity will consist of intra-regional trade in a larger region than in a smaller region, even when there is no intra-regional bias, merely because smaller regions tend by their nature to trade across their boundaries more than larger ones. In the limit,

⁶⁰ Including F & W (1997), F, W & S (1995), and F & W (1995c). In the first two studies, APEC members are again estimated to be significantly less open than predicted by gravity determinants. The third study shows the more expected result, where APEC members go from low openness to high openness during the course of the sample period. In all cases, the APEC bloc effect is high and significant.

⁶¹ We found similar results in F & W (1993a, 1993b) and F, W & S (1995).

⁶² F (1993, 1994).

when the unit is the world, 100 per cent of trade is intra-"regional." Our gravity results show that the APEC effect is genuine. Thus the United States and Canada appear to be full partners in the strongest grouping in the world, the Pacific bloc, even while simultaneously belonging to the significant but distinct Western Hemisphere bloc.

As with previous groupings, we have estimated the APEC effect on disaggregated trade data as well. The bloc effect is always highly significant in agriculture and other raw materials. Trade in manufactured goods shows an APEC bloc effect that is statistically insignificant in 1965, borderline-significant in 1970, and only becomes significant at the 99 percent confidence level from 1975 on. Thus the upward trend that one would have expected from the APEC effect is concentrated in manufactures.⁶³ The APEC bloc effect in manufactures is attenuated slightly throughout the period, if one allows both for the fact that APEC countries have a significantly lower-than-average level of openness in manufactures and one allows for the effect of linguistic links. Of the linguistic links, Chinese is the strongest, and English is the second strongest; both languages are relevant for the Pacific. The bloc effect does not turn positive until 1970, does not turn significant until 1975, and does not turn significant at the 99 per cent confidence level until 1980.⁶⁴ Even so, the bloc effect in manufactures toward the end of the sample period is as strong for APEC as for any grouping. In agriculture and other raw materials, it is the strongest.

A number of readers have found the bloc effect that we estimate for APEC surprisingly

⁶³ F, W & S (1995), Table 4a; and F, S & W (1994), Table 2. If one neglects to correct for a lower-than-average level of openness with respect to manufactures among Pacific countries, the bloc effect in APEC manufactures is highly significant in 1965 and 1970 as well; F, S & W (1994), Table 3b. This is also without the common language dummy.

⁶⁴ F & W (1995c), Table 3. If one allows for the languages, but not for the low level of openness, then the APEC bloc effect in manufactures again does not take hold in a significant way until the 1970s. (F, S & W (1994), Table 5).

high.⁶⁵ We must admit to a bit of the same reaction ourselves, though this does not necessarily mean the result is wrong. One will never learn anything from statistics if one always throws out any results that differ from one's preconceptions. But it is worth trying to think of alternative explanations.

One explanation that appears plausible is that the costs of transporting across and around the Pacific Ocean might be overstated by our distance measure. If this were true, then the estimate of the APEC coefficient would be biased upward: the amount of trans-Pacific trade would look surprisingly high because we would have exaggerated the cost of shipping across the ocean. Recall that our variable, the log of distance, already captures the declining marginal cost of transportation with respect to distance. One possibility is that the log specification does not allow the marginal cost to fall off quickly enough. But the other functional forms we tried made little difference.

⁶⁵ E.g., Polak (1996).

A more likely possibility is that the cost of ocean transport is less than the cost of land transport. Pacific trade consists disproportionately of ocean transport, while shipping among European or Western Hemisphere countries is more often overland, whether by rail or road. As already noted, we obtained data that allowed a distinction for land versus sea travel, from Winters and Wang. The APEC coefficient did indeed decline some when we estimated the equation with the Winters-Wang measure. The strong effect that had been estimated for 1980 fell only modestly, from 1.3 to .9.⁶⁶ A weaker 1990 bloc effect that had been estimated for 1990 in that study (possibly due somehow to the allowance for trade diversion effects), loses its significance altogether with the Winters-Wang measure. The APEC point estimate drops from .55 to .32. The coefficient estimates on the East Asia, EC, and Western Hemisphere blocs are undiminished by the use of the Winters-Wang distance measure for 1990. In any case, the issue of water versus land transport should not affect results regarding changes in intra-regional trade bias in the 1980s, however, given that the nature of shipping costs does not appear to have changed over as short a time span as five or ten years.

If one believes that distance has a smaller effect for Pacific countries' trade than for others' (e.g., shipping costs are lower per kilometer, even after holding constant for per capita GNP and the other variables), one might want to allow them to have different distance coefficients. Dhar and Panagariya (1995, Table 1A) allow each exporting country to have its own distance coefficient. The average coefficient for the 13 APEC members is 1.53 (-1.74 for ten East Asian members, including Australia, and -.82 for the three North American members). By comparison, the other nine countries that are estimated show an average distance coefficient of .92 (.67 for six European countries and 1.43 for three other LDCs). This piece of evidence tends to contradict the hypothesis

⁶⁶ F, W & S (1995), Tables 2 and 3. For the 1980 results, the EC bloc effect loses its previous borderline-significance when the Winters-Wang distance measure is introduced.

that distance brings lower costs for APEC members than for other countries.

The finding that intra-APEC trade is high relative to the natural benchmark appears to be a correct conclusion. It supports the APEC enthusiasts, who describe trade within the group as an example of *market integration*, "where the initiative has remained primarily with enterprises acting separately from state decisions, and where official encouragement of regional integration does not include major elements of trade discrimination"⁶⁷. This is by contrast with *institutional integration*, which we have referred to as formal or explicit regional policies.

5.4 Trading Blocs Among Other Less Developed Countries

In all the excitement about a possible system of three trading blocs, Europe, Americas, and East Asia, it is easy to forget that a large fraction of the planet's population lies outside these three groupings. Here we consider the rest of Asia, and then Africa.

South Asia

Seven countries of the Indian subcontinent formed the South Asian Association for Regional Co-operation (SAARC) in 1985. (The members are India, Pakistan, Bangladesh, Sri Lanka, Nepal, Bhutan and the Maldives.) Their past talks had been even more fruitless than ASEAN's. In May 1995, however, the members agreed to put a preferential trading arrangement into place in December. How much substance there will be in this PTA remains to be seen.

In our sample, the term "South Asia" refers to only two countries, India and Pakistan. One conjectures that the trade between these two countries is negatively impacted by their historical animosity. Estimates show that this is indeed the case: their trade is 70% lower than two otherwise

⁶⁷ Garnaut (1993, p.5). Drysdale and Garnaut (1993) borrow the terms from Cooper (1974).

identical economies.⁶⁸

Unfortunately Bangladesh, Sri Lanka, and Nepal are not in our sample of available bilateral trade data. But Srinivasan and Canonero (1995, p.29) do have data on trade between these countries and other major trading partners. They note that Bangladesh and Sri Lanka trade very little within the South Asia region. (Much of Nepal's trade is with India; but then Nepal has few alternative routes to the outside world.) The no-bloc finding that we obtained for South Asia might well generalize, even if all the members were represented.

Pan-Asian groupings

⁶⁸ F & W (1995d). Srinivasan and Canonero (1995, p.29) also find a negative effect for India-Pakistan trade, as do Dhar and Panagariya (1995, p.12-13). India-Pakistan is probably undercounted however, by some \$2 billion according to *The Economist* (Jan. 27, 1996, p.36); it is estimated that half this amount is smuggled across the border, and half goes by third-country ports.

South Asians wonder if they should not be included in Asia. The habit of speaking of Asia-East-of-Burma as a separate region called "East Asia," almost as a separate continent, has not always prevailed. It has become standard only in the last few decades, largely in response to the superior growth performance of most of these countries.⁶⁹ We have tried treating South and East Asia collectively as one candidate trading group. The coefficient for the East-and-South Asia group is 0.65 and significant, indicating that two countries in this group trade 90% [$=\exp(0.65)-1$] more than a random pair of otherwise identical countries. Given the aforementioned India-Pakistan finding, the positive coefficient on the East-and-South Asia bloc must reflect higher-than-average trade between East and South Asian economies.⁷⁰ If we add the Hong Kong and Singapore dummies to the regression, the coefficients on East Asia and East-and-South Asia dummies remain quantitatively large (1.36 and 0.37, respectively) and statistically significant.⁷¹

⁶⁹ Easterly (1993) and Easterly, Kremer, Pritchett, and Summers (1993) see the drawing of the line that separates East Asia from the rest of Asia as having been endogenous.

⁷⁰ Dhar and Panagariya (1995), however, find a negative effect for India-China trade, where there has been political and military animosity in the past.

⁷¹ F & W (1995d).

The next question to arise is whether the right place to draw the line dividing up Asia, if not between Burma and Thailand, is between Pakistan and Iran, or whether the proper grouping instead goes all the way west to Turkey. We include in the regression the whole of the continent of Asia (i.e., adding Asian countries in the Middle East to the above list) as a potential bloc. Two results are noteworthy. First, East Asian economies continue to show certain inward bias among themselves. Second, even after controlling for a special East Asia effect, Asian economies as a group appear to trade more among themselves than one would expect based on their economic and geographic characteristics. There is no reason to draw a line between South Asia and the Middle East. Part of the pan-Asia trade concentration undoubtedly has to do with the fact that many Asian economies have to import a substantial amount of oil from the Middle East.⁷²

Africa

There are fewer applications of the gravity model to Africa than to other continents. One is Foroutan and Pritchett (1992). They find, contrary to the usual hypothesis that trade among sub-Saharan African countries is artificially low, that intra-African trade is somewhat higher than one would expect, given the poverty of the countries and other standard determinants. This is consistent with what we found in Chapter 2 {previously 3}, in the intensity coefficients. Wang and Winters (1991) and Winters and Wang (1994) find the effects of two regional arrangements, ECOWAS and the South African Developing Coordination group, to be statistically insignificant however.

5.5 : Are These Regional Trading Arrangements Trade-Diverting?

The gravity equation is useful for estimating magnitudes of trade creation and trade diversion.

⁷² Adding the Hong Kong and Singapore dummies does not change the qualitative features of the picture. F & W (1995d).

When a PTA is created or expanded, one cannot base conclusions on simple observed changes in trade flows, without holding constant for changes in the countries' incomes over the same period. The bloc effects that we have already discussed can be interpreted as trade creation. In this section, we focus on trade diversion: the apparent effects of the regional arrangements on trade with non-members.

Diversion by European Groupings

There has long been a fear among non-Europeans that regional integration on the continent would create a *Fortress Europe*, which raised barriers against imports from non-members. We begin by considering some results from studies by others.

Typical of studies of the earlier stages of regional integration in Europe is research by Kreinen (1972, 1982) on the formation of the European Community and on its enlargement, which found relatively little trade diversion, with trade creation larger by a multiple of five to seven times.

Gundlach, et al (1993, p. 212-219) summarize some recent studies, such as Davenport (1991) and Page (1992), of the effect of the 1992 Single Market on less developed countries. Within the category of primary commodities, very little trade diversion is expected, because EC countries do not produce primary commodities or close substitutes. Even for the case of manufactured products, the authors themselves [Gundlach, et al (1993, p. 218)] are relatively optimistic. They argue that the Single Market is likely to open up substantial new export opportunities that outweigh trade-diversion.

The 1995 enlargement of the EU -- to take in Austria, Finland and Sweden, formerly members of the European Free Trade Area (EFTA) -- could impact skill-intensive Japan, the United States and Canada, while another future enlargement to include the poorer Czech and Slovak Republics, Poland, and Hungary, should be felt more by the labor-intensive developing countries. This would be a repeat of the earlier assimilation of Spain, Portugal and Greece into what was then

the 9-member European Economic Community.

The studies of the external effects of EC 1992, cited above, emulate some earlier studies of the effects on income within Europe, in the respect that they allow dynamic effects on European growth. This approach tends to yield a rosier outlook for everyone. The dynamic effects, in contrast to earlier static (and generally small) estimates, are maximized under the assumption that the investment rate will be stimulated. The classic references, the Cecchini Report (1988) and Baldwin (1989), estimated that EC GNP by the end of the century would go up on the order of 2.5 to 6.5 per cent as the result of the 1992 Single Market. This higher European income would raise imports from all trading partners. If the elasticity of import demand is about 2, then exports from non-members would go up at least 5 per cent. This effect is to be netted against the negative effects of trade diversion. The grounds for the dynamic estimates are unusually uncertain however.

We take GNPs as given in our analysis. To the extent that incomes are in fact endogenously determined by trade, our estimates understate the benefits of PTAs on outsiders.

The results for Europe from our gravity estimation are mixed. On the one hand, as we noted above, when the EC Twelve are considered as a group, they generally appear open to trade with non-members,⁷³ and that openness appears to increase over time.⁷⁴ On the other hand, specific tests of the effect of the expansion of the EC Six to Nine in 1973 and the expansion of the Nine to Twelve in the early 1980s, find negative effects in the change in trade between existing members and the rest of the world.⁷⁵ At the end of Chapter 10, our goal will be to evaluate the long-term tendency for regionalizing groupings to liberalize with respect to others, rather than to quantify the immediate

⁷³ Table A4.2 and F and W (1996, 1997).

⁷⁴ F & W (1995c) and F, W & S (1995). Not, however, in Table A5.2 .

⁷⁵ W & F (1995). The effects are estimated at -0.1 (borderline-significant) and -.3 (highly significant), respectively.

static trade-diversion effects of a specific regional initiative that holds tariffs toward non-members held constant. Thus we may want to take encouragement from the positive evidence on the long-run tendency toward openness among EC countries, as opposed to the negative evidence of short-term trade-diversion.

In the case of EFTA, the dummy variable for trade with non-members always shows a negative effect. This is the tendency under all three approaches to the test: the coefficient on average, the trend in the coefficient, and the estimated effect of the expansion of the EFTA from four to five in 1970.⁷⁶ If the question is the effect of formal PTAs, then the results from the broader groupings that were not operative during the sample period are less relevant. For the record, however, tests for the European Union Fifteen or for all of Western Europe, tend more often toward a low degree of openness than a high degree. Examples are Tables 4.2 and 4.3, reported earlier. Even though Western European countries engage in more trade than the average countries in our sample, this can be explained by the high levels of income, geographical proximity to other high-income countries, and so forth. The residual to shows a lower level of openness than one would expect. The same is true of the EU 15. Evidently it is the three new members of the EU, Austria, Finland and Sweden, that gave the Fifteen a negative openness coefficient in the preceding section.

Diversion by Western Hemisphere Groupings

Studies of the Canadian-U.S. FTA (CUSFTA) feared a decline in trade with third countries in general.⁷⁷ However, Clausing's (1995) careful analysis of the Canadian and U.S. data, disaggregated by sector, finds no evidence of a negative effect on imports from other countries.

⁷⁶ F & W (1995c, 1996), W & F (1997); F, W & S (1995); and W & F (1995), respectively. Trade with non-members showed an estimated coefficient on the expansion to include Iceland of -0.2.

⁷⁷ E.g., Harris and Cox (1984) and Primo Braga (1994).

It is too soon to tell the actual effects of NAFTA, but some authors have predicted likely effects. Studies such as Hufbauer and Schott (1993) predict a quite small amount of trade diversion. In part, this is because U.S. tariffs were already very low to begin with (and were already slightly lower against some Mexican goods than against imports from other industrialized countries, due to various U.S. tariff preference programs such as the Generalized System of Preferences and the maquiladora program.

The major barriers remaining in the United States (as in other industrialized countries) are not tariffs, however, but Non-Tariff Barriers (NTBs) and administrative protection (such as Anti-Dumping Duties). Canada and Mexico are to a greater extent buffered from U.S. NTBs under the NAFTA, by dispute-settlement provisions. Indeed this was the major objective in pursuing the FTA, from their viewpoint. Thus other countries' concerns about diversion of trade are still quite relevant.⁷⁸ Mexico ran a large and growing trade deficit in 1993 and 1994, which was initially perceived as beneficial for its trading partners.

Following the peso crisis that broke in December 1994, the Mexican deficit disappeared in 1995. The private capital flows to support it had begun to dry up after February 1994, and the central bank's reserves virtually ran out in December. The unexpectedly large peso devaluation that took place at that time, together with the subsequent sharp recession in Mexico, were the principal instruments of this adjustment. The United States and Canada experienced in 1995 a decline in Mexican demand for their imports. Americans should not, however, view NAFTA as having been a mistake in light of the Mexican crisis. To the contrary. U.S. exports to Mexico were still higher in 1995, the year after the crisis, than they were in 1993, before NAFTA (by 11 percent). U.S. imports from Mexico rose even more. But it has never been a good idea for the United States to judge the advantages of regional trading arrangements by the impact on the trade balance. The standard

⁷⁸ Krueger (1993).

legitimate argument in favor of NAFTA, that it helped to lock in the recent beneficial Mexican economic reforms, including trade liberalization, has already shown its virtue in this crisis. In contrast to the 1982 debt crisis, when Mexico raised import barriers – tariffs and licensing requirements – in an effort to obtain the foreign exchange to service its debts, it did not raise trade barriers against the United States and Canada after the 1994 crisis. For the reason the decline in Mexican imports from the United States from 1994 to 1995 (9 percent) was less than the decline in Mexican imports from Japan and Europe, and much less than the decline at the time of the earlier crisis (a 50 per cent fall in imports from the United States from 1981 to 1983).

Next come the potential trade-diversion effects of an enlarged Free Trade Area of the Americas. Anticipating the Miami Summit, Hufbauer and Schott (1994, p.163-64) made estimates of the effects of a hemisphere-wide FTA. They calculated by commodity groups how much of the increased U.S. imports from the rest of the hemisphere would represent diversion of trade that would otherwise come from other countries. The resulting numbers, while calculated to be somewhat biased upward, are still small. One reason already noted is that U.S. tariff barriers are already low, and will be even lower after the Uruguay Round: below 3 per cent by the year 2000. (The estimates do not include the loss of exports to Latin America.) If tariffs in Latin America were currently as high as they were ten years ago, the trade-diversion in these markets might be substantial. But tariffs in these countries have already come down a lot, and will probably come down a lot more. This fact, together with the fact that the Latin American market is not as large as the United States, implies that trade diversion should not be that large.

For the Canadian-U.S. pair, we find a negative openness coefficient in 1970, turning to a positive one in 1992.⁷⁹ For NAFTA, we find a negative openness coefficient throughout, though it

⁷⁹ F & W (1996).

must be repeated that the sample period precedes the date that the agreement took effect.⁸⁰ Mercosur often shows a negative openness effect early in the sample period, but a positive openness effect late in the sample, particularly in 1990, when the customs union was actually being negotiated.⁸¹ The Andean Group shows a similar mix of estimated coefficients, looking encouragingly open in some estimates, but less so in others.⁸² In the cases of these two South American groupings, the most encouraging results come in the tests for the changes in trade with non-members. Openness effects of +0.2 after 1991 for Mercosur and +0.4 after 1990 for the Andean Group are both significant.⁸³ When examining the Americas in the aggregate, however, the fairly consistent finding is a low level of openness.⁸⁴

Diversion by East Asia and Pacific Groupings

⁸⁰ Table 4.2 [here; W & F (1997); and F, W & S (1995).

⁸¹ Table 4.2 here; W & F (1997); and F, W & S (1995).

⁸² F & W (1996), F, W & S (1995), and W & F (1997); versus Table 4.2 [.

⁸³ W & F (1995).

⁸⁴ W & F (1997), F, W & S (1995), and Tables 4.3 and 5.1 [.

East Asian groupings show up as the most open to trade with the rest of the world. ASEAN shows little or no evidence of trade diversion. To the contrary, given their stage of development, the ASEAN countries consistently show a level of openness that is higher than for other countries in the sample.⁸⁵ If 1990 is taken as the key date for ASEAN, the estimated effect on the change in trade with non-members is also positive.⁸⁶ The same openness is revealed for the broader grouping of East Asia.⁸⁷

The results for the Australia-New Zealand CER are mixed. When estimated for individual years, the evidence points to trade-diversion; when the data are pooled, the opposite sign emerges, suggesting a positive effect on trade with other countries. A test of the effect on the change in trade after 1983 also shows a positive coefficient.⁸⁸ The results for APEC are also mixed. One set of estimates shows a pattern whereby the coefficient progresses from significantly negative in 1965 to significantly positive in 1990.⁸⁹ Others show coefficients that are more consistently negative.⁹⁰ It is important to realize that these estimates already hold constant for the openness of East Asian (and Western Hemisphere) countries. Thus the additional APEC effect largely reflects the level of openness of Australia and New Zealand.

The results regarding trading blocs are easily summarized. We have seen that the specifics

⁸⁵ Table 4.2 [, F & W (1996), and W & F (1997).

⁸⁶ W & F (1995).

⁸⁷ Table 4.3 , F (1994), F & W (1995c, 1995d), F, W & S (1995), and W & F (1995, 1997). These results are similar to those of Dhar and Panagariya (1995), who use the gravity model to find that East Asian countries are open with respect to outside countries, contrary to the usual view.

⁸⁸ Table 4.2 [, W & F (1997); and W & F (1995), respectively.

⁸⁹ F & W (1995c).

⁹⁰ Table 4.3 [, F, W & S (1995), and W & F (1997).

of the relative magnitudes and trends of the bloc effects in different parts of the world are different from what one would conclude by looking at the simple statistics on intra-regional trade shares. Nevertheless, the overall conclusion is the same. Even after holding constant for such natural determinants of bilateral trade as size and distance, intra-regional concentrations of trade are appearing in various parts of the world.

Looking at explicit preferential trading arrangements, we find the strongest effects for ASEAN and the Australia-New Zealand CER: each apparently serves to boost trade among its members by an estimated five-fold or more (though the magnitude of the ASEAN effect diminishes somewhat when one makes allowance for the special role of Singapore). We also find very strong effects for the Andes Pact and Mercosur in the 1990s: each shows an estimated effect on trade of roughly 2 1/2 times, in 1992. The EC effect becomes statistically significant from 1985. Estimates suggest an effect on intra-EC trade of about 65 per cent, with the membership expansions of 1973 and 1983 each boosting trade by about half that amount. There is not yet enough data for us to have obtained meaningful estimates of the effects of North American regional trading arrangements, though others are beginning to do so.

If we wish to look at broader regional groupings, as opposed to formal preferential trading arrangements, we again find statistically significant effects. Western Europe is generally statistically significant throughout the period. The Western Hemisphere effect becomes significant and large in 1990. The Asia effect appears even more significant, and is equally large from the beginning of the sample (1965). Unlike the European and American groupings, there is no evidence of an upward trend. When allowance is made for the special role of Hong Kong and Singapore, or for the openness of the entire East Asian region, the magnitude of the Asian bloc effect is reduced. APEC appears to have a very strong bloc effect as well, which so far has been robust to various attempts to dislodge it.

We should not expect always to be able to tell from the data whether true concentration effects are coming from the formal RTAs or from the broader geographic groupings, given the high correlation in membership. Some readers will prefer to make the choice on a priori grounds. But in the case of Europe, the true effect appears to be coming from the leading RTA on the continent, the European Community, while in the case of Pacific Asia it appears to be coming from the larger groupings (East Asia and APEC), rather than from ASEAN.

None of these results is claimed to be definitive. We will want to see how they stand up to additional possible perturbations and extensions of the analysis.

VI . EXTENSIONS OF THE EMPIRICAL ANALYSIS

The gravity equation that we have estimated leaves out many factors. If an omitted factor is correlated with the existence of a regional trading arrangement, distance, or one of the other variables that we are interested in, we will get biased coefficient estimates. Thus it is important that we take into account as many other factors as we reasonably can. It is inevitable, of course, that we cannot take into account everything. Often the work of others will help us learn what we want to know. We will see that the basic bottom line is unchanged, that statistically significant bloc effects are arising in many parts of the world.

6.1 : The Importance of Political Factors in Trade

An interesting trivia question illustrates the importance of historical factors in trade: Who is Zaire's biggest trading partner? Criteria of proximity or size will give the wrong answer. The country with which the former Belgian Congo conducts the most trade, even in absolute terms, is Belgium.

We have already captured some of this effect, in that our language variable counts the links between former colonies and colonizers (although Zaire is missing from the set of countries considered in our estimates, for lack of data). The colonial link was severed in 1960. Even when the original reason for a high level of bilateral trade has disappeared however, the stock of capital that firms have invested in the form of marketing and distribution networks, brand-name loyalty among customers, and so forth, lives on for many years thereafter. The word *hysteresis* is sometimes applied to this phenomenon, suggesting that the effect is considered to be permanent.

Political alliances and enmities

We noted earlier that trade between India and Pakistan is impeded by their historical animosity. Their trade is 70% lower than trade between two otherwise identical economies. We also noted in Chapter 2 that Bhagwati (1992, 1993a) is suspicious of the claim that proximity is an important determinant of trade, citing the example of India and Pakistan, and is therefore skeptical of the notion of natural trading blocs. He asserts that the high levels of intra-regional trade that are already observed in such areas as Europe must be the result of FTAs and other preferential trading arrangements that are already in place.¹ The issue becomes an important one for policy when other economists, such as Krugman and Summers, argue that proximity does promote trade, and propose that regional trading arrangements be pursued on the grounds that it is *natural* for neighbors to trade

¹ Panagariya (1995, pp. 9-10) echoes Bhagwati's suspicions. He attacks Summers' argument that an FTA among natural bloc partners is less likely to be trade-diverting, with "natural" defined by a low level of trade with countries outside the group. To the extent that a low observed level of trade reflects natural barriers, such as distance, our results support Summers and Krugman. Bhagwati and Panagariya (1995) are correct to point out, however, that high bilateral trade volume alone does not imply a natural partnership, if it exists for reasons *other* than low transportation costs or geographical proximity. Summers (1991) seems to imply otherwise. The proposition that one criterion for a trade-creating welfare-improving FTA is that a high proportion of trade be conducted among partners *ex ante* goes back to Lipsey (1960). But it seems evident that, to the extent such trade is the result of preferences that are already in place, it cannot be used to justify further preferences. These issues are addressed more formally in the next chapter.

with each other. (In Chapter 8, we derive the idea of an *optimal degree* of regionalization that can be justified by natural geographic factors.)

The gravity equation estimates convinced many of us some time ago that distance is in fact a very important determinant of trade. But special historical attractions or repulsions also matter, independently of distance. In South Asia (and the Mideast²), it is in fact true that neighbors do not necessarily trade more with each other. Historical enmity has indeed reduced trade between India and Pakistan. Perhaps the root of Jagdish Bhagwati's skepticism regarding the role of proximity in trade is that he has been heavily influenced by this one observation.

To repeat, the gravity model clearly shows that proximity is in general an important determinant of bilateral trade around the world, notwithstanding exceptions like India-Pakistan, Iran-Iraq, Israel-Syria, and other cases. Ideally, one would use a dummy variable to represent all pairs with a recent history of strong political or military conflict, especially including embargoes and boycotts. This variable would in essence be the antithesis of the dummy variable for linguistic and colonial links. The distance and adjacency effects are so strong however, that they show up as highly significant statistically even when no account is taken of the antagonist pairs.

The effects on bilateral trade of politico-military alliances, wars, colonial relationships, and so forth, have been extensively examined by Mansfield (1993), Mansfield and Bronson (1994, 1997), and Gowa and Mansfield (1991). Theoretically and empirically (in the gravity framework) they find, as one might expect, that trade is generally higher among countries that are allies, and lower among countries that are actual or potential adversaries. If two countries are currently at war, there is usually a negative effect on trade. It runs as high as a highly significant 99 per cent reduction [=exp(-4.90)] in 1965. More typical is a highly significant 82 percent reduction [=exp(-1.74)] in 1990.

The significance of distance and the other gravity variables tends to be unchanged. The log

² Fischer (1993, pp. 434-436).

of distance still has a negative and highly significant effect in every cross-section, always in the range from -0.51 to -0.69 (estimated at five-year intervals, from 1950 to 1990). The formation of a preferential trading arrangement is found to promote trade between two countries by about 100 percent if they were not already military allies, and by about 240 percent if they were. (The effects are different if the alliance involves either of two major powers: the Soviet Union and China.) Once again we see that *both* proximity and existing preferences are important determinants of trade patterns.

The Example of Canadian provinces

We noted in Chapter 4 that there would be some great advantages of having data at the level of states or provinces within countries. We would be able to ascertain how trade between two geographical entities is affected by their common membership in a political union. It would, for example, help us predict increases in the volume of trade among the members of the European Union, in the event that they achieve full political federation. We have learned that when two geographical units share such links as speaking a common language, their bilateral trade is clearly boosted. It stands to reason that when two units share a common cultural heritage or legal system, their trade will be enhanced by even more.

Data are not generally available on trade among U.S. states, Japanese prefectures, German lander, British counties, or French departments. But there do exist data on trade undertaken by Canadian provinces, among each other and with major American states. They show a strong intra-national bias to trade. Ontario exports three times as much to British Columbia as to California, even though the latter has ten times as many people. (The figures are for 1988.)

McCallum (1995) has applied the gravity model to trade among the provinces and states. The usual effects of size and distance show up. The fascinating result is the effect of a dummy

variable to represent when two states or provinces lie in the same country. Two such provinces trade 22 times as much with each other as would a province and a state that are otherwise similar but lie on opposite sides of the border. Helliwell (1995) has updated this test. He finds that the intra-Canadian bias factor is 20.5 in 1988, 18.5 in 1989, and 24.8 in 1990. There is not enough data to read much meaning into the trend; one should settle for an average estimate of 21.³ These results do not hold constant for provinces' adjacency, per capita incomes, or remoteness. It is possible that their omission explains some of the intra-country bias, though this cannot be much of the explanation.

³ The Canadian bias for Quebec trade is even higher than for other provinces, at 26.0. This comes as a surprise, in light of Quebec's insistence on the separateness of its culture. The explanation is that, even though Quebec's trade with other Canadian provinces is, as one expects, estimated to be lower than the gravity variables predict, Quebec's trade with American states is lower still. Quarrelsome Quebec is even more separate from the United States than it is from the rest of Canada.

The result is reminiscent of the striking finding, in the paper "How Wide is the Border?" by Engel and Rogers (1994), that crossing the Canadian-U.S. border adds as much to the relative price variability between two cities as does traversing a physical distance of 2500 miles within either country. This tendency for Canadian provinces to trade with each other is all the more surprising because they tend to maintain trade barriers against each other, never having had the advantage of a Constitution like the one in the United States that reserves trade policy exclusively for the federal level. Reasons for the intra-Canadian bias in trade include the ease of doing business within the same legal system, an integrated media and advertising sector, nationwide store chains, and an East-West railroad network.⁴ Helliwell and McCallum (1995) "suspect that the answers lie in a whole host of educational and geographic ties based on migration and family ties and supported by networks of transportation, communication and education, along with portability of health care and pension rights -- if not completely of beer." Presumably the sources of intra-national bias are even stronger for other countries that do not share the cultural proximity and liberalized trade relations of Canada and the United States.

This experiment is an instructive one for thinking about the likely effects of political integration or disintegration in other parts of the world. The Canadian federation of provinces or American federation of states provide possible models for the European Union. The very high effects of political union estimated in the Canadian-U.S. case tell us that trade among European countries would increase dramatically (21-fold) if the EU attained the same degree of political integration that Canada and the United States have each achieved within their borders. (Our results

⁴ According to Rousslang and To (1993), domestic costs of transporting, wholesaling, and retailing foreign-produced goods are in general so much greater than the analogous costs for domestic-produced goods, as to add an estimated 12.7 per cent to the total cost. By comparison, they estimate the barriers imposed by tariffs and international transportation costs combined at only 7.2 per cent.

indicate that as European countries continue to speak mostly different languages, intra-European trade will fall a bit short of intra-U.S. trade. Canada is a slightly closer standard of comparison in this regard.) At the same time, however, the exercise tells us that the European Union has a very long way to go before attaining that degree of integration. The formation of a common market turns out to be a relatively small step, by comparison.

Four historical examples of the break-up of federations

Historical examples of political unions that split apart offer further insights into the effect of federation on trade, and into the lags involved after a major change in such political ties.

The Austro-Hungarian Empire

One interesting example is the breakup of the Hapsburgs' Austro-Hungarian Empire at the end of World War I. Pre-dissolution data are of limited availability. (This is also the problem even with some more recent examples, such as the separation of Pakistan and Bangladesh or the breakup of Yugoslavia.) A League of Nations study compared 1924 data on trade between Hungary and Austria-Czechoslovakia (combined -- i.e., the former Kingdom of Austria), with 1913 data on trade between the corresponding regions.⁵ Its findings imply that if the Austro-Hungarian union had been preserved, trade flows would have been 2 1/2 times what they actually were in 1924. A more recent study by de Menil and Maurel (1994) applies the gravity methodology, and finds that, five years after the break-up of the Empire, trade among Austria, Hungary and Czechoslovakia was still almost four times as large as what would be expected from the fundamental determinants. They conjecture that the pre-war effects of the Austro-Hungarian Empire on trade must have been similar to the estimated

⁵ Layton and Rist (1925).

effects of the British Empire, and on this basis concur with the League of Nations finding that the breakup reduced trade by more than half. The two statistics together imply a total Austro-Hungarian pre-war effect on trade of about eight-fold. The authors also remark that the partial persistence of the effect into the 1920s must reflect "a network of business and personal relations that it takes time to build, and that does not decay instantaneously." However, the concentration in trade among the constituents of the former empire, which was still highly significant in 1924-29, had apparently disappeared by 1930-32.⁶

The Federation of Malaya

Another example is the 1965 breakup of the Federation of Malaya, which had at the time only recently attained its independence from Great Britain, into the two sovereign countries of Singapore and Malaysia. Trade between the two countries grew subsequently, but only because each of the economies grew. Adjusting for their size in world markets, trade between Singapore and Malaysia fell by 2 per cent in the year of dissolution (relative to 1964), and had fallen by 8 per cent by 1967. The Malaysia-Singapore example illustrates again that political federation or dissolution matters. The effect is much smaller than for the Austro-Hungarian case, perhaps because the Federation of Malaya had only existed a short time, while the Hapsburg Empire had been around for many centuries.

The Former Soviet Union

⁶ Maurel (1995). The effect reappeared in 1933-1938. The fluctuations in the 1930s are probably influenced by new shifts in regional relationships, at least as much as by gradual unwinding of the legacy of the Austro-Hungarian Empire.

The most important break-up of our time is the dissolution of the Soviet Union. The data on trade among the republics of the former Soviet Union (FSU) exist, pre-breakup as well as post-breakup. Unfortunately, there are severe measurement problems, associated in particular with the proper valuation of the trade. Furthermore, we must recognize that the thoroughly non-market nature of trade within the FSU makes it a fundamentally different case from federations in the rest of the world. Between 1990-91 and 1992-93, the share of trade that the FSU republics conducted among themselves fell by more than half, as one would expect. Trade with the rest of the world also fell sharply, however. The most evident explanation why all trade flows fell is that income fell sharply.⁷ The simple trade concentration ratio [which divides the intra-FSU trade share of the FSU's share of world trade, and which was very high to start with], more-than-doubled between 1990-91 and 1992-93.

⁷ Inter-republic trade on the whole tended to fall much more rapidly than income from 1991 to 1992, but some of the republics' trade with Russia actually rose relative to their incomes. Stern (1994, p.15, 16, 19).

This figure runs counter to all the other examples. One likely possibility is that the FSU's share of world trade is overstated by the use of official exchange rates in 1990-91, and perhaps understated in 1992-93.⁸ Another likely possibility is that a severe loss in trade credit and foreign exchange reserves impeded trade with the rest of the world more than it impeded trade among the republics, especially with Russia itself, much of which continues on a counter-trade basis. A similar effect was observed over this period among Eastern European states: intra-regional trade did not fall as rapidly after the dissolution of the Soviet Bloc as the decline in incomes and the other determinants of the gravity model would lead one to expect.⁹ Both within the former Soviet Union and within Eastern Europe, the barter-like arrangements the countries had long had with each other appear initially to have withstood the tremendous disruptions of the period better than did their fragile trade links with the rest of the world. In this case, we would expect to see a more normal pattern of trade to have begun emerging subsequently -- a pattern more oriented to the rest of the world. One measure {which holds the weight in world trade at 1990 levels, to avoid valuation problems, but allows trade-to-GDP ratios to vary} shows intra-FSU concentration peaking in 1992, and then falling in 1993 and 1994.

The "Velvet Divorce" Between the Czech and Slovak Republics

Czechoslovakia officially divided into two republics on January 1, 1993, resuming the dissolution of the old Hapsburg Empire that had taken place 75 years earlier. Data on trade between the Czech and Slovak halves exist both before and after the "velvet divorce." As Table 6.1 shows, the

⁸ An alternate measure based on implicit exchange rates, however, gives similar results for the concentration ratio [while not even showing a decline in the simple intra-regional trade ratio]. A third measure, at domestic prices, is the only one available for the years 1987-1991; it confirms the simple intra-regional trade share at .7 in 1990 [the same as the other two measures], and reports that this statistic and the concentration ratio had been steady over the preceding four years.

⁹ Brada (1993).

ratio of intra-Czechoslovak trade to total trade was a high .38 in 1991. Trade between the two republics declined a bit in 1992, as the dissolution was contemplated, and then fell sharply in 1993, when it took place. Meanwhile, trade with the rest of the world was growing. Thus, the intra-Czechoslovak ratio fell sharply in 1993, to half its 1991 level. (It then recovered slightly in 1994.) The intensity ratios, which adjust for Czechoslovakia's share of world trade, tell the same story. The net effect of the dissolution, from 1991 to 1994, was a decline in trade between the Czech and Slovak regions of about 44 percent, relative to what could be expected had the federation endured. Stated in reverse, Czech-Slovak trade in 1994 would have been higher by a factor of about 2.3, if the union had continued.

[Table 6.1 about here]

An Example of Reunification: Germany in 1990

While a number of countries have split up over the last three decades or so (the Malayan Federation, Pakistan, the Soviet Union, Yugoslavia, Czechoslovakia), there is one major example of reunification: the re-absorption of the eastern Lander into West Germany in 1990, after the fall of the Berlin Wall. Data are available on trade by the two halves before and after. It is an important experiment, because cultural and geographic proximity of the two Germanies dictate a high natural level of trade, and yet there was a severe unnatural barrier between the two, until it was suddenly lifted. It is also a chance to check for an asymmetry, under which trade could respond more or less quickly when barriers are removed than when they are erected.

Table 6.2 shows a six-fold increase in the level of intra-German merchandise trade, from 1979 to 1994. The share of intra-German trade increased four-fold over this period. This change reflects not only the redrawing of national borders, but also the radical switch of the East German economy to a market system.

[Table 6.2 about here]

There were other major influences on East Germany's trade at this time: on the one hand, trade with all Western countries was opened up at the same time as trade with West Germany, while, on the other hand, a recession in East Germany may have depressed trade generally. Computing the trade concentration ratio is the simplest way of adjusting for these influences. It shows that intra-German trade increased by four-fold between 1989 and 1994. The biggest increase comes in 1991, as expected, with steadily diminishing increments thereafter.

The four-fold positive effect of the reunification on trade after four years is much greater than the negative effect that we estimated for the breakup of the Malaysian Federation (or the effect for the Soviet Union, which we were unable to find at all). The effect is also somewhat greater than that of the dissolution of Czechoslovakia (a factor of 2.3), after the passage of three years). The large effect in the case of Germany reflects that the pre-1990 barriers consisted of far more than a simple political border, and that the reunification constituted the restoration of a state of high integration that had existed until 45 years earlier. The German case is probably too special to tell us much about asymmetries between the erection of political barriers, in general, and their removal.

All these estimates, even the German example, suggest, when compared to the much larger estimate for the Canadian federation, that only a small fraction of the effect is felt within five years of a change in political status. One suspects it may take as long as a century to attain the full effect. This is an important lesson for the European Union: it should not expect that political union would be followed within three-to-five years by anything like the full twenty-fold increase in intra-regional trade that the Canadian example suggests.

Evidence from the home-country bias to trade

While the few available examples of intra-federation trade are fascinating, they leave one

hungry for more information on the effect of national political boundaries. An alternative source of information is the statistical tendency of all countries to undertake economic transactions domestically versus internationally. We saw in Chapter 3 that countries are much less open than most people think: the ratio of trade to income is much smaller than it would be if economic transactions were in fact as easily undertaken with counterparties around the globe as with counterparties around the corner. Much of this difference is due to international trade barriers and to costs of doing business at a distance. But much of the difference remains after taking these factors into account, to be attributed to the apparently-inevitable costs associated with doing business across national borders.

The intercept terms in our gravity model reflect openness. By inspecting them we can compare openness across time or across countries.¹⁰ To put an interpretation on the absolute level of openness, however, requires a separate set of tests, as carried out by Wei (1996). There are (at least) two bothersome new data problems in this approach, which have not arisen in our international gravity equation, where we do not care about the absolute value of the constant term. First, the bilateral merchandise trade data and output data are not comparable because output includes services, and bilateral data on trade in services are not comprehensively available. Wei thus subtracts output in the sectors of construction and other services from the national GNP data, to make them comparable with the trade data. This means limiting the sample to 19 OECD countries; most other countries lack the necessary data. Second, we need some measure of the average distance between buyer and seller for domestic transactions, comparable with the international distance data. Much of trade within Russia or Canada takes place over a longer distance than trade between European countries. He estimates the average distance of domestic transactions as 1/4 of

¹⁰ To compare across time, look, for example, at the intercepts in Tables 6.1-6.3, discussed in the appendix to Chapter 4 [. To compare across countries, look at the intercepts in Appendix Table A4.5 [, which estimate openness relative to the United States.

the distance to the nearest neighbor.

Wei finds that the bias toward trading domestically is high and significant, though not as high as McCallum (1995) and Helliwell (1995) found for the Canadian federation. First, controlling only for country sizes and distance, to make the results comparable with the McCallum-Helliwell estimates of a 20-fold effect in the Canadian case, Wei finds that the home-country bias is an approximate factor of 10. Then, when he holds constant for adjacency, common language, and remoteness, the home-country bias to trade falls sharply, to something like a factor of 2. This estimate is much closer to those that emerge from the historical episodes of federation break-ups. It appears that part of the striking McCallum-Helliwell estimate may be attributable to the effects of adjacency and remoteness.¹¹ Still, the remaining intra-country bias is large and important.

Wei's second finding is that the downward trend in the home-country bias to trade, over the sample period, 1982-1994, is small in magnitude and significance. His third finding is that the bias against trading with foreigners is somewhat reduced when both countries are members of the EC. This difference does not show up in the case of EFTA.

6.2 The Importance of History

In this section we pursue the longer-term historical perspective. There are a number of

¹¹ Perhaps as much as a factor of five [$10/2 = 5$]. This would apparently leave a factor of four to be attributed to intra-Canada bias [$20/5 = 4$]. But the only way to know would be to estimate the gravity model on provincial data, including variables for adjacency, remoteness, and common language. Adjacency refers to the fact that Canadian provinces share borders with each other somewhat more often than they do with American states. Remoteness refers to the fact that Canada is one of those countries that is located rather far from the rest of the world, so that it is more dependent on trade with its immediate neighbor and with itself. The remoteness effect is examined later in this chapter.

lessons for the present.

The history of colonial links

Studies of the historical statistics illustrate the tendency for bilateral trade ties to change relatively slowly over time. Long-span historical studies usually use the simple measures, such as the intensity or concentration ratios, which have less demanding data requirements than the gravity equation.¹² Anderson and Norheim (1993, p.38), examining intra-regional trade intensities from 1928 to 1990, find numbers above 1, and steadily rising, for Europe, the Americas, and each of their halves (East vs. West, and North vs. South, respectively). Asia, Africa, and the Middle East each show even higher intra-regional intensities, but without the clear long-term trends. Recall that an intensity measure above 1 shows intra-regional concentration, but cannot distinguish whether it is due to natural factors such as proximity, or to policy factors such as preferential trading arrangements. The time-series comparison does bear on the latter question.

Kleiman (1976) finds that the bias in trade toward the colonial power around the time of independence (the early 1960s for many French and British colonies) was approximately a factor of two to four, on average across all colonies and ex-colonies. The factor was 2 to 3 for trade between the United Kingdom and its colonies, somewhat higher for France, and higher still for Belgium, Italy, and Portugal (in 1960-62). He also finds that the bias, which he attributes to enforcement (imperialism), decays rapidly after independence: by about one-third for countries that have been independent four to six years, and by about three-quarters for those who have been independent for two decades.

¹² Flandreau (1993), however, manages to estimate a simple gravity model for Europe, during the period 1860-1880. He finds that trade within the Latin Union and Scandinavian Union can be explained by proximity, and need not be attributed to the effects of these early regional trading arrangements.

Examining British trade intensities going back to 1913 (and simple trade shares going back to the 1850s), Anderson and Norheim find that the concentration of trade with the countries that made up the British Empire peaked in 1938, and has declined monotonically since 1958, falling below the critical level of unity by 1984. The intensity of French trade with the countries that were French colonies peaked at a very high level in 1938 and has declined monotonically ever since, while remaining substantially above 1 as recently as 1990. North America shows a gradual but fairly steady upward trend in its intra-regional intensity throughout the period 1929 to 1990. Latin America (excluding Mexico, which is classified in North America) shows an intensity during this period that has a greater average trend, but suffered a major step backwards between 1948 and 1958. The intensity of Japan's trade with Asia peaked at a very high level in 1938, and has been declining fairly steadily ever since. This is precisely the same result found by Petri (1993).

Historical estimates with the gravity equation show links and trends similar to the intensity coefficients. One approach is to add a separate variable representing trade of Britain with the Commonwealth countries, but excluding trade among the latter. The intent is to capture a special metropolitan relationship between the core country and its periphery. Linnemann found different coefficients for the two kinds of transactions, for 1958-1960, and even more so for the relationship between France and its former colonies. Wang and Winters (1991) and Hamilton and Winters (1992) find significant effects for UK Ex-colonial relationships (though not French) as late as 1984-86.

Blocs in the 1930s

The 1930s are generally considered to have been the height of regionalism at its worst. Existing tariff preferences among members of the British Commonwealth were strengthened by the Ottawa Agreements of 1932. This British preference system was associated with an increase of tariffs

on outsiders. From 1931 onward there was an identifiable sterling zone. The maintenance of fixed exchange rates among this group, in a world of increasingly variable exchange rates, may have helped promote trade within the group. The sterling bloc overlapped imperfectly with the Commonwealth.¹³ For those countries that stayed on the gold standard, trade may again have been promoted relative to the alternative, although the widespread use of exchange controls and other trade barriers to protect the balance of payments obviously had the opposite effect.

A bloc of Central and Southeastern European countries formed around Germany, particularly under the barter-like "Schacht Agreements". An attempt to liberalize trade within a group of Benelux and Scandinavian countries was undermined by a split with regard to currency policies between the gold bloc and silver bloc.

In the aftermath of the trade-decimating [though non-discriminatory] Smoot-Hawley tariff, U.S. Secretary of State Cordell Hull negotiated liberalization agreements bilaterally under the 1934 Reciprocal Trade Agreements Act. Countries that pegged their currencies to the dollar were fewer and less important, however, than the membership of the gold and sterling blocs (chiefly Spanish-speaking countries that were subject to extensive American political influence).

Eichengreen and Irwin (1995, 1996) apply the gravity model to the interwar period, in part to show the importance of long-lasting effects of past patterns of bilateral trade on current patterns. Eichengreen and Irwin (1995) find statistically significant intra-regional bias for the British Commonwealth and Central Europe blocs in 1935 and 1938, which is as expected. Less expected is their finding that much of this effect is simply a continuation of significant bloc effects in 1928 [intra-regional trade approximately double the level of trade with otherwise-similar non-member countries]. To be sure, the estimated effect of the British Commonwealth preferences does go up

¹³ For example, Canada was in the Commonwealth but not sterling zone, and three Scandinavian countries vice versa; Eichengreen and Irwin (1995).

between 1928 and 1935, presumably as the result of the Ottawa Agreements.

In Eichengreen and Frankel (1995), the Commonwealth bloc is slightly stronger than the German bloc in the 1930s, and the dollar bloc is not statistically significant. The only strong significant evidence of trade-diversion is in the German bloc in 1938.

Eichengreen and Irwin (1995) test for trade blocs and currency blocs in the 1930s with dummy variables. Their results point to a number of conclusions: (1) Worldwide trade was depressed in the 1930s, not only by trade barriers, but also by exchange rate variability; and (2) trade was also increasingly regionalized in the 1930s by the sterling area, Commonwealth preferences, a Reichsmark bloc, a Benelux bloc, and a Western Hemisphere bloc. The authors choose to emphasize the evidence that (3) these blocs were already beginning to affect trade flows in 1928, earlier than the regionalization is conventionally dated.

Lags in trade patterns

We have noted the tendency for trading patterns to change relatively slowly over time, even when the change in regional trading arrangements or political links is sudden. Eichengreen and Irwin (1996) take the bull by the horns and include lagged values of bilateral trade in their gravity estimates for the period from 1928 to 1965. They find, for example, that trade links among British colonies in 1954 and 1964 that might otherwise be attributed to Commonwealth preferences, are in fact simply the lagged effects of trade flows of 1949, when the countries belonged to the British Empire. They use the term hysteresis to refer to these effects that seem to linger long after the original reasons for the bilateral trade have vanished. They conclude that one should always include lagged variables in the gravity equation.

A possible alternative interpretation is that one should be careful to look out for missing variables. These variables are likely to change slowly over time, and so can produce significant

coefficients on lagged trade. One possible missing variable that addresses the hysteresis issue directly is the accumulated stock of Foreign Direct Investment. If firms of the mother country develop a large stock of capital in a colonial dependency, this investment might be manifest in a heightened level of trade between the two countries long after the colonial relationship has been severed. By including bilateral FDI as an explanatory variable, which we do below, we hope to capture the effect directly. The measured stock of FDI will necessarily omit such intangibles as brand loyalty among consumers in the importing country or familiarity with local languages and customs in the exporting country. But bilateral FDI should capture much of the essence of this relationship.

6.3: The Effect of Cumulative Foreign Direct Investment

If a firm inherits a stock of retail outlets in another country, it is likely to export more to that country, other things equal. Or, if the firm inherits a stock of factories or plantations or mines in another country, it is likely to import more from that country. Furthermore, subsidiaries are more likely than local firms to turn to the mother country for intermediate inputs. There is also the tendency that when an "imperialist" power builds the infrastructure in a country of the periphery, whether it is the railroad system in colonial Africa or the port system in modern Southeast Asia, it will do so in such a way as to facilitate trade links with itself. Thus, for all these reasons, one expects a positive association between FDI (Foreign Direct Investment) and trade.

Theoretically, there is also an effect that can go the other way. If high barriers to trade, such as tariffs or quotas, prevent firms from exporting into the protected market, they may undertake direct investment as a substitute, jumping the barriers by building factories to supply the protected market directly. Empirically, however, most studies have found that the positive association between FDI and trade dominates.

The proponents of the yen bloc hypothesis tend particularly to emphasize the role of FDI as a tool that Japan uses in East Asia to divert economic relationships toward itself. They point to the rapid growth of Japanese FDI in East Asia in the 1980s. These arguments often neglect that the increase in Japanese FDI into East Asia was in proportion to the increase in trade, which we have in turn described as natural in light of the rapid growth of the economies involved. Moreover, Japanese FDI into East Asia is smaller than into the United States and Canada. The latter is more than twice what one would expect from North America's share of world trade. This pattern, evident in the official Japanese Ministry of Finance figures on the destination of FDI, applies even more strongly if one takes into account that they actually represent statistics on investment either approved by or reported to the government, and greatly overstate the extent of true Japanese investment in developing countries. The more accurate balance-of-payments data from the Bank of Japan show a smaller percentage of investment going to Asia.¹⁴

Nevertheless, it is well worth exploring the possibility that the trade flows we have been analyzing are influenced by past FDI. We wish to test the hypothesis that the stock of foreign direct investment which country *i* has placed in country *j* is an important determinant of exports from *i* to *j*. Thus we try adding cumulative bilateral FDI as an additional factor in the gravity model of bilateral trade.

¹⁴ Ramstetter (1991a, pp. 8-9; 1991b, pp. 95-96) and F (1993, pp. 67-69).

Bilateral FDI data are only available on a limited basis. Many previous studies have looked at bilateral FDI between the United States and other countries, or between Japan and other countries.¹⁵ But there are few worldwide estimates. We have put together a relatively comprehensive and recent bilateral FDI data set.

There are serious problems of endogeneity: all the factors that affect bilateral trade could also be expected to affect bilateral FDI. The simplest way to try to address this problem is to use the lagged stock of FDI as an explanatory variable. Results for 1992 trade, determined as a function of the 1990 stock of FDI, are reported in Table 6.3.¹⁶ Bilateral FDI has an apparent high and very significant effect on bilateral trade. The point estimate is .17. In other words, a 1 per cent increase in the stock of FDI leads to a subsequent .17 per cent increase in exports.¹⁷ The coefficients on distance, adjacency and language decline slightly, but the gravity results are basically unaffected. The estimated coefficients for all bloc effects decline slightly, suggesting the possibility that the bloc effects earlier might have been appropriating a bit of the FDI effect. The bloc effects for APEC, the Western Hemisphere (when holding constant for openness), and the EC (when not holding constant for openness) are still significant.

[Table 6.3 about here]

Simply lagging the stock of FDI is not, however, truly a satisfactory way of addressing

¹⁵ Sachs and Shatz (1994, 45-49) show in a gravity model of U.S. bilateral exports and imports that the bilateral activities of multi-national corporations are a significant determinant.

¹⁶ Taken from Table 4 of W & F (1997). This test necessitates introducing the distinction between the importing country and the exporting country, which is discussed at length later in the chapter. The FDI tests are extended to 1990 and 1994, with a special emphasis on Southeast Asia, in F & W (1996).

¹⁷ An extension of these tests, with an extra emphasis on Southeast Asia, finds a somewhat smaller effect in 1990, and a somewhat larger effect in 1994 (F & W, 1996). The effects are significant in all three years.

endogeneity. Even in a time series context, precedence does not insure causality. Firms could undertake direct investment in rational anticipation of a market for trade that will expand in the future. In a cross-section context, the simultaneity problem is even worse. We have obtained what seems to be a good instrumental variable for bilateral FDI: a dummy variable indicating the existence of a bilateral tax treaty. This is a good instrumental variable in that it is a significant determinant of bilateral FDI, as a ("first-stage") regression confirms. When the tax treaty variable is used to isolate exogenous variation in the stock of FDI, we find that the effect of FDI on trade vanishes. This finding suggests that, even though bilateral FDI is highly correlated with bilateral trade, there is in fact no causal relationship running from the first variable to the second. (An alternative possibility is that the endogeneity problem is even worse for the tax treaty variable than for FDI itself.)

One can estimate a gravity model of the determination of FDI, analogous to the standard one for trade.¹⁸ Results presented in Frankel and Wei (1996) are a start. They omit many of the possible determinants that have been identified in the literature on FDI. Nevertheless, the gravity framework has its attractions, notably a much larger number of observations than in the typical study of FDI.

¹⁸ Eaton and Tamura (1994, 1996) estimate bilateral gravity models for FDI. But they include only two source countries: the United States and Japan. They find that features of a country associated with more trade with the U.S. or Japan are also associated with more FDI from those countries.

To the extent that the motive for FDI is to sell into the local market, one might expect distance and transport costs to have, if anything, a positive effect on FDI, thus reversing a key plank of the gravity model. On the other hand, to the extent that the motive is exporting back to the source country, distance should have a negative effect, just as it does for trade. The same is true if distance matters because it breeds unfamiliarity with local culture.¹⁹ We find that the coefficient on distance is even more significant and negative than is the case in the gravity model of trade. Similarly, the coefficient on language (which also includes former colonial links) is extremely high and significant. In 1992, the existence of linguistic links raised the stock of FDI by about nine-fold ($\exp(2.24) = 9.4$). In this light, it is not surprising that the addition of FDI into the trade equation, in the results reported above, deprived the language variable of its statistical significance. The coefficients on GDP are also highly significant.²⁰ To sum up the results on the determination of bilateral FDI, they seem to be similar to the determinants of bilateral trade, which explains the effect that lagged FDI had in the earlier trade equations in the preceding section.

A full exploration of the determinants of FDI is outside the scope of this book, though it is relevant to the study of regionalization. Many of the regional economic arrangements include liberalized rules for Foreign Direct Investment. A full model of the simultaneous determination of bilateral trade and investment would raise many interesting questions, but will have to wait for future extensions.

6.4: Rich and Poor Countries

¹⁹ Eaton and Tamura (1996) find in their gravity model that distance inhibits FDI much less than it inhibits trade.

²⁰ But one knows that there is probably a bad misspecification in this equation in this regard. We have not yet included terms for GDP per capita, which would capture the fact that rich countries tend to be the source of FDI.

We have been treating countries with perfect symmetry. This would surprise trade theorists of the past, as international trade has always been thought to be motivated by countries' *differences*. Poor countries produce goods intensive in unskilled labor, and trade them to rich countries, for goods intensive in capital and skilled labor.

The Linder Effect

We tried including bilateral absolute differences in GNP/capita figures, to test for Linder or Heckscher-Ohlin effects. The variable did not have the positive effect that one would expect if countries traded capital-intensive products for unskilled-labor-intensive products. Rather, it had a moderately significant negative effect, as in the Linder hypothesis that similar countries trade more than dissimilar ones. This undercuts support for the Heckscher-Ohlin hypothesis. Meanwhile the positive coefficient on the product of per capita GNPs remains, as before, very highly significant, suggesting that the most powerful effect of the three is the relationship between development and trade.²¹

Factor endowment differentials

We also tried better to capture classic Heckscher-Ohlin effects. We estimated a gravity equation that included more direct measures of factor endowments: the two countries' differences in capital/labor ratios, educational attainment levels, and land/labor ratios. The data (for a subset of 656 of our 1,953 pairs of countries) were generously supplied by Gary Saxonhouse (1989). The results are not quite as bleak for Heckscher-Ohlin as was the negative coefficient on income differentials. There is a bit of support for some of these terms, specifically for capital/labor ratios and

²¹ F, S and W (1995, Table 3).

educational attainment in 1980. In the other years, however, the results are poor. The coefficients on the other variables are, in any case, little affected.²²

The endogeneity of incomes

We have treated both economic size and income per capita as exogenous variables in the gravity equation. Yet there are good reasons to suspect that they are endogenous, influenced by the level of trade. Economic theories, from the basic principle of comparative advantage to modern theories of technology transfer through trade, have long said that countries that are more open to trade will have higher or more rapidly increasing levels of real income. There is good empirical support for this effect as well.²³

But trade is the dependent variable in our gravity equation. Thus the apparently significant effect of income on trade in our equation could be spurious, a correlation that is in truth attributable to the influence of trade on income. Not only would estimation of the income coefficient be biased in the event of such a simultaneity problem, but so would the estimation of the other coefficients. If, for example, FTAs are more likely to be successful among rich countries than poor ones, improper measurement of the effects of income on trade could also lead to improper measurement of the effects of FTAs.

[new Table 6.4 about here {from Cyrus}]

²² The results are reported in table 4 of F, S & W (1995); also in Table 5 in F and W (1993b).

²³ Frankel and Romer (1995) and Frankel, Romer and Cyrus (1995) give extensive references to the empirical literature. Those two papers also show that the statistical effect of openness on growth holds up well, even when allowing for the possible endogeneity of trade. Thus, while we are about to use exogenous variables from the growth equation (such as investment) to correct for simultaneity bias in the trade equation, the two earlier papers use exogenous variables from the trade equation (such as distance) to correct for simultaneity bias in the growth equation.

While many variables determine countries' income levels, the most important quantifiable determinants are their levels of factor accumulation: labor force, stock of physical capital, and stock of human capital (education and technological know-how). It is plausible that these determinants of income are exogenous, and thus good instrumental variables.²⁴ Table 6.4 {had to renumber later tables} reports along these lines the results of instrumental variables estimation of the gravity equation.²⁵ We follow the standard neoclassical growth model in listing the three factor accumulation variables as determinants of income in IV1, and we follow the "conditional convergence" literature in adding 1960 income to the list in IV2. The use of the instrumental variables technique sometimes reduces the coefficient on income or income per capita very slightly, but the differences are not statistically significant. The other coefficients are usually little changed as well, whether on other gravity variables, narrowly defined regional trading arrangements, or more broadly defined blocs. Evidently the endogeneity of income makes little difference.

6.5: The Role of Currency Links

In addition to the strategy of enacting regional trading arrangements, governments sometimes seek to promote trade among a group of countries by first linking their currencies. The link might be relatively loose, as in the Exchange Rate Mechanism of the European Monetary System, or somewhat tighter, as in the fixed exchange rates that some Latin American countries have

²⁴ Wei (1996) also allows for the endogeneity of income in the gravity equation, using simple population as an instrumental variable. Harrigan (1991) uses a more complete set of factor endowments as instrumental variables for income. Unfortunately, he omits any measure of distance from his calculations, which must have a major effect on the results.

²⁵ Drawn from Cyrus (1996). Results for 1980 and 1985 are reported there as well.

established against the dollar, or very tight, as in the common currency on which the members of the European Union have set their sights. The idea is that currency links reduce the costs to doing business with the partner country, in the form of foreign exchange risk (including the cost of hedging such risk) and transaction costs. The reduction in costs should promote trade. This was certainly a major motivation on the part of the fathers of the proposal for EMU (European Economic and Monetary Union).

Some European leaders believe that currency links are not just a desirable supplement to a successful common market, but are actually a necessary component of it. Others have also read the Latin American experience as suggesting that exchange rate stability is a prerequisite for successful regional trading arrangements: extreme instability helped to do in some of the PTAs of the 1970s, and large swings in the real exchange rate between the Argentine peso and the Brazilian cruzeiro have put strains on Mercosur.²⁶ The Mexican peso crisis of December 1994 (rightly or wrongly) soured some Americans and Mexicans alike on NAFTA and its expansion. On the other hand, exchange rate stability does not seem to have been necessary to Canadian-U.S. trade links. The exchange rate between the Canadian and U.S. dollars has floated more cleanly and over a higher fraction of the post-war period than any other bilateral exchange rate in the world.

We can readily investigate the extent to which the bilateral stabilization of exchange rates within major regional groupings has contributed to the regionalization of trade. The question can be decomposed into two: to what extent have bilateral exchange rates in fact been stabilized within the major regional groupings, and to what extent does reduced bilateral exchange rate variability promote trade. As regards the first question, the unsurprising finding is that the values of currencies of European countries are tied far more closely to the value of the mark than to other major currencies, and the values of Western Hemisphere currencies tend to be tied to the value of the

²⁶ E.g., Abreu and Bevilaqua (1995).

dollar.

The finding that surprises some people, however, is that out of nine currencies in East Asia, only a few show a significant or growing link to the yen, and all show that links to the dollar are much stronger than to the yen. Although only Hong Kong pegs formally to the dollar, no currency pegs formally to the yen. Those that have at times adopted loose basket pegs (Malaysia, Singapore and Thailand) give far more weight in the basket to the dollar (about .8) than to the yen (about variability that prevailed in 1980, a standard deviation of 2 per cent, had been eliminated altogether, the volume of intra-EC trade would have increased by 14.2 per cent. This OLS estimate should be regarded very much as an upper bound. For one thing, the 1980 point estimate of the effect of exchange rate volatility is the largest of all the years. In the earlier observations, the magnitude of the estimated effect is 1/5 to 1/2 the size.

Interpretations of the OLS estimates are complicated by the likelihood of simultaneity bias in the regressions. Governments may choose deliberately to stabilize bilateral exchange rates with their major trading partners. This has certainly been the case in Europe. Hence, there could be a strong observed correlation between trade patterns and currency linkages even if exchange rate volatility does not depress trade. To address this problem, we use the method of instrumental variable estimation, with the standard deviation of relative money supply as our instrument for the volatility of exchange rates. The argument in favor of this choice of instrument is as follows. Relative money supplies and bilateral exchange rates are highly correlated in theory (they are directly linked under the monetary theory of exchange rate determination), and in our data as well, but monetary policies are less likely than exchange rate policies to be set in response to bilateral trade patterns.

The Instrumental Variables results show the same sign pattern across the years as the OLS estimates, but the negative effect is statistically significant only in 1965. The coefficient for 1980 is (a completely insignificant) 0.28. Even if the point estimate is taken at face value, it would imply that the

elimination of exchange rate variability worldwide would increase trade by only 0.9 per cent ($=0.28 \times 3.22$). The weak results when correcting for simultaneity is the second strike against the hypothesis that the stabilization of exchange rates within the EU or other groupings has been an important factor in promoting intra-regional trade.²⁷

To summarize, these results are generally consistent with the hypothesis that real exchange rate volatility depressed bilateral trade a bit in the 1960s and 1970s. But the evidence for a negative trade effect, which starts out relatively strong in 1965, diminishes steadily in the 1970s and 1980s, especially if one takes due account of the simultaneity. Even as the exchange rates were becoming more volatile, the effect of any given level of volatility fell. Presumably, importers and exporters learned to cope with uncertainty. The proliferation of currency options, forward contracts, and other hedging instruments over this period may explain why the effect that appears once to have been there, has more recently disappeared.

In any case, the exchange rate variability term does little to reduce the strength of the trade bloc effects among such relevant groupings as the EC, the Americas, or APEC.

6.6]: **Bilateral exports and imports: The roles of relative prices and relative distance**

In the interest of ensuring that the results from our gravity model are as robust as possible, we have considered a number of further extensions. Those in this section concern the distinction between the importing country and the exporting country.

Disaggregating imports from exports

²⁷ The estimates are taken from F & W (1996), Tables 4, 5 and 6. Similar results are reported in the earlier F & W papers.

We have focused primarily on estimation results where imports and exports are aggregated together. Our goal has been to uncover the deep underlying structure of bilateral trade patterns. We think that exports plus imports (say, divided by 2, to arrive at the average) are a good measure of this underlying trade structure. It is common in gravity models to treat imports and exports separately, without constraints forcing the coefficients on domestic and foreign income, or income per capita, to be equal. Most studies of the model in this form find income coefficients that are slightly different on the import side and the export side. Linnemann, for example, found that the estimated income elasticities were different for the two, but that constraining them to be equal in spite of this produced little change in the results. If one concludes that the coefficients are equal, then it is legitimate to impose this constraint. Coefficients on distance and the other bilateral variables are necessarily equal for imports and exports.²⁸ Thus it is in turn legitimate to add imports and exports together, to simplify the estimation and the presentation, as we have done.²⁹

A further motive for adding imports and exports together, aside from the simplicity of dealing with fewer equations, is that to treat them separately requires that one enter the realm of macroeconomics. As soon as one considers imports and exports separately, one has to recognize that their difference is the trade balance. The trade balance is not determined by the deep underlying structure of trade. Rather it is the difference between saving and investment, and so is determined by such macroeconomic factors as the real exchange rate, business cycles, government

²⁸ For a given country, proximity or common language could conceivably have a bigger effects on imports than exports. With our general formulation however, where the variable to be explained is simply T_{ij} , no such distinctions are possible. [When we use distance to measure transport costs, for example, we abstract from the fact that sailing time or flying time from east to west may be different than from west to east.]

²⁹ It is, moreover, legitimate to take logs. One has simply added $\log(2)$ to the righthand side of the equation that describes the log of exports and the log of imports.

spending, investment opportunities, and intertemporal optimization by savers.

Exploring the realm of macroeconomics has many rewards. Robust econometric results do not tend to be among these rewards. Statistical estimates often change wildly every time one extends the sample period, or modifies the specification slightly, or applies a different technique.³⁰ Until this point in our study, the hope (perhaps a vain one) has been that when a country undergoes a real appreciation, the positive effect on its imports and the negative effect on its exports approximately cancel out. This would justify omitting a term for the real exchange rate, in the equation that explains the sum of exports and imports.

These arguments notwithstanding, if the income coefficients were in truth different for the importing country and the exporting country, then it would be no good to pretend otherwise. Aggregating the two equations together would not be legitimate. As a further check on the robustness of our results, we have now tried the estimation on imports and exports disaggregated.

We begin by relaxing the constraint that the income coefficients are equal in the importing and exporting country, but without the real exchange rate term. The results are reported in Table 6.5 (Table 6.5a for 1965-75, and 6.5b for 1980-90; Table 6.3 reported the estimates for 1992). As others have found, the GNP coefficient is a little greater for the importing country than the exporting country. This coefficient is consistent with the old-fashioned Keynesian idea of a

³⁰ A few authors have sought to estimate relative price terms in their bilateral trade equation, on the grounds that they appear in a gravity-type equation derived from a general-equilibrium theory, without thinking of themselves as having entered the realm of macroeconomics. In fact, however, relative price variation is in practice so heavily dominated by exchange rate variation, that there is really no way of escaping that this is a macroeconomic phenomenon. This is especially true because conventional indices measure prices only relative to a base year, preventing comparison on a pure cross-section dimension. These studies tend to get unsatisfactory results for the relative price terms.

demand-determined marginal propensity to import, expressed in elasticity form. But the difference between the import elasticity and export elasticity is small, and often insignificant statistically. The difference in coefficients on GNP per capita is larger, and more significant, with the exporting country always having the higher value. This might suggest that, of the various explanations for the role of per capita income in determining trade, the tendency for poorer countries to engage in more import protection does not seem to be the dominant one. One must always recall, however, that averaged over the very long run, a country's exports equals its imports.

[Tables 6.5a and 6.5b [formerly 7.1a and 7.1b] about here;

The magnitudes of the estimated coefficients on the per capita incomes is higher than in the earlier estimates. The same is true of the coefficient on distance. It still has a trend that, if anything, rises rather than falls. This is particularly true in the 1970s. (Over the decade 1980-1990 it is fairly steady at .9.) The estimated coefficient on the adjacency variable now has a strong upward trend from year to year. Evidently the costs to doing business with close neighbors have been falling more rapidly than the costs to doing business with distant ones.

What are the effects on the bloc variables? As before, the strong effects belong to East Asia and APEC. As before, the trend in the East Asian bloc is, if anything, rising in terms of j ; so the real exchange rate is the relative price of i 's goods in terms of j 's. In this table (as in the preceding one) we use the Summers-Heston measure of real GDP, as well as their measure of the real exchange rate.

[I am now omitting the former 7.2a and 7.2b];

The coefficient on the real exchange rate fluctuates in sign and is usually insignificant statistically. This is similar to what others have found. Boisso and Farrantino (1993) also had little

success with a relative price term. Bergstrand (1985, 1989) emphasized the importance of prices in theory, but the estimated coefficients were again usually not significant statistically. We will not concern ourselves with the real exchange rate term further.

Remoteness

In this section, we estimate a modified version of the gravity model. To explain bilateral trade between a country and a specific trading partner, we incorporate the distance of each country from its average trading partner, which we call remoteness, in addition to the direct bilateral distance. This extension of the basic gravity formulation is based in part on a new formulation in Deardorff (1997).³¹

³¹ Also on suggestions from Jacques Polak (1996) and Ed Leamer, and a theoretical formulation in Stein (1995).

The remoteness variable measures how far an exporting (or importing) country is from all other countries. It is a measure of "overall distance." An exporter's remoteness is defined as its average distance from its trading partners, using partners' GNPs as the weights. An importer's remoteness is defined analogously. The hypothesis is that the remoteness of an exporter from the rest of the world has a positive effect on bilateral trade volume in the equation (conditional on bilateral distances). An example will illustrate the intuition. The distance between Australia and New Zealand is the same as the distance between Spain and Poland. Spain and Poland have lots of other natural trading partners close at hand, but Australia and New Zealand do not. One might thus expect the antipodean pair, who have less in the way of alternatives, to trade more with each other, holding other influences constant, than the European pair. The idea is that it is not just the absolute level of bilateral distance that matters, but also bilateral distance expressed *relative* to the distances of each of the pair from their *other* partners. As a check for robustness, we have also computed the remoteness of exporters and importers as the equally-weighted distance from their trade partners, with little effect on the results.³²

Results are reported in Appendix Table A6.4.2]. The coefficient on the exporter's remoteness is always positive and, three out of four times, statistically significant. Other things equal, if country Z is further away from the rest of the world than country S by one percent, then Z's exports to a common third country, say A, will be higher than that of S by 0.3 to 0.6 percent. Another way of stating this result is to break down the coefficient on bilateral distance (.9), into two roughly equal effects: an effect of bilateral distance *relative* to the average distance of the exporter (.3 to .6) plus an *absolute* distance effect (.6 to .3).

The coefficient on the importer's remoteness is consistently negative, surprisingly, and

³² The results are reported in F & W (1997), Table 2. (Table A7.2 here is taken from Table 1 there.)

statistically significant at the five percent level. Apparently, if Z's average distance from the world is greater than S by one percent, then Z's imports from M, other things equal, is less than S by 0.4-0.8 percent. We have not yet figured out why this might be.

In Appendix Table 6.4 we confine ourselves to testing formal FTAs. The bloc variable is defined the same as in earlier tables. (The dummy variable labelled "non-bloc," however, is defined as trade between a member of the grouping and a non-member.³³) The regional arrangements tested were the European Community (EC), European Free Trade Area (EFTA), Canada-U.S. FTA, Mercosur, Andean Group and ASEAN. Year-by-year point estimates of the EC bloc effect and their dynamics tell an interesting story. In terms of levels, within-EC trade has always been below the prediction of the gravity model. In terms of trend, however, the degree of within-EC bias has clearly risen. This suggests that while the European countries were more open to trade than many countries, for historical or other reasons, their trade pattern exhibits evidence of increasing bias among members and increasing trade diversion away from member countries. Trade among the EFTA countries seems below what one would have expected based on their economic, geographic and linguistic linkages, although the difference is not statistically significant. The Canada-U.S. effect is insignificant. The four South American countries that constitute Mercosur traded more intensely among themselves than the gravity model prediction. There appears to have been an increase in the intra-group trade intensity in the 1970's and 1980's. The Andean group also appears to show a certain degree of intra-group trade bias. The intra-ASEAN trade bias is positive and significant in every year. However, the bias appears to diminish in the last decade, from 2.85 in 1980 to 1.80 in 1992.

³³ Its coefficient has the same interpretation as the coefficient on openness in earlier tables. However the coefficient on the bloc variable in this table exceeds the bloc coefficient in the equation specification used before, by the amount of the openness coefficient. (This follows from the definitions of the dummy variables: $\text{openness} = \text{non-bloc} + \text{bloc}$.)

6.7 : Econometric Extensions

Finally, we consider a few extensions concerning the functional form of the equation and the nature of the error term. While these inquiries are a bit more on the technical side, to have explored them will reassure us about the robustness of our basic results.

Interactive effects: Are blocs more effective when the members are close together?

One might wonder whether there are *interactive effects* between bloc membership and the basic gravity variables, distance and adjacency.³⁴ Perhaps membership in a bloc has a greater effect when trading partners are neighbors than when they are far-removed, an effect that is greater than the sum of the two separate effects of bloc membership and distance. We looked at these interaction terms as a robustness check. The coefficient on distance interacted with the bloc effect was always negative, and highly significant statistically in two years out of seven (1965 and 1970). The finding suggests that the combined effect of proximity and bloc membership is greater than the sum of the two individual effects. (The negative effect of distance is, of course, a positive effect of proximity.) The coefficient on adjacency interacted with the bloc effect fluctuated in sign and was not statistically significant.

The results are reported in Appendix Tables A6.1a and A6.1b. The dummy variable for regional blocs had a positive coefficient, and was highly significant in four years out of seven. This

³⁴ Brada and Mendez (1985) find that the bloc effects interact with other effects, that preferential trading arrangements are more likely to boost trade if the members (a) are close together, and (b) have relatively high per capita incomes. Mansfield and Bronson (1995) find that bloc effects interact with the effect of common membership in regional geographical groupings, as well as in multilateral alliances.

result suggests that formal preferential trading arrangements in general promote trade even after taking into account the continental effects.³⁵

Zero-valued entries

For some country pairs the data entry is zero, normally due to levels of trade too small to be recorded. There are 245 missing values in our data set in 1985, for example. These are generally countries that, by virtue of small size and remoteness, would be expected to have little trade with each other. It is not always possible to ascertain whether their trade is literally zero, or is very small and has been rounded down to zero. Either way, these zero pairs present a problem for econometric estimation of the gravity model. The reason is that the standard procedure is to take the logarithms of the original multiplicative gravity equation so as to be able to estimate it in linear form. But one cannot take the log of zero.

³⁵ The Regional Bloc (RB) dummy variable in Table A6.1 is the sum of the narrowly-defined FTAs -- ASEAN, Mercosur, etc. -- not the continental blocs.

There are three approaches that have been most commonly taken in the literature. First one can simply omit the zero pairs from the data set. This is the strategy followed by Brada and Mendez (1985), Bikker (1987) and others. It is also the strategy that we have followed in most of our results. One must be concerned, however, that the exclusion of these data points might bias the results. If so, one could argue that a sample selection problem arises inevitably, from leaving very small countries out of the data set altogether. Even the few studies that include a broader set of countries than our set of 63 nevertheless leave out the smallest countries.³⁶

Second, one can substitute arbitrary small numbers in place of the zeroes, such as \$1,000 or some other minimum unit. This is the technique used by Linnemann (1966) and Wang and Winters (1992), among others. (They found that inclusion of the missing values made little substantive difference to the results.) The virtue of this strategy is that it allows the computer to run the linear OLS regression. The obvious drawback is that it is ad hoc. A less obvious drawback, perhaps, is that the log of a small positive number, though not negative infinity, is still a negative number that is very large (in absolute value). OLS regression in effect gives larger weight to extreme values, whether large or small. For this reason, the zero pairs might then receive too large a weight in the estimates. This is closely related to the problem of heteroscedasticity, which is addressed below.

Third, one can express the dependent variable, bilateral trade, in levels rather than logs (the semi-log formulation), and then use Tobit to estimate. Tobit is a technique that estimates separate parameters to determine whether an observation is non-zero, and then to determine what the coefficients are conditional on the observation being non-zero. This is the technique used by

³⁶ Boisso and Farrantino (1995) use a data sample of over 260,000 export-import pairs. This sample only goes up to 1985, however. The price of including so many countries is that more recent data is not available.

Biessen (1991), Havrylyshyn and Pritchett (1991), and Eaton and Tamura (1994), among others.

Fourth, Eichengreen and Irwin (1995, 1997) express the dependent variable as the $\ln(1+TRADE_{ij})$. Their logic is as follows. On the one hand, when $TRADE_{ij}$ is large, the dependent variable is approximately equal to the usual one, $\ln(TRADE_{ij})$, so that the coefficients can be interpreted as elasticities in the usual way. On the other hand, when $TRADE_{ij}$ is small, the dependent variable is approximately equal to $TRADE_{ij}$ itself [as in the semi-log formulation], which can be interpreted appropriately. Unfortunately this solution entails some inelegant econometric complications.

Our preferred way of dealing with the problem of zero observations is a simple robustness check. The only reason that the zeroes cannot be included is that the log of zero is minus infinity. Why let a pesky detail like that stop us? We tried running the equation in multiplicative form, instead of log-linear, so as to allow the inclusion of pairs of countries that are reported as undertaking zero trade. When the sizes of the countries are close to (or equal to) zero, the predicted level of bilateral trade will also be close to (or equal to) zero, exactly as it should be. Once the equation has been estimated in nonlinear form, we can test whether the results are sensitive to the exclusion of the zero observations. We find that the inclusion or omission of such countries in the multiplicative specification makes little difference to the results. This finding offers some assurance that the omission of the zero observations from our standard log-linear regressions does not lead us far astray.³⁷

Heteroscedasticity

³⁷ The results were reported in F & W (1993b); the Appendix Tables A2 and A3 to CIDER Working Paper 93-025, U.C. Berkeley; and Appendix Table A2-A3 to F & W (1993b). The use of the multiplicative form itself changes the results some, however.

The OLS regression technique attaches the same importance to a pair of large countries as to a pair of small countries. It figures that each piece of information is useful. We might figure, however, that the information contained in the number for U.S.-Canada trade, or Japan-Germany trade, is far more valuable than the information contained in Ecuador-Peru or Iceland-Kuwait trade. A good way to think about this is that a country like the United States is made up of hundreds or thousands of economic units, each of which is no bigger than Iceland or Kuwait. Hypothetically, if we had the trade data for each of these units, their accuracy would likely be not very much better or worse than the data for Iceland or Kuwait. When we add up the U.S. trade data into an aggregate number, it is far more reliable than it would be for a smaller unit: measurement errors tend to cancel out, because of the Law of Large Numbers. Technically, the problem is heteroscedasticity (larger error terms for some observations than others). The appropriate correction in this case is the technique of Weighted Least Squares, based on the size of the countries. As with so many of the other econometric extensions that we have tried, the correction turns out to make little difference.³⁸

6.8 : Conclusions from the Gravity Model

This completes our econometric extensions of the gravity model. We have confirmed that statistically significant regional trading arrangements are indeed springing up in a number of places, both at the level of formal arrangements, and at the broader level of informal continent-wide groupings. The Andean group, Mercosur, and ASEAN remain effective PTAs. At the continental level, so do the EC, the Americas, and East Asia or the Pacific. Perhaps this chapter's most

³⁸ Reported in F & W (1993b) and appendix, *ibids.*, Table A1. Techniques that will correct for heteroscedasticity without requiring any information regarding the structure of the variance matrix have become fashionable among econometricians. When one has a good idea as to the source of the heteroscedasticity, however, as in this case, it seems much better to apply that a priori knowledge.

interesting findings, in their own right, were the trade effects of historical-political links and of bilateral Foreign Direct Investment. The absence of systematic factor-endowment effects is also memorable. The larger message, in any case, is that the previous chapter 5's [previously 6] findings of significant regional effects have now been subjected to a great many perturbations and robustness checks, and, on the whole, have stood up well.

The next question is whether the regionalization trend constitutes an undesirable threat to the world trading system. We will be returning to the gravity estimates toward the end of the book, when it is time to evaluate the regionalization of trade policy according to what can be justified on the grounds of proximity and the other natural determinants that we have analyzed here.

*Table 4.5 [formerly 5.5]:**Estimates of Coefficient on Log Distance in Gravity Equation**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 (1995)	-.51 to -.69	1950-1990

Controlling for adjacency

Aitken (1973)	-.35 to -.51
Biessen (1991)	-.74 to -.86
Bergstrand (1985)	-.75 to -.78
Tinbergen (1962)	-.56

**Appendix to Chapter 4:
A Note on the Intercepts in the Pooled Estimates**

In our pooled time-series cross-section estimates, if one wishes to interpret the intercept terms for each year, one should take into account how much the dollar price level went up over the period 1970-1992, a factor of 3.45 for U.S. prices, and how much real Gross World Product went up, a factor of 2.002.¹ The theory of Chapter 7 suggests that national dollar GNPs in each year should be deflated by dollar Gross World Product. We have not bothered to do so in the cross-section context because it is not necessary. But here we should add 1.9 to the intercept on the 1992 cross-section,² to see if there has been a secular increase in real worldwide trade, relative to 1970, beyond what can be attributed to inflation and growth. As an illustrative example, the estimated intercept term in 1990 is about -1.3.³ Thus the increase in real trade, adjusted for GNP and the other factors (population is the other variable that changes over time) is $-1.3 + 1.9$, or 0.6. Expressing the trend on an annual basis, trade has increased at about 3 per cent per year. This increase could be attributed to declining transport costs, or to worldwide liberalization of trade policy.⁴

¹ The log of the dollar price level increased by 1.24, or .056 at an annual rate, while the log of real income increased by .694, or .032 at an annual rate.

² Where does this number come from? $1.93 = 1.24 + .69$. (See preceding footnote).
If one takes first differences of the gravity equation estimated, sets the GNP coefficients to one, and assumes no change in the other variables, then the estimated equation looks like:
growth in nominal trade = estimated time trend + 2(nominal GNP growth).

The true equation is:

$$\text{growth in real trade} = \text{true time trend} + 2(\text{real growth}) - \text{real world growth}.$$

Thus *true time trend = estimated time trend + real world growth + inflation*.

³ Tables 6.1-6.3; F & W (1995d); or F, S & W (1996).

⁴ The reported intercept term for 1992 is not comparable with the others. [It needs to be increased by the log of 1,000, to be comparable with the earlier numbers (because the trade and income data for that year differed from earlier years by a factor of 10^6 and 10^3 , respectively).] Perhaps the most reliable estimate would be the sum of the three intercept terms in the estimates based on changes in trade, from W & F (1995). It is 1.24. Adjusted for inflation, the estimate is $1.24 - (.25 + .38)1.24 = .46$, a trend increase in trade of 2.1 percent per year.