REGIONAL TRADING BLOCS

Jeff Frankel, Senior Fellow

with the assistance of
Alan Kackmeister
Ernesto Stein
and Shang-Jin Wei

COPYRIGHT 1995 BY THE INSTITUTE FOR INTERNATIONAL ECONOMICS

This book was written with the assistance of Alan Kackmeister of the IIE (Appendix 1, in particular), Shang-Jin Wei of Harvard University (Chapters 4-6), and Ernesto Stein of the InterAmerican Development Bank (Chapter 7-8). Without them, it could not have been written. Many others provided useful inputs at various stages of this project. The author would like to thank Benjamin Chui, Teresa Cyrus, Xiong Bai Fan, Jung Shik Kim, and Barbara Kotschwar for research assistance, and Warwick McKibbin, Gary Saxonhouse and Alan Winters, for supplying data. The list of those who contributed valuable comments at some stage is so long that many will have been accidentally omitted. The manuscript was reviewed at length by Jacques Polak, Jeffrey Schott, Alan Winters, Paul Wonnacott and Ronald Wonnacott. The list also includes Claude Barfield, Tamim Bayoumi, C.Fred Bergsten, Jagdish Bhagwati, Rudiger Dornbusch, Barry Eichengreen, Peter Drysdale, Jeffry Frieden, Ross Garnaut, Reuven Glick, Gene Grossman, Tim Hazledine, C. Randall Henning, Gary Hufbauer, David Hummels, Miles Kahler, Paul Krugman, Robert Lawrence, Edward Leamer, Philip Levy, Zenta Nakajima, Peter Petri, Richard Portes, J. David Richardson, David Romer and T.N. Srinivasan.
Regionalism is in fashion. Every week, it seems, brings news of yet another agreement among a group of countries, or between one group and another, to strengthen their economic links, particularly by removing barriers to trade and investment among themselves.

This is certainly not the first time in history that regionalism has been on the march. Most recently, there were widespread attempts at regional trading arrangements in the 1960s, which largely failed. But before that, there was in the 1930s a major fragmentation of the world trading system into competing blocs, which, in the standard view, succeeded only too well.

There is little point in trying to identify the earliest regional trading arrangement in history. For as long as there have been nation states with trade policies, they have discriminated in favor of some valued neighbors and against others. Regional trading arrangements have at times played major roles in political history. The German zollverein, the customs union that was formed among 18 small states in 1834, was a step on the way to the creation of the nation of Germany later in the century. This precedent has not been lost on those

---

1 That period has been called the *First Regionalism*, and the current period the *Second Regionalism* (Bhagwati, 1991).
Europeans who today wish to turn the European Union into a single nation state in the twenty-first century.

This book seeks to address three big questions regarding the current regionalism. First: how big an influence are these arrangements having on the actual patterns of trade flows? Second: within the approach of traditional customs-union theory, which takes as given the level of trade barriers between blocs, is the within-bloc liberalization good or bad? Third: after, taking into account the possible effects of regional liberalization on global liberalization, we again ask, is the trend good or bad? These three questions are addressed in Chapters 2-6, 7-9, and 10-11, respectively. There is a geographic emphasis throughout the book. Contrary to standard analyses, it matters that a regional trading arrangement links countries that are located in a common geographic region, rather than linking an arbitrary set of countries.

**MFN and Article XXIV**

It is somewhat easier to identify the historical origins of the obverse of preferential trading arrangements, the principle of non-discriminatory trade policies. The principle goes under the name of MFN policies (Most-Favored Nation). The United Kingdom adopted non-discrimination as its trade policy early in the nineteenth century, when it undertook unilateral trade liberalization in 1846
-- the famous repealing of the Corn Laws.\(^2\) The principle of non-discrimination began to spread to other countries with the Anglo-French commercial treaty of 1860. This treaty incorporated a provision that "Each of the contracting powers engages to extend to the other any favor, any privilege or diminution of tariff which either of them may grant to a third..."\(^3\) This MFN clause was then repeated in many other bilateral treaties.

Some newspaper readers today wonder why the United States decides to favor a country such as China so much as to grant it Most Favored Nation status, missing the point that this simply means treating China the same as the overwhelming majority of U.S. trading partners. It is easier to understand how such a misleading term could have arisen, once one understands that MFN clauses originally applied to a small number of trading partners.

The principle says that when a country extends trade concessions to one partner, it must extend them to all. Nineteenth-century negotiators hoped that the procedure would eliminate a potentially harmful incentive that would otherwise hamper negotiations. This was the incentive, for those who negotiated early, to withhold concessions, for fear that a later partner would get a better bargain from which the early partner would be excluded. In the event, the system worked well, as it helped reduce tariffs among an ever-growing number of countries.

After World War I, strenuous efforts of Britain and the League

\(^2\) Irwin (1993, pp. 92-95).
\(^3\) As cited by Maurel (1995).
of Nations to reinstate the MFN clause as the basis of trading arrangements were unsuccessful. The world divided into separate blocs such as the British Commonwealth, Central Europe, and others.

The victorious allies who planned the world economic system after World War II, particularly the United States, were of the view that the discriminatory trade practices in the 1930s had contributed to the collapse of world trade and in turn to the Great Depression. Accordingly the MFN principle was built into the post war trading system, in the form of Article I of the General Agreement on Tariffs and Trade (GATT). The United States opposed discriminatory tariff policies such as the British Commonwealth preferences at the time of the founding of the GATT. It soon dropped its opposition to preferences, however, in the context of European integration. The Americans considered the political desirability of peaceful European integration to be important enough to warrant making an exception to the MFN principle.

The GATT from the beginning allowed for a major sort of deviation from the MFN principle. In the form of Article XXIV, it incorporated the possibility of regional trading arrangements. This Article says that a group of countries may form a Free Trade Area or Customs Union, dropping barriers among themselves, subject to several requirements. The first requirement is that "substantially all" barriers among

---

4 Recently, an opposing view has held that regional trading arrangements in the interwar period in fact offered what little scope for trade liberalization there existed (Oye, 1992, or Maurel, 1995).
5 Irwin (1993) and Finger (1993) review the history.
the members be removed, neither stopping short at partial preferences (except in transition), nor excluding major sectors: "A free-trade area shall be understood to mean a group of two or more customs territories in which the duties and other restrictive regulations of commerce [with some exceptions...] are eliminated on substantially all the trade between the constituent territories..." The provision does not define "substantially all." The restrictions are looser in the case of preferential trading arrangements among developing countries.

The second requirement is that trade barriers against non-members not be made more restrictive. When the members go into the arrangement with different levels of tariffs against non-members, a process of averaging can go on to set the new external tariffs. If the net effect is to raise some barriers in some sectors, non-members who are adversely affected claim compensation.

Finally, subsequent progress toward economic integration is supposed to be expeditious. The relevant provision says that an agreement "shall include a plan and schedule for the formation of such a customs union or of such a free-trade area within a reasonable length of time." In the Uruguay Round negotiations concluded in 1994, this length of time was finally defined as normally not to exceed 10 years. Article XXIV does not mention a requirement of geographical contiguity or proximity, allowing such preferential arrangements as

---

6 From early on, the phrase was interpreted both as liberalization covering a high percentage of total trade (80 per cent in the case of the formation of the EEC in 1957), and as liberalization occurring in most major sectors (as in the formation of EFTA in 1960). [Analytical Index of the GATT, pp. 738, 766-68.]
the British Commonwealth, which will below be deemed "unnatural".

In practice, these provisions have been interpreted very loosely. No trading arrangement has been rejected by the GATT. In most cases, they have not even been ruled on. Later in this book (Chapter 8), we will take Article XXIV at face value, and ask whether it could be expected to raise economic welfare, as compared to an alternative world of strict MFN.

The Recent Move to Regionalism

Between 1990 and 1994, the GATT was informed of 33 regional trading arrangements, nearly a third of all deals since 1948. By now there are only a few countries, including Japan, that do not belong to any formal regional trading arrangement, according to the survey of the field taken by the World Trade Organization (the successor to the GATT), on the occasion of its inauguration. If the Asia Pacific Economic Cooperation forum (APEC) is counted as a planned regional trading arrangement of sorts, then virtually all countries now belong to at least one such club.

The surge in regional trading arrangements that has taken place over the last ten years constitutes a break with preceding post-war history. Previous regional agreements had been neither so numerous, nor so successful, as those of recent years have been. Perhaps most importantly, the United States had tended to oppose them, choosing

---

7 More precisely, reports such as those on the EEC and EFTA were neither officially adopted nor rejected by the GATT. E.g., Fieleke (1992).
8 WTO (1995, p.27).
rather to emphasize the multilateral liberalization process through the GATT. Now the United States is at the forefront of some of the largest regionalist initiatives. Why the change, and why now? Late-1980s developments in four essentially independent arenas seem to have come together to create a movement toward regionalism: European integration, U.S. strategy, developments in Canada, and attitudes toward trade in the developing world.

The influence of the European example

The European Community (EC) took a major step with the Single Market initiative, which was adopted in 1987 and took effect in 1992. The ambitious plan was to turn a free trade area into a true common market. European integration has faced some setbacks, most notably the September 1992 and August 1993 crises in the Exchange Rate Mechanism, which put a crimp in the plans for Economic and Monetary Union that had been agreed at Maastricht only nine months before. Nevertheless, the continued expansion of the EC, now the EU (European Union), in terms of scope, depth, and geographical area, is a truly historic achievement. This success has undoubtedly had a demonstration effect, encouraging emulation in the form of regional initiatives in other parts of the world.

The reversal of the American position on regionalism

The second key force underlying the global move to regionalism is American strategy. Until the 1980s, European steps toward integration had produced two reactions in the United States. First, Americans would override their instinctive aversion to regional trading arrangements by taking a dose of geopolitical medicine. The
premier motive behind European economic integration -- particularly the formation of the EEC in 1958 -- was the political one of assuring that no more wars be fought in the heart of Europe, after three Franco-German wars over the preceding century. The United States approved this logic, and indeed pushed it forward. Second, American governments would respond to European actions on the regional front -- the expansion of the EC -- by proposing a new round of liberalization negotiations in the GATT, so as to keep the momentum in the multilateral direction. This pattern played a role in the Dillon Round, Kennedy Round, and Tokyo Round.

In 1982, this pattern switched. Trade Representative William Brock encountered European resistance to American proposals at a GATT Ministerial Conference in Geneva for a new round of multilateral negotiations. He responded with regional initiatives. He let it be known that the United States was "willing to dance" with interested partners. The swift outcome was the U.S.-Israel FTA and the Caribbean Basin Initiative. (Negotiations with Canada took longer.) Where the Americans had previously reacted multilaterally to European actions on the regional front, now they were reacting regionally to European actions (or lack thereof) on the multilateral front. The logic -- "if the multilateral road is obstructed, then we will just have to explore these other roads" -- continued to hold at least up until the ultimate conclusion of the Uruguay Round in 1994.

The new willingness to consider departures from multilateralism

---

as a matter of policy, carried over from the Republican Administrations to the Democrats. Lawrence Summers (1991), later to become Bill Clinton's Deputy Secretary of the Treasury, gave what has come to be seen as a statement of Administration policy: that there should be a "rebuttable presumption in favor of all the lateral reductions in trade barriers, whether they be multi, uni, tri, plurilateral."

All the "lateralisms" are viewed benignly. In other words, any progress toward removing trade barriers is progress. This has remained the policy even after the successful conclusion and implementation of the Uruguay Round and the establishment of the World Trade Organization.\(^\text{10}\)

If one seeks a deeper historical significance to the new U.S. policy, it might be found, like so many other recent developments, in American concerns with declining economic and political hegemony.

In this view, a non-discriminatory global trading regime, enforced by an institution such as the WTO, is an international public good. That is, the non-discriminatory regime benefits everyone, but the support to initiate and maintain it will not be provided if each country acts to maximize its self-interest under the assumption that it is too small to affect the actions of others. In this view, the cooperation to support such a regime will only be forthcoming if there exists a hegemonic power that is large enough to recognize the effect its actions have on the global equilibrium, and therefore is large

\(^{10}\)The language was repeated as recently as Summers (1996, p.48): "...it is important that the march towards lower trade barriers -- unilaterally, regionally, and multilaterally -- continues." See also Council of Economic Advisers (1995, 217-231).
enough to organize the others into collective action.\textsuperscript{11}

Great Britain was the necessary hegemon in the 19th century. The absence of a hegemon in the interwar period exacerbated the collapse of trade and income in the Depression. The United States has been the necessary hegemon since World War II. Over the last half-century, however, the U.S. share of Gross World Product has fallen from nearly one-half to close to one-quarter. American perceptions of the decline in economic status are even worse than the reality, particularly in the trade arena. In response, the United States may in effect have decided to "circle the wagons," to abandon active support for the non-discriminatory regime, in favor of pursuing its own regionalist agenda. In doing so, it would be repeating the pattern of Britain fifty years earlier, when the declining hegemon instituted imperial preferences. In other words, in the face of a loss of confidence and in the face of the entry of possible new claimants to superpower status, the U.S. wishes to guarantee that its dominance of its own region is secure.

\textit{Canada enters the picture}

Israel and the Caribbean are too small to constitute in themselves a major trend in U.S. policy. Indeed, it is possible that Brock did not realize in 1982 he was initiating a long-term shift in American policy emphasis. But the U.S. frustration with the lack of multilateral progress happened to coincide with an independent development, the third link in the chain. Canada instigated

\textsuperscript{11} Seminal references include Keohane (1980), Kindleberger (1973), Krasner (1976), and Olson and Zeckhauser (1966).
negotiations on a Free Trade Area with the United States. In doing so, Canada reversed over 100 years of explicit rejection of proposals for such links with its larger neighbor. An important motive was the development of a view among Canadian businessmen that their domestic market was too small to exploit economies of scale. On the one hand, the new American willingness to consider regional agreements was certainly a necessary component of the bilateral deal, which was completed in 1988. On the other hand, the change in Canadian attitudes was also a necessary component. That the initiative came from the Canadian side is borne out by the fact that the U.S. allocation of negotiating resources was low-level and, in some key respects, last-minute.

The abandonment of the import-substitution philosophy among developing countries

The fourth component of the new regionalism, and the one that has made it a worldwide phenomenon, is the spread of serious regional initiatives to the developing countries. Most dramatically, Mexican President Carlos Salinas in 1990 reversed 150 years of active resistance to the yanqui embrace, and asked to follow the Canadian precedent in forming an FTA with the United States. The resulting North American Free Trade Agreement (NAFTA) was concluded in 1992 and ratified in 1993. But there has also been a simultaneous proliferation of regional trading arrangements among developing

---

12 And among economists; e.g., Wonnacott and Wonnacott (1967) and Harris (1984, 1991).
14 Two references are Wonnacott (1987) and Schott and Smith (1988).
countries themselves, far more serious than failed attempts along these lines in the past. In 1991 alone, Mercosur was inaugurated among four countries in the eastern half of South America (Argentina, Brazil, Paraguay and Uruguay); Venezuela and Colombia reinvigorated the Andean Pact in the western half of that continent, agreeing to establish a common market within two years; and Southeast Asian countries agreed to form the ASEAN Free Trade Area. Many more agreements followed as well, in virtually all parts of the world. They are recounted in detail in Appendix 1, and summarized in Table 1.1.

Even confined to Latin America alone, agreements are by now profuse, as Figure 1.1 illustrates.

For some developing countries, keeping a preoccupied United States engaged in their part of the world has been a motive for regional arrangements. President Salinas feared, after the fall of the Berlin Wall, that northern capital might be diverted from Latin America to Eastern Europe, and sought NAFTA in part as a way to counteract that threat. For many East Asians, APEC is a way of keeping the United States involved in the Pacific, militarily and otherwise.

There was a logical prerequisite for successful regionalism among developing countries. The prerequisite was that countries in Latin America and elsewhere had ditched the import-substitution model which had dominated thinking in the 1960s and 1970s, in favor of the market liberalization model. The origins of this sea-change are outside
the scope of the book, though the examples set by the phenomenal success of the East Asian dragons on the one hand, and by the collapse of the Soviet system on the other, were certainly key factors. Whatever explains the philosophical shift in the 1980s, it preceded the FTA craze chronologically, and was even more sweeping and fundamental as a movement.\footnote{References on the trade-and-growth literature are given in Frankel, Romer and Cyrus (1995). See also Sachs and Warner (1995).}

That outward orientation is generally a pre-requisite for successful FTAs can be seen in the failure of those attempts made by countries in Latin America and elsewhere under the import-substitution philosophy of the 1960s. As long as the goal was to find regional markets for favored industries that could not compete on world markets, and as long as each country tended to favor the same industries (steel was a typical favorite), the enterprise was doomed to failure. When the goal is to make industries better able to compete internationally, regional arrangements are more likely to succeed.\footnote{E.g., Lawrence (1996, p.5-6 in draft).}

**A trend toward continent-sized trading blocs**

In the 1990s, the talk has moved to expansion of the regional Free Trade Agreements within their respective continents. The Bush Administration proposed the possibility of bilateral deals throughout the Western Hemisphere, under the Enterprise for the Americas Initiative in June 1990. Concretely, NAFTA contains provisions to
add other Western Hemisphere countries, and Chile has been announced to be the first to join. At a hemisphere-wide summit in Miami in December 1994, President Clinton proposed, and the other leaders accepted, the idea of forming a single bloc. It was given the inelegant name, Free Trade Area of the Americas (FTAA).

A similar expansion is underway in Europe. The European Economic Area, an arrangement of (relatively) free movement in goods, services and labor comprising 17 countries, came into being in January 1994. Meanwhile, and more importantly, three new members joined the European Union in 1995 bringing its total membership to 15 (all of them also members of the EEA). Other countries from Central and Eastern Europe hope eventually to join their western neighbors in the EU. German Chancellor Helmut Kohl has told the citizens of Poland (perhaps rashly) that they will be able to join the EU by the year 2000.

Explicit FTAs are much less popular in Asia than in the Americas or Europe, but many observers see an East Asia bloc forming nonetheless. ASEAN admitted Viet Nam in 1995, and decided to admit the countries of IndoChina in the future. Malaysian Prime Minister Mahatir had already proposed expanding the group further, into an East Asian Economic Group, which has become the informal East Asian Economic Caucus. A still-less formal version of the putative East Asian bloc is the network founded on links among businesspeople who are ethnically Chinese. A third possible version is the yen bloc that Japan is reported to be establishing in the region.

All this regional activity leaves some observers concerned that
the world is dividing into three continental trading blocs, one in the Americas centered on the United States, one in Europe centered on the EU, and one in Pacific Asia, centered on Japan. This book considers regionalism at two levels: both the formal regional trading arrangements that are already in effect, and these broader continent-sized groupings that are under discussion. There should be an a priori presumption that the existing formal arrangements are more likely to have already had substantive economic effects than are the broader informal groupings, but the latter have also generated a lot of interest.\(^{17}\)

In Chapters 2 and 4-6 \textbf{[formerly 3 and 5-7]}, we examine the statistical evidence on how much effect these blocs have had on actual trade patterns, before going on to consider whether the observed trends are good or bad. We make an evaluation of the famous trade-off between trade-creation and trade-diversion operational, by parameterizing it along a geographical dimension. The geographical dimension would seem indispensable in an analysis of "regional" trading arrangements, but in the past has been relatively neglected. Traditional customs union theory pays no attention to the physical location of a union's members. The key result in the theoretical part of the study is that the desirability of regional trading arrangements depends on whether

\(^{17}\) Another possibility is that the East Asians and North America form a trans-oceanic bloc via APEC (Asia Pacific Economic Community), dividing the world into two: APEC and a Europe-centered bloc. In part to counter this possibility, some European leaders in 1995 proposed a second trans-oceanic bloc in the form of TAFTA (Trans-Atlantic Free Trade Area). We will also be considering these prospective blocs.
the extent of regionalization exceeds an optimal level that is determined by the magnitude of transportation costs between regions. This exercise, however, takes the extent of multilateral liberalization as given. Ultimately, a judgement on the desirability of regionalism hinges on whether it is more likely to help build political support for liberalization more generally, or to undermine it. The arguments are considered in the last few chapters of the book. The balance of the evidence seems to support the first view, that regional trading blocs can be stepping stones toward worldwide free trade. On these grounds, the ultimate verdict is positive.

First, some terminology.

**Terminology**

Formal regional trading agreements can cover a spectrum of arrangements, from small margins of preference in tariffs, to full-scale economic integration. Five levels can be distinguished.\(^{18}\)

**Preferential Trading Arrangement**

The loosest type of arrangement is the granting of partial preferences to a set of trading partners. As already noted, if preferences are less than 100 per cent, then the discrimination against non-members in general violates Article XXIV of the GATT. But such partial preferences are common nonetheless. If the concessions are

\(^{18}\) Concise definitions are offered by Balassa (1987).
one-way, then we term this a Preferential Trading Arrangement. Where the concessions are granted by an industrialized country to less developed countries, which is the usual case, they have been widely tolerated by the GATT. Examples include the Lome preferences toward African, Caribbean and Pacific countries granted by European countries, and the Caribbean Basin Initiative on the part of the United States.\textsuperscript{19}

If the concessions are \textit{reciprocal}, then we may apply the term \textit{Preferential Trading Area (PTA)} to describe the club of countries covered. Much of the econometric and theoretical analysis in this book will treat potential regional trade blocs as PTAs.

\textbf{Free Trade Area}

If the members of a preferential trading area go so far as to eliminate all tariffs and quantitative import restrictions among themselves (100 per cent preferences), then they form a \textit{Free Trade Area (FTA)}. Typically they retain varying levels of tariffs and other

\textsuperscript{19} The GATT grants LDCs exceptions in two areas. Preferences from other countries in the form of "more favorable treatment to developing countries," and Preferential Trading Areas in the form of "regional or global arrangements entered into amongst less-developed" countries are both allowed by the Enabling Clause of the Decision of the Contracting Parties of 28 November 1979 (paragraphs 1 and 2(c), respectively). There is no requirement for elimination of barriers on substantially all trade. The first provision forms the legal basis for the Generalized System of Preferences (GSP) and the second for such arrangements as ASEAN, LAIA, MERCOSUR, SPARTECA, and the Gulf Cooperation Council. Before 1979, EC preferences toward developing countries under the Lome Agreement were justified by Articles XXIV-XXVIII of the GATT, which sets a principle of non-reciprocity between developed and developing nations, but does not allow regional arrangements. A few other systems of preferences, such as the U.S. Caribbean Basin Initiative (CBI), have been justified under a waiver voted by a 2/3 majority, as allowed under Article XXV.
barriers against the products of non-members. Uncompetitive industries in the more highly-protected member countries may have sufficient clout to prevent their shields from being lowered to the same levels as in the less-protected member countries.

While terminology is "only a question of semantics," semantics can often have a subtle but material effect on the political debate. Words like "discriminatory" and "free trade" come with ready-made connotations. If PTAs and FTAs were routinely called "discriminatory trading arrangements," the immediate connotation might be pejorative [conjuring up images of racial discrimination].

The term "Free Trade Area" has a meritorious sound to it. Almost all professional economists are in favor of free trade in general. A good majority of the American public and its politicians consider themselves philosophically disposed in favor of it as well. Even aggressive trade policy speeches begin: "I am in favor of free trade. But...." (It is never: "Let me tell you why I am philosophically a protectionist....") Thus free trade has a positive connotation. Jagdish Bhagwati has suggested using the term Preferential Trading Area in place of FTA. He is fully aware that "Free Trade Areas" has the same positive connotation as "Free Trade," and thinks that the name should be changed precisely for this reason. While a leading proponent of true free trade, he considers FTAs (and PTAs) harmful, for reasons that will be discussed in later sections. We adhere to standard definitions here, as the distinction between partial preferences and complete elimination of barriers is important.
A bit more terminology is necessary before we leave FTAs. Those who design FTAs fear that, in the absence of special provisions, imports from outside will come into the member country with the lowest tariff, and then be re-exported to the higher-tariff members, thus obviating the goal of retaining higher barriers in the latter. The special provisions that are thought necessary to deal with this problem are called rules of origin. They generally prevent goods re-exported from the lower-tariff country to the higher-tariff country from receiving the FTA preferences. The pro-trade temptation would then be to administer a little value-added to the imported good in the low-tariff country, so that it can be re-exported in the guise of a domestic product. This could be the assembly of a product out of imported parts or, in an extreme case, a simple affixing of labels and packaging. Rules of origin prevent this, typically by specifying how much value-added is needed before the product can genuinely be classified as domestically produced. They get very complicated. The anti-trade temptation is to specify that any product with less than 99 per cent domestic value added in the low-tariff country cannot be exported to the high-tariff member.  

20 Rules of origin take up 194 pages in the NAFTA document. During the 1993 NAFTA debate in the United States, Pat Choate challenged the economists who had signed a letter supporting ratification to say whether they had read the entire document. They were unsure how to answer, since the truth in most cases was "no." A good answer might have been to agree that legal documents tend to be too long and complex these days, and to suggest that the agreement be replaced with the sentence, "There will be free trade between Mexico, the United States and Canada." Most of the rules of origin and other complicated provisions are put into such agreements for political reasons: to buy off sectors that would otherwise oppose the agreement. Without the political opposition of the anti-NAFTA forces, these provisions would not have been necessary. Some free-trade economists would be happy to have goods trans-shipped from low-tariff members to high-tariff members.
Rules of origin also allow the development of a set of FTAs that would otherwise make little sense: a *hub-and-spoke system*. This would be the outcome if the United States were to negotiate separate FTAs with a number of other countries. Such a hub-and-spoke system seemed a possibility around 1990, after the CBI, U.S.-Israel FTA and Canada-U.S. FTA were in place, and talk turned to adding new partners in Latin America and East Asia. In this case, each partner would be able to export freely to the United States, but not to the other partners. Rules of origin could turn out to prevent importers in the hub re-exporting goods from one spoke into another. As events unfolded in the early 1990s, the United States has instead chosen the route of bringing Mexico, Chile and possibly other Latin American countries into the North American FTA. But North America would be a hub with three huge spokes, if APEC, TAFTA, and FTAA all came into being. As a more modest example, Mexico currently belongs to a number of different FTAs (as Figure 1.1 shows).

**Customs Union**

The next level of integration occurs when the members of an FTA go beyond removing trade barriers among themselves, and set a common level of trade barriers vis-a-vis outsiders. This at a minimum entails a common external tariff. Under the terms of Article XXIV, as already noted, the level of the common external tariff can be no

---

higher than an average of the previously existing tariffs of the member countries, or else the group must offer compensation to adversely affected non-member countries. A full customs union would also harmonize quantitative restrictions, export subsidies, and other trade distortions. Indeed, it would set all trade policy for its members as a unified whole. It would, for example, engage in any future trade negotiations with other countries with a single voice.

The distinction between trade creation and trade diversion is crucial to the classic theory of the welfare effects of customs unions, while also applicable to other kinds of preferential trading arrangements. We shall not review the literature here. Despite the ubiquity of the terms, they do not have standard definitions that are entirely agreed upon. We will use the term trade creation to describe an increase in trade between members of a PTA that occurs as the result of the preferences, and trade-diversion to refer to the decrease in trade between members and non-members that occurs as the result of the preferences.

The presumption that trade creation and trade diversion can be identified with the good and bad effects of PTAs, respectively, is oversimplified for a number of reasons. Trade diversion can under certain circumstances be beneficial. This can happen, for example, when the new availability of imports from fellow members reduces distortions in consumers' patterns of consumption. It can also

---

22 The seminal reference was Viner (1950), followed by Meade (1955) and Lipsey (1960). Reviews are available in many places, such as Bhagwati and Srinivasan (1983) and Bhagwati and Panagariya (1995), and Wonnacott and Lutz (1989).

23 Meade (1955), Gehrels (1956), and Lipsey (1957).
happen when economies of scale allow producers to operate at lower cost.²⁴ Or when the new competition reduces the market power of inefficient domestic monopolies. Some economists have defined trade diversion so as to rule out such beneficial effects, but others reasonably point out that this flirtation with tautology would make the terms less useful.²⁵ The distinction between trade diversion and trade creation also omits other important effects, particularly those on the terms of trade, the relative prices of countries' exports. This effect is beneficial for members of the PTA, but harmful to non-members.²⁶

**Common Market**

Each of the stages considered so far — PTA, FTA, and Customs Union — falls within a range that the Brookings Institution has recently characterized as shallow integration. The direct effects of such arrangements lie exclusively within the realm of international trade. More advanced stages, which we now consider, constitute deep integration.²⁷

A Common Market entails, beyond the free exchange of goods and services among members, also the free movement of factors of production: labor and capital. The dividing line is admittedly sometimes blurred between the free exchange of services and the free movement of factors: labor includes services such as construction

---

²⁴ Corden (1972) and Venables (1987).
²⁶ Mundell (1964).
²⁷ Lawrence (1995) and Lawrence, Bressand and Ito (1995).
or consulting, and capital includes banking and other financial services. The free movement of capital applies to portfolio capital as well as to foreign direct investment (FDI): the purchase and sale across national boundaries of real estate, plant, and equipment. This is deep integration, in that it impacts some laws and institutions that could have been preserved as domestic prerogatives even with a high level of trade integration. Examples include policies governing the licensing of providers of professional services, such as doctors, lawyer, architects, etc., and anti-trust or competition policy. Migration is of course an especially difficult and sensitive subject in many countries.

**Economic Union**

Going beyond the free movement of goods, services, and factors, Economic Union involves the harmonization of national economic policies, including typically taxes and a common currency. Belgium and Luxembourg formed an economic union in 1921. The decision of the European Community to change its name to the European Union in 1994 represented a determination to proceed to this higher stage of integration. The full unification of economic policies typically would in turn require political federation.

Country groupings often choose names that are far too ambitious, or are otherwise unsuited, for the kind of integration or cooperation that they are actually prepared to undertake. Neither NAFTA nor (especially) the ASEAN FTA expect literally to eliminate all interior
barriers to trade as one would think from the appellation "Free Trade Area." The European Community adopted the phrase "Common Market," as a goal in the Treaty of Rome, long before the acceptance of serious plans to integrate trade and factor movements fully. Typical of more extreme examples is the Central American Common Market, which has yet to aspire to become an FTA, let alone a customs union or common market. Nevertheless, in this book we shall allow groupings to call themselves whatever they choose.
CHAPTER 2

RECENT HISTORY OF REGIONAL TRADING ARRANGEMENTS

THIS CHAPTER

(DRAFTED BY ALAN KACKMEISTER)

HAS NOW BEEN CHANGED INTO AN APPENDIX.

THE CHAPTER NUMBERING has been CHANGED, as indicated (6/14/96).

...
CHAPTER 2 [PREVIOUSLY 3]
SIMPLE MEASURES OF REGIONAL CONCENTRATION IN TRADE

Global trade has expanded steadily. During the postwar period, worldwide trade has grown much faster than income. This phenomenon is reproduced in most countries as an increase in a simple measure of openness, the ratio of trade to GDP. The most obvious explanations for this trend include secularly declining transportation costs and secularly declining tariffs. Our econometric tests will shed light on these two key factors, as well as on a third factor: poor countries trade more as they become richer.

Has the growth in trade been biased toward trade within regions, and away from trade between regions? This would constitute a movement toward regional trading blocs. It is the central question for the next few chapters of this book. This chapter begins the analysis, by reviewing the measures that are in most common use to gauge tendencies toward regional concentration in trade.

The preceding chapter recounted the recent history of explicit preferential trading arrangements. Nominal or explicit preferential trading arrangements and operational regional trading blocs are not necessarily the same thing, however. On the one hand, one can't be sure that explicit proclamations of regional trading arrangements

---

28 See Rose (1991) or Krugman (1995). This was also true over the two centuries preceding World War I (e.g., Council of Economic Advisers, 1992, p.194); the period 1914-1945 was the exception.
de jure are carried out de facto. As is clear from the review of recent regional trading arrangements in Appendix 1, there is a long history of grandly proclaimed agreements falling short in practice. It is a story of failure to translate visions into specific plans, of delays in implementation, of rampant sectoral exclusions or escape clauses, and of poor enforcement of nominal agreements. This is particularly true among less developed countries. For example, the implementation record of ECOWAS (the Economic Community of West African States, set up in 1975 with the objective of becoming a customs union) has been rated "progress negligible" over the subsequent two decades. Similar ratings have been given to LAFTA, WACU, CACM, and EACM. ASEAN and the Andean Pact repeatedly postponed their plans for regional integration in the 1970s and 80s.29

Conversely, even where explicit preferential trade arrangements do not exist, some observers see the same effect as if they did, through other means. Such means include foreign direct investment (FDI), overseas development assistance (ODA), currency links, and even political intervention or threats of military force. These tools are thought to have the effect of distorting and re-directing trade toward the regional hegemon. Certainly the hypothesis put forward by heralds of an emerging yen bloc in East Asia, or modern "Greater Asia Co-Prosperity Sphere," is not that Japan has organized an explicit Free Trade Area. To the contrary, Japan is the only industrialized country that does not have reciprocal preferential trading

29 E.g., de la Torre and Kelly (1992, 26) and Balassa (1987). Hoekman and Leidy (1993) detail loopholes that typify this history, and other "slips twixt cup and lip."
arrangements with any neighbors. Rather the hypothesis is that Japan is forming an economic bloc in the same way that it runs its economy: by means of policies that are implicit, indirect, and invisible. Earlier opponents of U.S. "imperialism" in Latin America had similar views.

In this study, we will consider both kinds of groupings: more narrowly defined groups of countries representing formal regional trading arrangements, as well as broader geographical groupings where trade has been alleged to have been artificially concentrated. Under the narrow definition, we look chiefly at NAFTA, Mercosur, the Andean Pact, ASEAN, and the EC and EFTA (for the 1970s and 1980s) or the EU (for today). Under the broader definition, we look at Western Europe, the Western Hemisphere, and East Asia. Momentum for expansion of the EC into the EEA and EU and the expansion of NAFTA into the FTAA provides some motivation for looking at these continent-scale blocs. Finally, we also consider two potential transoceanic groupings: APEC and TAFTA.

---

Intra-Regional Trade Shares, and What is Wrong With Them

We want to know what effect all these impressive-sounding acronyms have had on the actual pattern of trade. Many studies of regional trading arrangements attempt to document effects on trade by reporting simple statistics on shares of intra-regional trade. The denominator in this ratio, for a member of a particular grouping, is its total trade, and the numerator is the subset of that trade it undertakes with others members of the grouping. This is supposed to reflect tendencies for countries to direct their trade toward other members of their grouping. We begin our exploration with this measure, though we shall argue that it is unserviceably primitive.

The intra-regional trade shares are reported in Table 2.1 [formerly 3.1].\(^{31}\) The trade shares are also graphed in Figures 2.1a (depicting the formal FTAs) and 2.1b (for the larger continental groupings). The most serious problem with this measure shows up right away. The share is very high for large groupings like APEC, Western Europe or TAFTA, and very low for small groupings like the Andean Pact countries. Is this because APEC has been very successful at promoting trade among its members and the Andean Pact has not? No, not necessarily. Rather, it reflects primarily that APEC is a large

\(^{31}\) Tables 2.1 and 2.2 [formerly 3.1 and 3.2] pertain to a relatively large data set of 163 countries (implying 26,406 import pairs plus export pairs if all were reported: 163x162). Estimates of the gravity model of bilateral trade that are reported in later chapters pertain to a narrower data set of 63 countries (implying 3,906 pairs, when all are reported: 63x62). Data for this set is more consistently available. For comparability with the gravity coefficients discussed in later chapters, the versions of tables 2.1 and 2.2 that correspond to this narrower data set of 63 countries are reported in a statistical appendix (Tables A2.1 and A2.2 [formerly A3.1 and A3.2]). The intra-regional concentration ratios are generally within 4 per cent of those reported in Table 2.2 [formerly 3.2] here.
group of countries, both in the sense of the number of members that belong and in the sense that many of them are quite large trading countries, while the Andean Pact represents a relatively small group of small countries. It is a necessary property of the intra-regional share measure that the bigger the set of countries around which one throws the lasso, the higher will be the apparent concentration of trade within. In the limit, if one throws the lasso around all countries of planet Earth, one would find a ratio of 100 percent.

[Table 2.1 [formerly 3.1] about here]

[Figures 2.1a and 2.1b [formerly 3.1] about here]

To illustrate the point, consider an arbitrary set of countries: those having names that do not begin with the letter F. This large group of countries conducts 94 per cent of their trade with each other, as compared to the three countries that do begin with the letter F (Fiji, Finland and France), who conduct only 1 per cent of their trade with other members of their group [in 1994]. This difference reflects nothing more than that only a small part of the world's trade is accounted for by the second group of countries. One would not want to conclude from these two statistics that countries that do not begin with F are more inclined to preferential trading arrangements with each other than countries that do begin with F.

The fallacy often arises in comparisons across different groups. Boosters of APEC, for example, sometimes point to its intra-regional share of 71 per cent, forgetting to acknowledge that members of any grouping that comprises a large share of the world economy will
inevitably trade a lot with each other. Only after one takes into account APEC’s share of world trade (41 per cent) can one consider its intra-regional trade to be noteworthy.

The fallacy arises even more often in comparisons across time. The shares are often used to illustrate the increasing intra-regional bias in world trade. It is pointed out that the intra-regional trade share among western European countries increased substantially in 1965, when six were in the European Economic Community, and increased further after 1973, when Denmark, the United Kingdom and Ireland joined. As the table shows, the EC Twelve's ratio of internal trade to total trade rose from 49 per cent in 1962 to 60 per cent in 1990, by which time the accessions of Greece, Portugal, and Spain were complete. In the early 1990s, however, the intra-EC trade ratio levelled off, showing no evidence of the effect of the Single Europe Act that went into force in 1992. This might be puzzling if one considered the intra-regional trade ratio to be a good measure of the effects of regional trading arrangements.

Similarly, it is pointed out that regional groupings in the Western Hemisphere also have shown a bit of an upward trend in intra-regional trade shares, and that the intra-regional trade share in East Asia has increased the most rapidly of all, by as much as 1/2 between 1970 and 1990, as Table 2.1 shows. The intra-regional

---

32 Early studies of the effect of the original formation of the EEC and EFTA on trade shares were reviewed and improved upon by Williamson and Bottrill (1971). (See also Truman, 1969, and Kreinen, 1974.)

33 If one restricts the data set to the 63 countries used in the gavity estimation of later chapters, then the ratio rose from 53 per cent in 1962 to 64 per cent in 1990. Appendix Table 2.1.

34 E.g., Table 1 in Frankel, Wei and Stein (1994) or Table 1 in Frankel, Stein and Wei (1995).
trade share in East Asia reached .50 in 1994 (versus .74 in Europe and .50 in the Western Hemisphere). This rapid trend is cited as evidence that Japan is building a trade bloc in East Asia, even without explicit policy steps toward a preferential trading area. We shall see that these inferences regarding the speed with which trade is becoming intra-regionally concentrated in Europe and Asia are incorrect.

It is worth noting that levels of intra-regional trade, as reflected for example in Table 2.1 [formerly 3.1], are indeed useful for some purposes. Let us say we are interested not in the effects of preferential tariffs and other policy determinants on bilateral trade patterns, but rather in the effects of bilateral trade. Such effects would be of interest, for example, to businesspeople, macroeconomists, and political scientists. Then it would be perfectly appropriate to look at the intra-regional trade shares.

A businessperson, particularly one in a trade-related industry like shipping, might want to know in what parts of the world is bilateral trade increasing the most rapidly, so that he or she can plan where to invest. A macroeconomist might want to know the sensitivity of a particular small Asian economy to sudden cyclical fluctuations emanating from the United States or Japan, which depends importantly on the magnitude of its trade links with these two countries. The old principle that East Asian economies are highly dependent on North American growth is rapidly becoming less true,

[There the share was defined as intraregional trade divided by total trade involving countries of the region, which is slightly different that the ratio reported in Table 3.1.]
as trade within Asia becomes more important. For that matter, Japan is itself declining in importance, as compared to a "third growth pole" on the East Asian mainland.

Intra-regional trade may have important political implications as well. Hirschman (1980), in a classic study, pointed out the international influence that arises from trade. In time of political or military conflict, a country may be reluctant to side against a large trade partner. Hirschman made it clear that the trade need not be the outcome of a preferential trading arrangement. "For the political or power implications of trade to exist and to make themselves felt, it is not essential that the state should exercise positive action, i.e., organize and direct trade centrally; the negative right of veto on trade with which every sovereign state is invested is quite sufficient" (p.16-17). Thus to observe from Table 2.1 [formerly 3.1] that intra-regional trade shares for groupings that include such large countries as the United States and Japan will necessarily be large, is to observe accurately that the United States and Japan are powerful players. To repeat the central objection to the trade shares, however, they cannot be used to assess whether trade is in any meaningful sense necessarily concentrated or biased, toward the United States or Japan, or toward all the members of APEC, beyond what would be expected from the size of these countries.

Concentration or Intensity Ratios

To obtain a usable measure of regional concentration, we need to adjust the intra-regional trade shares by a measure of each group's
importance in world trade. We want to know if a typical member of group \( i \) trades more with other members of group \( i \), than does a typical country located anywhere in the world. The simplest way to accomplish this is to divide each intra-regional trade share in Table 2.1 [formerly 3.1] by that region's share of world trade. We shall call such numbers simple concentration ratios. The intuitive idea is that if bilateral trade takes place in geographic patterns that are simply proportionate to the distribution of total trade, then the concentration ratio should be close to 1. If trade is concentrated within a given grouping of countries, that grouping should show a ratio in excess of 1. Petri (1993, p.23) calls this ratio the "double-relative" measure of intensity, to indicate that bilateral trade has been deflated both by the total trade of the importing country and the total trade of the exporting country. (In a time series context, one also divides by total world trade, so that intensity is not affected by uniform worldwide inflation, or by uniform growth in real trade.)\(^{35}\) The statistics are illustrated in Figures 2.2a and 2.2b.

\(^{35}\) Petri (1991, 1993) also notes that our concentration ratio (his "double-relative"), has in the past literature commonly been called a gravity coefficient. We think that the term "gravity" should be reserved for a measure that substitutes GNP for total trade as a measure of size, and preferably takes into account the role of distance as well. Both of these variables are crucial to the gravity model and to the rest of this book. We shall reserve the term "gravity" for an equation that includes these terms, and other determinants of bilateral trade as well, and that estimates their coefficients rather than constraining them. It will turn out that the total estimated coefficients on GNPs will in fact be insignificantly different from one. This finding would help justify computing a "double relative" that deflated bilateral trade by both partner's GNPs. The estimated coefficients on distance, population, and the other variables, however, are clearly very different from zero. This shows the limitations of all the measures based on simple ratios, if our goal is to detect true bias in trade patterns.
As observed in the footnote, an alternative correction would be to divide each bilateral trade quantity by both partners' total GNPs (Gross National Products), instead of their total trade. Adjusting by total trade rather than total GNP, however, has the advantage that at the same time as taking out the unwanted influence of size, it also takes out the unwanted influence of openness. We want to know whether trade of a given country with fellow group members is high relative to non-group members, not whether all its trade is high.

There is another measure, which is closely related to the concentration ratio or double-relative, and has been called simply the intensity coefficient, by some Australian economists in particular. When applied to a single pair of countries, the measure is the same as what we have called the simple concentration ratio. But when evaluating intra-regional trade within a group of three or more countries, there is an issue as to how to aggregate trade among the pairs. Their intensity coefficient is an improvement in that it makes an adjustment in the denominator of the concentration ratio. We wish to know whether a given country trades more with a fellow bloc member than one would expect from the latter's weight in world trade. To allow this assessment, the intensity coefficient

---

\(^{36}\) Anderson and Norheim (1993a,b), Drysdale (1988), and Drysdale and Garnaut (1982, 1992).
deletes from the second country's weight in world trade its trade with the first country. The reason is that the numerator does not count the first country's trade with itself, and one wishes to make the numerator and denominator comparable. The resulting measures of intra-regional trade intensity are reported in Table 2.3 [formerly Table 2.3, formerly 3.2], under the name "corrected concentration ratios".  

These authors also go on to use more sophisticated measures designed to adjust for trade complementarity by commodity, presumably dictated by innate comparative advantage, and for openness. We shall not pursue these extensions, being anxious to proceed to the full-fledged gravity equation in later chapters. This in part reflects a suspicion that the differentiated products view of trade, overall, offers a better description of aggregate patterns than do classic theories of comparative advantage. It also reflects a strong belief that distance and the other factors in the gravity equation are essential to uncovering evidence of true bias in trade patterns, which requires estimation of the parameters in an equation, instead of computation of ratios. It should be noted, however, that some of the important lessons to emerge from our econometrics are the same that the Australian group and others have uncovered with their

---

37 Anderson and Norheim (1993) and Drysdale and Garnaut (1982, 1992) explain the intensity coefficient at greater length. Their adjustment in the denominator gets more complicated when the intensities of members of a region are aggregated, as they are here. (Although a country's trade with itself does not appear in international trade statistics, international trade among members of the region does.)
intensity ratios. This includes the lesson that intra-regional concentration in East Asia was slightly lower in 1994 than in the 1950s, 1960s or 1970s, not sharply higher as widely assumed from the simple trade shares.\textsuperscript{38}

**The Proximity School versus the Discriminatory Policies School**

The most striking thing to see in Table 2.2 [formerly 3.2] is that almost all the numbers are greater than 1. This reflects that members of each group trade more with each other than would random pairs of countries. The clear conclusion is that trade is geographically concentrated.

Various economists have observed this fact, and have drawn varying inferences from it. The key difference in interpretation centers on whether the evident regional concentration in trade should be attributed to the natural factor of geographical proximity, or to the artificial factor of preferential trade policy. Two eminent economists, while admitting that existing trade policies must play a role in such statistics, have asserted that the dominant explanation for the high concentration ratios must be geographical proximity (Krugman, 1991b, 19-20, and Summers, 1991, 297-299). Krugman adjusts bilateral trade among a small number of large countries for national incomes, and then comments on the resulting estimates: "To nobody's surprise, they point out very strongly the local bias of trade: the United States and Canada, according to the regression, do thirteen times as much trade as they would if

\textsuperscript{38} In addition to the authors cited in the preceding two footnotes, others who have come to the same conclusion using concentration or intensity ratios include Yamazawa, Hirata and Yokota (1991), Petri (1991, 1993), and Goto and Hamada (1994).
they were not neighbors, while the four major European countries do seven times as much. Of course, these results are in part due to the fact that there are already special trading arrangements between the United States and Canada, on one side, and within the EC on the other. Yet the results are so strong that they make overwhelmingly clear that distance still matters and still creates natural trading blocs."

At the opposite extreme, also-eminent economists have dismissed the role of geographical proximity, and asserted that therefore the explanation for the observed concentration must be existing discriminatory trading arrangements (Bhagwati (1992, 1993a), Panagariya (1995, pp. 9-10)). Bhagwati (1992, p.544-545) describes as wrong-headed, Krugman's proposition "...that countries sharing borders, or closer geographically to one another, have higher proportions of trade with one another than countries further apart do...If I had access to captive research assistance and funds, I could examine whether, for all conceivable combinations of countries and distances among them, and for several different time periods, the premise is valid. As I do not, I must rely on casual empiricism and a priori arguments. Compare for instance the trade through the 1960s between India and Pakistan with that between India and the UK or the USSR. The former trade has been smaller than the latter. Borders can breed hostility and undermine trade, just as alliances among distant countries can promote trade..."

This issue is important, because each of the two camps engages in a line of reasoning that runs from the positive statements, regarding the effect of policy on trade, to normative statements regarding the desirability of regional trading arrangements. Chapter 8 of this book will explain the logical connection between trade determinants and policy prescription, in developing the notion of natural and supernatural trade blocs. The central task for the book will then be to ascertain econometrically which view seems to be closest to right. Fortunately, one need not rely on "casual
empiricism". It is possible to quantify the extent to which intra-regional concentration is attributable to proximity, as in the Krugman-Summers view, versus existing preferences, as in the Bhagwati-Panagariya view.

In Chapters 4-6 [formerly 5-7] we shall adjust the bilateral trade figures for the effects of geographical proximity and other non-policy variables that naturally link countries. In this way we hope to isolate the effects on trade of preferential trading policies in different parts of the world.

Even before we turn to the full model, however, we can draw from the information we already have one tentative sort of lesson: conclusions from comparisons of regionalization across time. The missing factors such as distance, common borders, and common language do not change over time. (We abstract here from such slow processes as continental drift. More seriously, per capita GNPs do change non-negligibly during the course of our sample. But their effect is unlikely to alter major conclusions, as can be verified when we come to the full analysis.)

Thus it is legitimate to inspect the concentration ratios and infer that trade is indeed becoming more regionalized over time in some parts of the world. Among the EC 12, concentration has increased over every half-decade since 1965. The largest jumps came in the late 1960s, shortly before the United Kingdom, Ireland and Denmark joined (the agreement did not formally take force until January 1, 1973), and the early 1980s, during the drawn-out period of accession for Greece, Spain, and Portugal (the dates of accession were 1981
for Greece and 1986 for the Iberians). Interestingly, the greatest increase in intra-regional trade concentration often seems to take place after an agreement has been decided, but before it actually takes effect. Firms rush to open business, in order to establish a stake in what they expect to be an important market in the future.

The concentration ratios for the PTAs are graphed in Figure 2.1 [formerly 3.1], and for the larger geographic groupings in Figure 2.2. As Table 2.2 [formerly 3.2] and Figure 2.1 show, the intra-EC concentration continued to rise during the first half of the 1990s, the period when the Single Market Initiative went into effect. This result is more sensible than that from the simple intra-regional trade shares. The pattern is quite similar if one looks at the current EU 15. EFTA shows a comparably high level of concentration as the EC, but without any positive trend. Moving to the continental level of aggregation, Western Europe shows an upward trend similar to that of the EC or EU, especially during 1992-94.

When we turn to the formal regional trading arrangements in the Western Hemisphere, we see an illustration of the importance of correcting for the size of the countries in question. Whereas the Andean Pact, Mercosur, and NAFTA have much lower amounts of intra-regional trade than does the EC or EU (refer back to Table 2.1 [formerly 3.1]), we get a very different answer when we normalize for their shares of trade with the rest of the world. Mercosur has the highest concentration ratio of any group examined. As noted,
we cannot distinguish from this statistic whether this concentration of trade in the eastern half of South America is due to proximity or to the emerging Customs Union. (That the concentration ratio for these four countries was also very high in the 1960s, long before Mercosur, suggests that proximity plays a large role.) More meaningfully, Mercosur also has the concentration ratio that increased by far the most rapidly between 1990 and 1992. The Andean Pact also shows a concentration ratio that is high and rapidly increasing in recent years, reaching in 1994 a level almost as high as Mercosur. The NAFTA coefficient is lower, and shows little upward trend. Thus we learn from Table 3.2 that the increase in intra-NAFTA trade evident in Table 2.1 [formerly 3.1] can be mostly explained by a general increase in trade on the part of North American countries. At the hemispheric level, the level and trend of an Americas bloc looks much like NAFTA, reflecting the importance of the North American countries in the region.

Turning to Asia, ASEAN shows a level of concentration that is high, but has actually decreased in recent years. When the grouping is expanded to the East Asian Economic Caucus (EAEC), or to the entire [15-country] membership of APEC, the concentration ratio again shows a slight downward trend over time. The membership of the EAEC grouping is the same as the membership of the yen bloc in most popular characterizations. Thus, even on the basis of statistics no more sophisticated than those in Table 2.2 [formerly 3.2], we begin to suspect that fears of an intensifying yen bloc in East Asia are overblown. Yes, Japan trades far more with Malaysia than it did 20
42

years ago. But there is nothing surprising in this. Almost every country in the world trades more with Malaysia than it did 20 years ago, because that country has grown so rapidly.\footnote{This conclusion was the basis of an earlier essay, Frankel (1991), and has been confirmed by various authors, including in the first three chapters of Frankel and Kahler (1993).}

The findings for Asia and the Pacific are also very different from what the simple intra-regional trade shares showed, which was a strong upward trend. Correcting for the importance of the region in world trade makes a big difference, just as it did for the South American groupings. In the case of the Asian groupings, the correction makes a difference because the rapid growth of intra-regional trade can be seen to be nothing but the outcome of the rapid income growth of the countries. In the case of the South American groupings, it makes a difference because, once one takes into account that these countries \textit{did not} grow rapidly during this period, the modest growth in their intra-regional trade becomes quite impressive.

The concentration ratio for Africa starts out much lower than for the other continents, but shows an apparent steep upward trend. Again, what would otherwise look like very slow growth in intra-African trade is more impressive when one adjusts for the even slower growth in the trade or GNPs of these African countries.\footnote{We must note that the data set omits more countries in Africa than in other parts of the world, so that the statistics are not as reliable.}

Each of these conclusions, in addition to others, will be revealed with greater reliability and completeness when we apply the full econometric analysis of subsequent chapters. The full analysis will
require first an excursion into the metaphysics of distance, followed by the econometrics of the gravity model.

One might ask, why all the bother? To discover whether bilateral trade patterns are biased by policy, why not just look at explicit agreements, tariff preferences, and so forth? In the first place, as noted, many nominal agreements haven't materialized in reality, and vice versa. We have already mentioned the hypothesized effect of implicit or invisible policies, such as ODA and FDI on the part of Japan in Asia or politico-military influence on the part of the United States in the Western Hemisphere. Artificial trade links were obvious in the old Soviet bloc, yet they could hardly be measured statistically by tariff preferences.

In the second place, even in the case of formal preferential trading arrangements, systematic and comprehensive data on bilateral tariffs are hard to get.

In the third place, the web of preferences is often complicated, varying from one commodity to the next. It is not clear how to aggregate across commodities. If the Common Agricultural Policy in the European Community is observed to discriminate against agricultural imports from non-members, should that get a low weight in the aggregate measure of intra-regional trade bias because agriculture goods have an (artificially) low weight in EC imports?

Furthermore, non-tariff barriers, which are increasingly important relative to tariffs, are notoriously hard to measure. More favorable agricultural quotas on favored partners could in principle be observed, it is true. Preferential procurement policies or a greater
reluctance to apply administered protection (such Anti-Dumping duties) against fellow bloc members, however, would be much harder to observe and aggregate.

CHAPTER 3 [formerly 4]: THE IMPORTANCE OF GEOGRAPHICAL PROXIMITY IN TRADE

It is a commonplace of international economics that the world is getting smaller. Indeed, it is the commonplace of international economics.

The proposition that technological progress has long driven costs of transportation and communication steadily lower is of course true. One has only to mention such important technical advances of a century ago as the automobile and airplane. In recent decades we have seen further cost-saving advances in shipping: supertankers, roll-on-roll-off ships, and containerized cargo. Between 1920 and 1990 the average ocean freight and port charges per short ton of U.S. import and export cargo fell from $95 to $29 (in 1990 dollars). Between 1930 and 1990, average air transport revenue per passenger mile fell from $0.68 to $0.11. Over this period the cost of a 3-minute telephone call from New York to London fell from $244.65 to $3.32. Where shipping costs were once a large fraction of the eventual price of an item, by 1988 the $200 price of a VCR imported to the United

States from Japan included less than one percent allocated to ocean transportation. Jet air shipping and refrigeration have changed the status of goods that had previously been classified altogether as not tradeable internationally. Now fresh-cut flowers, perishable broccoli and strawberries, live lobsters, and even ice cream are sent between continents. The continuing communications revolution, including faxes and the internet, speaks for itself.

In large part because of the shrinking costs of doing business at a distance, people are very aware of the economic effects that far-off events can have on them. An oil price shock originating in the Middle East, an upsurge in cheap labor-intensive exports from China, a retrenchment by banks in Japan, financial volatility in London, a debt crisis in Mexico... each has a tangible impact on the livelihood of citizens in countries, such as the United States, that had previously thought themselves relatively immune from foreign developments.

It is easy to carry the theme of a shrinking world too far, however. Book titles such as "The Borderless World," "The Death of Distance," and "Global Financial Integration: The End of Geography" are probably not intended literally. But if the titles are intended...
literally, then they carry the theme too far. Distance and national borders both still matter, as we shall document below.

One way of measuring the remaining obstacles to international integration is to look at the inability of arbitrage to equate prices internationally. Even a commodity like cocoa or copper delivered in London is not the same commodity as cocoa or copper delivered in Chicago; prices in the two markets reflect a difference. For most goods, which are far less easily traded than basic agricultural and mineral products such as cocoa or copper, integration is much more incomplete. The obstacles to goods market arbitrage across countries are so large that tests of the standard criterion of international integration routinely were, until relatively recently, unable to find any statistical tendency for relative prices to move in the direction called for (purchasing power parity). Even now, consensus estimates say that prices on average close only about 20 per cent of a gap in a given year.\textsuperscript{44} A recent historical study of prices in England and Holland since the year 1273 for eight specific commodities (barley, butter, cheese, eggs, oats, peas, silver and wheat) reveals a surprising conclusion: the deviations from the so-called Law of One Price are no smaller or less persistent now than they were in past centuries. Evidently, whatever tendency technological progress has had to enhance international arbitrage in these commodities has been logic that made them useful lines of demarcation in the first place.” (The latter two titles have been taken somewhat out of context. They are actually about telecommunications and financial services, respectively, not merchandise trade. Even in financial markets, however, integration is not literally perfect, as is borne out by empirical evidence of international differences in interest rates and of home country bias in the holding of securities.)

\textsuperscript{44} Frankel and Rose (1996) give further references.
counteracted by greater roles for government trade barriers or exchange rate volatility (Froot, Kim, and Rogoff, 1995).

The same statistic that most readily illustrates the increased openness of most countries, the ratio of trade to GNP, also illustrates how far we in fact are from a world without borders. U.S. merchandise exports and imports (averaged) have doubled as a share of GDP since 1971, to 8.4% in 1993, versus 4.2 per cent then. The openness numbers for Japan and the European Union [all fifteen countries, netting out intra-EU trade] are comparable: 7.2 per cent and 8.5 per cent, respectively. Does the increase in these figures mean we are close to a world in which people buy and sell goods from abroad as easily at home? Do businessmen trade around the globe as easily as around the block? Far from it. The United States is 26 per cent of Gross World Product, Japan is 15 per cent, and the EU is 30 percent (evaluated at recent exchange rates). If purchases and sales took place in simple proportion to income and output, unaffected by geography, then U.S. openness would be 74 per cent (=1-.26), EU openness would be 70 percent (=1-.30) and Japan's 85 per cent (=1-.15). But we just noted that these openness statistics are in fact about 8 per cent in each of the three parts of the world. By this measure, then, openness would have to increase by roughly a factor of ten, before we could claim a fully integrated world economy! Clearly, geography still matters.

3.1 Why Geography Should Be Part of Trade Theory

As surprising as it sounds to say it, most international economists until quite recently ignored distance and other
geographical factors as determinants of trade. Trade models vary along many dimensions: Some are empirical, and some theoretical; some are designed to describe a number of large countries, and some to describe one small country; some assume imperfect competition among firms, and some perfect competition; some are based on economies of scale as the motive for trade, and some on differences in factor endowments and other sources of inherent comparative advantage. They vary among other dimensions as well. Yet most trade models have until recently had one thing in common: they treated countries as disembodied entities that lacked a physical location in geographical space.

We will see that one cannot get very far into an empirical analysis of bilateral trade, that is, an analysis of trade between pairs of countries, without recognizing the strong inhibiting effect of distance on trade. How then could so much of earlier research have omitted this obvious factor? The simplest answer is that earlier research did not try to explain quantities or composition of bilateral trade. Its goal, rather, was to explain the quantities and composition of the total trade undertaken by a country, regardless with what trading partners. Classic trade theory seeks to predict, for example, what products China will export (answer: textiles, because they are intensive in unskilled labor, which is abundant in China). It also seeks to predict what products Japan will import (again textiles, because Japan no longer has much unskilled labor).

The classic trade theory does not usually pay attention to the question whether textiles are exported specifically from China to Japan. Rather, it is content with predicting that China adds a certain
quantity of textiles to the world's supply pool, and that Japan buys a certain quantity from that textiles pool. As to whether China exports its goods specifically to Japan or France, and as to whether Japan buys from China or Morocco, the theory is silent.

A deeper answer to the question why earlier research was not interested in explaining bilateral trade, or in addressing any geographical dimensions of trade, has been given by Paul Krugman. His point is that, until recently, the standard theories of trade were based on the assumptions of perfect competition and constant returns to scale. It is difficult to analyze many geographic influences with such models. The assumptions of imperfect competition and increasing returns to scale are needed for modeling questions such as where industries choose to locate. We have all heard about the person who loses a silver dollar one night, and looks for it on the next street over from where he lost it, because that block is the only one that has bright street lights. Similarly, international economists neglected to investigate the important questions of geography because they lacked the analytical tools.\textsuperscript{45}

Why should we care about the geographical dimension to trade? There are at least three sorts of reasons why we should care about bilateral patterns of trade, the role of distance in particular, and geography in general.

Let us first note a spurious reason. We are not, it is to be

\textsuperscript{45} Krugman (1991, 1995b, 1996). The model that we develop in Chapter 7 [formerly 8] is Krugmanic.
hoped, interested because we think that bilateral trade balances are important, except in the narrow political sense that they are important because the electorate thinks they are important. It is true that a U.S. trade deficit is a subject of concern to the extent that the excess of spending over income that this deficit represents does not have a worthwhile origin. A high trade surplus in Japan might also be a subject of concern (unless one thinks that the high saving rate that brought it about is a good thing). Even so, the bilateral trade balance between Japan and the United States is a needlessly inaccurate reflection of their respective overall balances. The overall U.S. deficit and Japanese surplus are the only valid subjects of interest.

To think otherwise is to think it important that each source of spending match each source of income. It would be as if one refused to pay one's drycleaner in money, and insisted instead that he accept payment in the form of the services of one's own profession. Life would be rather inconvenient, having to search for a drycleaner willing to do business under such terms. Better to sell one's services where they are wanted, and use the proceeds to pay the drycleaner.

Even worse than concerns in policy-making circles with bilateral balances overall, are concerns with bilateral balances in specific products. Some countries object when their citizens buy more automobiles from a particular trading partner than the citizens of the foreign country buy from domestic producers. But important gains from trade are to be had by trading in different products. Life would be particularly inconvenient if one gave up the option of taking shirts to the drycleaner, beyond the number of the cleaner's shirts that
one was prepared to wash in return.

The first valid reason that we care about the role of geography is that distance leads to agglomeration. Agglomeration in turn means that history can affect what goods are produced in a given region or country. Consider the case of an intermediate input. If there is a cost to doing business at a distance, there will be a tendency for customers to locate close to a supplier of the input, and then for other suppliers to locate in the same area. A given firm chooses to locate in a given area because other firms are already located there. The result, as scholars of geography and urban studies have long known, is regional agglomeration. Krugman (1991) has developed the implications theoretically, as well as illustrated the manifestations with examples that are all around us. Examples include the famous tendency for the computer industry to cluster in Silicon Valley, California, or around Route 128, Massachusetts, the concentration of the auto industry in the midwest, the carpet industry in Georgia and so on. In many cases, the tendency toward concentration originally began with a single chance event.

The effect works internationally as well. A city-state such as Singapore owes its entire existence to the forces of agglomeration, together with the benefits, as the real estate agent mantra goes, of "location, location, location." The very high amount of trade among the Benelux countries and Germany largely reflects a concentration of heavy industry analogous to that in midwestern American states (Krugman, 1991c, 1995). The end result is that the question what goods a country will end up producing is often answered
by historical accident, more than by innate comparative advantage based on god-given endowments.\textsuperscript{46}

The second reason why we care about distance and location is much less cosmic in scope, but is the central point that this chapter contributes to the logic of the book. Distance between a pair of countries is an important natural determinant of the volume of trade between them. We need to isolate this factor before we can look at others. Thus, when looking at trade data, correcting bilateral data for distance is a helpful way of answering other questions. Examples include the questions whether rich countries trade more than poor ones, or whether a country like Japan is an outlier in global trade patterns. In particular, we wish to measure empirically the extent to which preferential trade policies are influencing bilateral trade patterns. To do so, we must hold constant for the distance between the two countries, and related factors such as whether or not they share a common border. In terms that were introduced in the preceding chapter, our plan is to apportion the observed regional concentration of trade between the Krugman-Summers explanation (proximity) and the Bhagwati-Panagariya explanation (existing regional trading arrangements).

The final reason why we care about distance and location bears

\textsuperscript{46} This is a lively little area for research currently. Krugman and Venables (1995) show how transport costs affect international agglomeration in manufacturing. [When transport costs are high, all countries have some manufacturing, but as transport costs fall below a critical value, a core-periphery pattern forms.] Brainard (1993) shows how transport costs affect the decisions of multi-national corporations to locate production overseas. Hanson (1994) applies the agglomeration concept to show the implications of liberalization and NAFTA for the regional location of the Mexican apparel industry.
directly on the central point of this book, which will be developed in later chapters [particularly Chapter 8]. We will see that countries that are located close together constitute a natural trading bloc, by which we mean that a reduction in trade barriers between them can be economically beneficial. We will also develop the notion of the optimal degree of regionalization of trade policy, which is the maximum amount of preferences that can be justified by geography. It will turn out that the more sharply do transportation costs rise per kilometer of distance, the higher is the optimal degree of regionalization, if other things are equal.

3.2 Three Kinds of Costs to Doing Business at a Distance

In a classic study of bilateral trade, Linnemann (1966) identified three different categories of costs associated with doing business at a distance. We will call them physical shipping costs, time-related costs, and costs of unfamiliarity. The first, shipping costs, is the most obvious.

Shipping costs

Although transportation costs are obvious, how to measure them is not. The easiest way to get a comprehensive measure of shipping costs for a country's trade is the ratio of c.i.f. trade values (measured at the cost to the importing country including insurance and freight) to f.a.s. or f.o.b. values (measured as it leaves the exporting country, i.e., "alongside ship" or "on board"). The difference, insurance and freight, ranges widely, depending on the
context. In general, the U.S. customs data show that transport costs for international trade exceed the cost of duties. We display the recent figures pertaining to U.S. imports from 63 countries in Table 3.1 [formerly 4.1]. The c.i.f. margin ranges from 1.7 per cent for close neighbor Mexico at the low end, to 12.9 per cent for distant neighbor Chile at the high end.

An econometric attempt to explain the c.i.f./f.o.b. ratio (in log form) by distance from the United States to the trading partner produced the following results. First, the log of distance fits slightly better than the level of distance. Second, in most years (1989-1994), the point estimate of the coefficient on the log of distance is slightly above 1.0. This suggests that aggregate transportation costs rise very roughly by about one percent of value when distance increases by one percent. Third, the coefficient is not estimated with any precision, when the c.i.f./f.o.b. is computed for the aggregate of trade across all commodities, as in Table 3.1 [formerly 4.1]. Fourth, the coefficient is more often statistically significant when the dependent variable is computed for a disaggregated commodity. For example, the cost of transport for

---

47 The c.i.f. difference was estimated by Linnemann at 5 per cent for intra-European trade on average and 10-15 per cent for intercontinental trade. These are the same two parameters that enter into the analysis of our Chapter 8.
49 The figures pertain to 1994. Chile is followed by Ecuador (and then four Middle Eastern countries). In much of this book we speak of South American countries as being in the same region as the United States. In fact, however, transportation costs between South America and North America are quite high, due in part to a highly regulated and cartelized shipping industry. We return to this point in Chapter 11. The recent study by Amjadi, Winters and Yeats (ib.) finds that the transport costs are higher from South America to the United States than from Europe, and suggests that therefore a FTAA is not a natural trade bloc.
mineral fuels, mineral oils and products of their distillation [SITC code 27] has a significant coefficient of 5.3 in 1994. Clearly trade among distant countries tends to be relatively concentrated in goods that are relatively easier to transport (relative to their value), as compared to trade among close countries. This suggests that the aggregate c.i.f./f.o.b. numbers can be seriously misleading.

Table 3.1 (formerly 4.1) about here

The margins in the table pertain to the set of 63 countries that we use as the data base for our most important estimates in this book. A more complete set of all countries shows transport costs for imports to the United States in 1994 running as high as 21.9 per cent for the Solomon Islands and 25.8 per cent for Guinea. Interestingly, some countries that are even more remote than these two apparently have low aggregate shipping costs, such as the Central African Republic, Zambia, and Rwanda. The explanation is certainly again the commodity composition of trade: costs are in fact so high for these landlocked African countries that only the most easily-shipped goods are exported. This shows the limitation of using the c.i.f./f.o.b. numbers to measure transportation costs without disaggregating by commodity. The worldwide average for the complete set of countries is 3.75 per cent [down slightly from 3.87 per cent in 1993 and 3.93 per cent in 1992].

There is an even wider variation in shipping costs across commodities than across countries. We have broken down U.S. imports
by 2-digit category. The c.i.f./f.o.b. estimate of transportation costs start from levels as low as 0.7 per cent for pearls (which have a high ratio of value to weight), 0.8 per cent for aircraft (which can fly themselves to their destination), and 0.9 per cent for works of art and antiques (again a high ratio of value to weight; surprisingly, costs arising from insurance or the special needs of fragility seem to be minor in proportion to the value). At the opposite extreme, costs go as high as 19.3 per cent for fruits and nuts (which are perishable), 20.9 for straw baskets and similar manufactures (low ratio of value to physical volume), and 25.1 per cent for salt, sulfur, earths, stone and plastering material (heavy).

The c.i.f. margins for the full range of commodities are reported in Appendix Table B3.1 [formerly 4.1].

[Figure 3.1 (formerly 4.1) about here]

Figure 3.1 (formerly 4.1) shows the aggregate worldwide c.i.f. margin from 1950 to 1993. It is substantially higher than for U.S. imports considered alone. It has been falling steadily over the last 45 years. This reflects declining transport costs. One suspects that the decline in shipping costs per kilometer of a given commodity is offset to a degree by the (resulting) tendency to undertake trade in commodities that would have previously been deemed too expensive, and with partners that would have previously been deemed too expensive. This suggests that shipping costs have fallen more steeply than the Figure indicates.
Boisso and Ferrantino (1995, Fig. 1) show that weighted-average distances between trading partner, where the weights are bilateral trade shares, have increased steadily in the postwar period. This illustrates two things. First, transportation costs and the other costs to doing business at a distance have continued to fall. This must be the reason why trade is taking place over longer and longer distances. Second, as one suspected, a measure of weighted-average transportation cost, such as the ratio of c.i.f. value to f.o.b. value, will not give a good indication of the extent to which transport costs are falling over time. To repeat, precisely because trade takes place over increasingly large distances, the trend in the trade-weighted average of transportation costs will be biased upwards as an estimate of the trend in transportation costs at a given distance.

**Time elapsed in transporting**

The second category of costs associated with doing business at a distance is the time element, which like shipping costs derives from the process of transporting the commodity. It includes interest charges, perishability, and the loss of adaptability to changing conditions that comes from long delivery lags. A firm will find it difficult to practice "just-in-time" inventory control if there is a long lag between the time a shipment is ordered and the time it arrives. An increasingly important concern in such industries as high-fashion apparel and sound and video recordings is that the product may have gone out of style by the time a shipment arrives.
Cultural unfamiliarity

Linnemann called the third of his categories "economic horizon," or "psychic distance." Here he had in mind that familiarity with another country's laws, institutions, habits and languages is an important part of marketing. Drysdale and Garnaut (1982) call this subjective resistance to trade, as opposed to the other, distance-related, factors, which they call objective resistance. We have all heard about such classic marketing mistakes as trying to sell Japanese consumers appliances that are too big for their kitchens or cars that have the steering wheels on the wrong side, or to sell Latin Americans a car with a name that in Spanish means "doesn't go" (nova). Cultural familiarity is highly correlated with geographical proximity, though history, language, and other factors matter greatly as well.

One may hypothesize, as a way of rendering psychic distance more precise, that the cost of human travel between points is an important component of the cost of doing business. In the case of some services, such as tourism or consulting, the point is almost tautological; the trip is virtually what is being bought or sold. Human travel is a factor for other goods as well, however, whether they are commodities where physical shipping costs are high, such as machinery, or zero, such as financial and insurance services. Consider a kind of export important in the United States: high-tech capital goods. To begin sales in a foreign country may involve many trips by various personnel: engineers, marketing people, higher-ranking executives to clinch a deal, and technical support staff to help install the equipment or
to service it when it malfunctions. Anecdotal evidence includes the example of a Korean company who would rather buy such equipment from a manufacturer in Japan than in the United States, because service personnel can respond to distress calls in a matter of hours, rather than days. One could conceivably try to distill the slightly nebulous psychic distance factor into hard numbers, by counting the dollar cost of such human travel that firms incur in association with operating in particular foreign countries. One could try to add in the cost of mail, courier services, and telephone calls, all of which depend to some degree on distance.

Any such attempts at direct quantification, however, would inevitably neglect many intangible factors. One example is the inconvenience of communicating in real time with different time zones. Another is the high cost (as a percent of value) to circulating newspapers over large distances. A third is radio and television transmission. Residents of East Germany stayed attuned to the culture of West Germany and the rest of the capitalist world, during the years when physical travel across the border was impossible, by listening to radio and television programs from West Germany. This would have been much harder if the distance between the two Germanies had been greater. East German awareness of specific consumer products undoubtedly helped to speed economic integration with West Germany when the Wall came down. More fundamentally, it was also a factor in the easterners' rejection of communism. The farther away is the country in question, the less well-informed are domestic residents likely to be about its culture, politics, economy, and so forth.
3.3 Indirect evidence on the importance of distance

Whether air travel is considered an important component of the costs of doing business internationally, or merely an example of it, the volume of air travel between pairs of cities offers a fascinating illustration of the role of distance, and a first look at the principles of the gravity model. Table 3.2 [formerly 4.2] shows international air traffic, measured by the number of passengers traveling a route in 1992, for the top 25 pairs of cities. If international airline travel were based solely on the size of the two cities in question, one would expect traffic among the ten largest cities in the world to account for most of the top 25 most-traveled routes. Yet, of the 25 most-traveled international air routes, only 7 involve pairs of the 10 largest cities, such as London-New York. Most of the list involves such proximate city pairs as Singapore-Kuala Lumpur, Dublin-London, and Taipei-Hong Kong. The lesson is that proximity is important, and in some sense is even as important as size, in determining economic transactions.

[Table 3.2, formerly 4.2, about here]

---

50 Ten cities implies 45=10x9/2 pairs of cities. Six of these 45 pairs turn out to lie within the same country; they should be excluded from the list because Table 3.2 shows only international city pairs. Thus, of these there are 39 among the top ten cities, enough to fill the top 25 slots and then some.

51 We have chosen to rank cities by economic size (population times the country's GDP per capita, at current exchange rates). If we had instead ranked cities by population, then only one pair would have entered the top 25 air routes (and that one is a proximate pair: Seoul-Tokyo). Distance would look even more important.
Perhaps the simplest proof that distance is still important is the observed tendency toward geographical agglomeration of industries, already noted above. The agglomeration occurs even in sectors where physical transport costs are negligible, as in financial services or computer software. Financial firms concentrate in Manhattan and computer firms concentrate in Silicon Valley. They do not choose to locate near each other because they are trading physical commodities with each other and wish to save on shipping costs. Rather, face-to-face contact is important for exchanging information and negotiating deals. Even the difference between a distance of one city block and two can be relevant sometimes. In corporate office buildings, government agencies, and universities, the proximity of one's office to the offices of key colleagues or superiors is considered a crucial aspect of one's job, and not solely for reasons of prestige.

The importance of distance is also revealed by analysis of data on prices of goods in different locations. If transport costs and other costs to doing business at a distance are important, then arbitrage should do a better job of keeping prices of similar goods in line when they are sold at locations close together rather than far apart. Engel and Rogers (1995) study consumer prices for 23 Canadian and U.S. cities. They look at 14 consumption categories over the period 1978 to 1994. They find that the distance between two North American cities has a significant effect on the variability of their relative prices (bi-monthly changes). The equation holds
constant for whether the two cities are on the same side of the border or not. The effect of a 1 per cent increase in bilateral distance on the standard deviation of relative price changes is an estimated .0000260 (annualized).\textsuperscript{52} The average distance between U.S.-Canadian city pairs is 1,346 miles. The upshot is that the estimated average cost that Canada-U.S. city pairs pay as a result of not being located close together is to increase the variability by about one fifth of the average standard deviation (annual).\textsuperscript{53} If the distance between the average pair of cities were to be doubled, then variability would further increase by about 3 per cent.\textsuperscript{54}

Engel and Rogers (1997) extend their results to a wider set of countries. They find that the log of distance has a statistically significant effect on relative price variability for seven out of nine industries tested. The impact of a one percent increase in distance is an increase in the annualized standard deviation of .0000273. This is nearly identical to the effect that Engel and Rogers found in their earlier study using U.S. and Canada data.\textsuperscript{55} The estimates hold constant for the effect of bilateral nominal exchange rate variability.

 Wei and Parsley (1995, Table 3) look at the variability in

\textsuperscript{52} This is the effect on bi-monthly variability (.0000106) times the square root of 6, assuming that the bi-monthly changes follow an approximate random walk. Table 3 [in the revised version].

\textsuperscript{53} The effect is an increase of .01825 in the annual standard deviation, that is, an increase in the monthly standard deviation of .00745 (=7.03* .00106. The average log distance is 7.03.)

\textsuperscript{54} In other words, if distance were increased by another 100 per cent (logarithmically, i.e., from 1,346 miles to 3,659), the monthly standard deviation would go up by another .00106. The annual standard deviation would go up by another .00260.

\textsuperscript{55} A one per cent increase in distance raises the monthly standard deviation by an estimated .00000789. This monthly number is multiplied by the square root of 12, to annualize it.
relative prices for given tradeable industries across fourteen industrial countries. Other factors that affect relative price variability, and thus must be held constant when evaluating the effect of distance, include bilateral exchange rate variability (when the cities are located in different countries), common membership in the EC or other trade blocs, and linguistic ties. Their estimate is that increasing the distance between two countries by 1 per cent raises the standard deviation of (yearly) price differentials by .000063. In other words, the distance effects that they estimate from their worldwide data set are about half again as large as the effects estimated for by Engel and Rogers.

In any case, the point is clear. The effectiveness of arbitrage diminishes as the distance between cities widens. Any model of bilateral trade must take distance into account.
CHAPTER 4 [formerly 5]: 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. 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,

---

56 Isard (1960), Steward (1948), and Zinf (1946).
in its most basic form, says that trade between country $i$ and country $j$ is proportional to the product of $\text{GNP}_i$ and $\text{GNP}_j$, 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 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."\textsuperscript{57}
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 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.

\textsuperscript{57} 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.
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. There is the problem, however, that even a data set of 1,953 country pairs has only so much information to give. If one tires it out by asking a lot of silly questions, the answers to the questions we really care about will become increasingly haphazard. 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.

---

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

59 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 entrepots), 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. But the results appeared unreliable when testing country effects and bloc effects at the same time. Thus our standard equation does not allow each country to have its own intercept.

The Gravity Equation

We now turn to the origins of the equation itself.

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

---

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

61 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.
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 [formerly 7].

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 (=63x62/2) 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. 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.

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).
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.\(^{63}\) But we shall focus here on population as the measure of size and self-sufficiency.

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

\[
\log T_{ij} = 0.7 \log (\text{GNP}_i \text{GNP}_j) + 0.3 \log [(\text{GNP/pop}_i)(\text{GNP/pop}_j)] - 0.7 \log (\text{Dist.}_{ij}).
\]

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:

\[
\log T_{ij} = 1.0 \log (\text{GNP}_i \text{GNP}_j) - 0.3 \log (\text{pop}_i \text{pop}_j) - 0.7 \log (\text{Dist.}_{ij}).
\]

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.

\(^{63}\) The estimate is highly significant statistically. Frankel and Romer (1995), Table 1.
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. 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 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:

64 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).
(3) \[ \log T_{ij} = 1.0 \log \left( \frac{\text{GNP/pop}_i}{\text{GNP/pop}_j} \right) + \cdot7 \log(\text{pop}_i) \cdot(\text{pop}_j) - .7 \log(\text{Dist}_{ij}) \].

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).²⁶

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 \textit{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 \textit{countries with similar levels of output/population will trade more than countries with dissimilar levels.} 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 \((GNP/pop_i)\) and \((GNP/pop_j)\) 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 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 \textit{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

\begin{footnotesize}
\footnotesize
\textsuperscript{66} 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.
\end{footnotesize}
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. 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.

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.)
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 of the Heckscher-Ohlin hypothesis. The results are described below, in Chapter 6 [formerly 7].

**Our Results for the Basic Gravity Variables**

\[
\log(T_{ij}) = \alpha + \beta_1 \log(GNP_i GNP_j) + \beta_2 \log(GNP_i / pop_j GNP_j / pop_j) \\
+ \beta_3 \log(DIST_{ij}) + \beta_4 (\text{ADJ}_{ij}) + \beta_5 (\text{LANG}_{ij}) + \gamma_1 (\text{WE}_{ij}) + \gamma_2 (\text{WH}_{ij}) + \gamma_3 (\text{EA}_{ij}) + u_{ij}.
\]

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

\[ (4) \]

The last five explanatory factors are dummy variables. \text{ADJ}_{ij}, short for Adjacency, is equal to 1 when countries \(i\) and \(j\) share a common border. \text{LANG}_{ij} is equal to one when they share a common language or past colonial links. \text{WE}, \text{WH}, and \text{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.

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. In most cases, a test would fail to reject the constraint that the sum of the coefficients on GNP and per capita GNP is 1.0. Holding constant for population, trade between two countries is simply proportionate to the product of their GNPs.

---

68 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.
The reported results measure GNPs and per capita GNPs at current exchange rates. An alternative is to measure them at PPP rates.\textsuperscript{69} 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.\textsuperscript{70} 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.

\textbf{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 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.\textsuperscript{71}

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

\textsuperscript{69} 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.

\textsuperscript{70} 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.

\textsuperscript{71} Boisso and Farrantino (1995) find very little difference in gravity equation results whether measuring distance between most populous cities or geographical centers.
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 [formerly 7] 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 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 [formerly 7.1].72 The data analyzed in this book, however, are all at the national level.

---

72 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. 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 Table 4.4.

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.

We noted when analyzing the c.i.f. measure of transport costs in chapter 3 that it does not seem to rise linearly with distance, but rather less rapidly than that. This is as one would expect: there is a large cost associated with loading cargo onto a vessel, and then a relatively small marginal cost per mile of distance traveled. The logarithm has the property that average cost diminishes with distance (assuming the coefficient comes out less than 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.

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

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.
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 [formerly 4], 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. 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 maquiladora 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.

---

75 Frankel (1993).
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.

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]\.

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.\(^76\)

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.\(^77\)

\(^76\) Reported in Frankel and Romer (1995).

\(^77\) 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.\(^7\) 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 own simpler distance measure. It had relatively little effect on most of the results.\(^7\) 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.\(^8\)

[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

\(^7\) 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).

\(^7\) Frankel, Wei, and Stein (1995). The coefficient of distance varies a bit over the course of the observations, but with no clear trend.

\(^8\) 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.]
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). 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). These findings confirm to us that physical transport costs are not necessarily the most important component of costs associated with distance. The costs associated with transport time and cultural unfamiliarity may be greater, and these costs are greater for manufactured goods than for agriculture.

---

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

82 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 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.\textsuperscript{83}

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 1880.\textsuperscript{84} 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.

\textsuperscript{83} 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).

\textsuperscript{84} 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.\footnote{I am most indebted to J. David Richardson for this point.}

**Linguistic links**
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.\textsuperscript{(86)} (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 **DROP5.4**)]. 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.\textsuperscript{(87)} Nonetheless, two languages, English and Chinese, seem to qualify. Two Chinese-speaking countries appear from the official statistics to trade four times as 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.\textsuperscript{(88)}

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

\textsuperscript{86} 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.

\textsuperscript{87} 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).]

\textsuperscript{88} Table 3 of F & W (1993a).
links are included than when they are not.\textsuperscript{89}

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.

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 [formerly 6.1 and 6.2] 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. [**DROP Table 9.1**])

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).

\textsuperscript{89} Estimates of the equation without the language effect were reported in F (1993).
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.\textsuperscript{90} Furthermore, our theoretical rationale for the gravity model (Chapter 7 \textit{formerly} \textsuperscript{8}) 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.\textsuperscript{91}

\textsuperscript{90} Bikker (1987, fn.3) points out that this is a problem with Aitken (1973). Brada and Mendez allow yearly constant terms.

\textsuperscript{91} A brief Appendix to Chapter 4 discusses some of the details of the interpretation of the intercept.
CHAPTER 5 [formerly 6]: 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 [formerly 6.1-6.3]). 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 trade disaggregated by major sector.

5.1 [formerly 6.1]: 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

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 statistically when estimated in this way.
The relevant grouping for western Europe has changed frequently. We start, in Table 4.2 [formerly 5.2], 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.  

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. 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.


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.] 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. Still the EC bloc effect remains usually significant in 1980 and thereafter.

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. The results show that both expansions of EC membership increased trade between the existing members and the newcomers. 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.

---

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.

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.]
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 [formerly 3.1] 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.

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 throughout, and turns statistically highly significant in 1980, when estimated on manufactures alone. 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.\textsuperscript{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.\textsuperscript{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 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.

\textbf{European Free Trade Area (EFTA)}

\textsuperscript{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.

\textsuperscript{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 \textit{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.
We have seldom found EFTA to be a statistically significant bloc.\footnote{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.\footnote{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 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.\footnote{18}

\textbf{Western Europe}


\footnote{17} F & W (1997).

\footnote{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. 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. 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 Twelve dominates the closedness of the EFTA Five. 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

---

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.


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.

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

---

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.

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.
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.

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 economies, 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

26 Collins and Rodrik (1991) and Brown, Deardorff, Djankov and Stern (1996) also forecast trade between these countries and the west.
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. The estimates are reported in Table 4.3 [formerly 5.3]. The coefficient is always negative. It even appears to be statistically significant in many years. In 1992, the coefficient is a significant -0.3, 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 [formerly 6.2]: 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 (=3x2/2) for NAFTA, 6 (=4x3/2) for MERCOSUR, and 10 (=5x4/2) 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. This constrains all PTAs 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 [formerly 5.2], 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, 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

\(^{28}\) [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).]
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 Claussing'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 [formerly 3.1]. 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.\textsuperscript{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.

\textsuperscript{31} In addition to Table4.2 [formerly 5.2] 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.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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.

\textsuperscript{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.

\textsuperscript{33} F, W & S (1995), Tables 4, 4a and 4b.

\textsuperscript{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 [formerly 5.2]. 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.\textsuperscript{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.\textsuperscript{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 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.


\textsuperscript{36} F, W & S (1995), Table 4a.
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 [formerly 5.3] 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

\(^{37}\) Surprisingly, the lack of openness appears in the Table to continue right up to 1992. The finding of a
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.\textsuperscript{38} When one allows for the low level of openness, the bloc coefficient registers .7 in 1991 {and .8 in 1992}.\textsuperscript{39}

When the data from 1970 to 1992 are pooled, in Table 5.1 \textbf{formerly} 6.1, 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.\textsuperscript{40}

\begin{quote}
\textsuperscript{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.)

\textsuperscript{39} Appendix Table A4.4a \textbf{formerly} A5.4a {and 5.2?, respectively}.

\textsuperscript{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).
\end{quote}
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.\footnote{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.}

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 [formerly 6.3]: 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

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.\textsuperscript{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.\textsuperscript{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.

\textsuperscript{42} F & W (1997, Table 1).

\textsuperscript{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 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

46 F & W (1997, Table 1).
47 Appendix Table A4.3 [formerly A5.3].
A question like "what is the effect of ASEAN on trade among its members" can change radically, depending on 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**

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 [formerly 5.3], 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.

---

49 W & F (1995), Table 1.
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 percent more with all parts of the world, other things equal, than do average countries \( \exp(0.20) = 1.22 \). 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 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 [formerly 6.1], 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


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

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

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.

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 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


\(^{55}\) 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).
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, transhipping 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. 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. 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 [formerly 5.3] 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.

When the years 1970, 1980, 1990, and 1992 are pooled, in Table 5.1 [formerly 6.1], 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. 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.

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, 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.

---

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).

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.\footnote{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.} 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.\footnote{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).} 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 high.\footnote{E.g., Polak (1996).} 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.
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 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

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.
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 [formerly 6.4]: 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 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


\(^{68}\) 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.
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**

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.

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.

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 [formerly 6.5]: Are These Regional Trading Arrangements Trade-Diverting?

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

72 Adding the Hong Kong and Singapore dummies does not change the qualitative features of the picture. F & W (1995d).
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, and that openness appears to increase over time. 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. 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 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.

---

73 Table A4.2 [formerly A5.2] and F and W (1996, 1997).

74 F & W (1995c) and F, W & S (1995). Not, however, in Table A5.2 [formerly A5.3].

75 W & F (1995). The effects are estimated at -0.1 (borderline-significant) and -0.3 (highly significant), respectively.
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.\textsuperscript{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 [formerly 5.2 and 5.3], 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.\textsuperscript{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.

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

\textsuperscript{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.

\textsuperscript{77} E.g., Harris and Cox (1984) and Primo Braga (1994).
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 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

---

78 Krueger (1993).
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.\textsuperscript{79} For NAFTA, we find a negative openness coefficient throughout, though it must be repeated that the
sample period precedes the date that the agreement took effect.\textsuperscript{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.\textsuperscript{81} The Andean Group shows a similar mix of estimated coefficients, looking
encouragingly open in some estimates, but less so in others.\textsuperscript{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.\textsuperscript{83} When examining the
Americas in the aggregate, however, the fairly consistent finding is a low level of openness.\textsuperscript{84}

\textit{Diversion by East Asia and Pacific Groupings}

\textsuperscript{79} F & W (1996).

\textsuperscript{80} Table 4.2 [formerly 5.2] here; W & F (1997); and F, W & S (1995).

\textsuperscript{81} Table 4.2 [formerly 5.2] here; W & F (1997); and F, W & S (1995).

\textsuperscript{82} F & W (1996), F, W & S (1995), and W & F (1997); versus Table 4.2 [formerly 5.2].

\textsuperscript{83} W & F (1995).

\textsuperscript{84} W & F (1997), F, W & S (1995), and Tables 4.3 and 5.1 [formerly 5.3 and 6.1].
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.\(^\text{85}\) If 1990 is taken as the key date for ASEAN, the estimated effect on the change in trade with non-members is also positive.\(^\text{86}\) The same openness is revealed for the broader grouping of East Asia.\(^\text{87}\)

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.\(^\text{88}\) 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.\(^\text{89}\) Others show coefficients that are more consistently negative.\(^\text{90}\) 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 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

\(^{85}\) Table 4.2 [formerly 5.2], F & W (1996), and W & F (1997).


\(^{87}\) Table 4.3 [formerly 5.3], 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 [formerly 5.2]; W & F (1997); and W & F (1995), respectively.

\(^{89}\) F & W (1995c).

\(^{90}\) Table 4.3 [formerly 5.3], F, W & S (1995), and W & F (1997).
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 looking 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 [formerly VII]. 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 [formerly 7.1]: 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 [formerly 3] 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. The issue becomes an important one for

---

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
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 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 Mideast\(^2\)), 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 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 addressed more formally in the next chapter.

\(^2\) Fischer (1993, pp. 434-436).
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 [formerly] 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. 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.

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. 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 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

---

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.
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. 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 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.

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

---

5 Layton and Rist (1925).

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.
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

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. 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. 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. 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 rations 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 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

---

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).
1994 would have been higher by a factor of about 2.3, if the union had continued.

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 [formerly table 6.1] 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.

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 [formerly 4] 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.\(^\text{10}\) 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 the distance to the nearest neighbor.

\(^{10}\text{To compare across time, look, for example, at the intercepts in Tables 6.1-6.3, discussed in the appendix to Chapter 4 [formerly 5]. To compare across countries, look at the intercepts in Appendix Table A4.5 [formerly A5.5], which estimate openness relative to the United States.}\)
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.\footnote{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.} 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 [formerly 7.2]: The Importance of History

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

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).

\textsuperscript{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.
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. 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

13 For example, Canada was in the Commonwealth but not sterling zone, and three Scandinavian countries vice versa; Eichengreen and Irwin (1995).
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 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 [formerly 7.3]: 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.\textsuperscript{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.

\textsuperscript{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. 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 [formerly 7.1]. 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. 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.

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

---

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.
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.\textsuperscript{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.

\textsuperscript{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.\textsuperscript{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.\textsuperscript{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 [formerly 7.4]: Rich and Poor Countries

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

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

\textsuperscript{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 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.\textsuperscript{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 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.\textsuperscript{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.\textsuperscript{23}

\textsuperscript{21} F, S and W (1995, Table 3).

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

\textsuperscript{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.
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.

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. Table 6.4 reports along these lines the results of instrumental variables estimation of the gravity equation. 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 [formerly 7.5]: The Role of Currency Links

In addition to the strategy of enacting regional trading arrangements, governments sometimes seek to

---

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.
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 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 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

26 E.g., Abreu and Bevilaqua (1995).
(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 (=0.28x3.22). 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

27 The estimates are taken from F & W (1996), Tables 4, 5 and 6. Similar results are reported in the earlier F & W papers.
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 [formerly 7.6]: 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

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.\textsuperscript{28} Thus it is in turn legitimate to add imports and exports together, to simplify the estimation and the presentation, as we have done.\textsuperscript{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

\textsuperscript{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 $T_p$, 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.]

\textsuperscript{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.
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 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. 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 [formerly 7.2] (Table 6.5a for 1965-75, and 6.5b for 1980-90; Table 6.3 [formerly 7.1] 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 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

---

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.
country's exports equals its imports.

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

{OR COULD DROP SINCE THIS SEPARATION IS NOW REPORTED
IN THE PRECEDING (TWO) TABLES}

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 anythini's currency 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 [formerly A7.2]. 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 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.

---

32 The results are reported in F & W (1997), Table 2. (Table A7.2 here is taken from Table 1 there.)
In Appendix Table 6.4 [formerly 7.2] 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.\textsuperscript{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.

6.7 [formerly 7.7]: 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?

\textsuperscript{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: openness = non-bloc + bloc.)
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 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.

---

\(^{34}\) 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.

\(^{35}\) 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.  

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

---

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.
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.  

6.8 [formerly 7.8]: 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 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 

---

38 Reported in F & W (1993b) and appendix, ibid., 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.
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**

<table>
<thead>
<tr>
<th>Study</th>
<th>Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brada and Mendez (1983)</td>
<td>-0.76</td>
</tr>
<tr>
<td>Bikker (1987)</td>
<td>-0.9 to -1.1</td>
</tr>
<tr>
<td>Boisso and Farrantino (1995)</td>
<td>-1.0 to -1.5</td>
</tr>
<tr>
<td>Linnemann (1966)</td>
<td>-0.77</td>
</tr>
<tr>
<td>Oguledo and MacPhee (1994)</td>
<td>-.76</td>
</tr>
<tr>
<td>Mansfield and Bronson (1995)</td>
<td>-.51 to -.69</td>
</tr>
</tbody>
</table>

**Controlling for adjacency**

<table>
<thead>
<tr>
<th>Study</th>
<th>Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aitken (1973)</td>
<td>-.35 to -.51</td>
</tr>
<tr>
<td>Biessen (1991)</td>
<td>-.74 to -.86</td>
</tr>
<tr>
<td>Bergstrand (1985)</td>
<td>-.75 to -.78</td>
</tr>
<tr>
<td>Tinbergen (1962)</td>
<td>-.56</td>
</tr>
</tbody>
</table>
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. 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, 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. 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.

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:
\[ \text{growth in nominal trade} = \text{estimated time trend} + 2(\text{nominal GNP growth}). \]
The true equation is:
\[ \text{growth in real trade} = \text{true time trend} + 2(\text{real growth}) - \text{real world growth}. \]
Thus \[ \text{true time trend} = \text{estimated time trend} + \text{real world growth} + \text{inflation}. \]

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 10^4 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.