Journal of Economic Literature
Vol. XLII (September 2004) pp. 691–751

Trade Costs

JAMES E. ANDERSON *and* Eric van Wincoop[1]

"*... the report of my death was an exaggeration.*" ~ Mark Twain, 1897

1. *Introduction*

The death of distance is exaggerated. Trade costs are large, even aside from trade-policy barriers and even between apparently highly integrated economies. Despite many difficulties in measuring and inferring the height of trade costs and their decomposition into economically useful components, the outlines of a coherent picture emerge from recent developments in data collection and especially in structural modeling of costs. Trade costs have economically sensible magnitudes and patterns across countries and regions and across goods, suggesting useful hypotheses for deeper understanding. This survey is a progress report. Much useful work remains to be done, so we make suggestions below.

Trade Costs Matter. (1) They are large—a 170-percent total trade barrier is constructed below as a representative rich country ad-valorem tax equivalent estimate. This includes all transport, border-related, and local distribution costs from foreign producer to final user in the domestic country. (2) Trade costs are richly linked to economic policy. Direct policy instruments (tariffs, the tariff equivalents of quotas and trade barriers associated with the exchange-rate system) are less important than other policies (transport infrastructure investment, law enforcement and related property-rights institutions, informational institutions, regulation, language). (3) Trade costs have large welfare implications. For example, Anderson and van Wincoop (2002) argue that current policy-related costs are often worth more than 10 percent of national income. (4) Maurice Obstfeld and Kenneth Rogoff (2000) argue that all the major puzzles of international macroeconomics hang on trade costs. (5) Details of trade costs matter to economic geography. For example, the home market effect hypothesis (big countries produce more of goods with scale economies) hangs on differentiated goods with scale economies having greater trade costs than homogeneous goods (Donald Davis 1998). (6) The cross-commodity structure of policy barriers is important to welfare (e.g., Anderson 1994).

Broadly Defined. Trade costs, broadly defined, include all costs incurred in getting a good to a final user other than the marginal cost of producing the good itself: transportation costs (both freight costs and time costs), policy barriers (tariffs and nontariff barriers), information costs, contract

[1] Anderson: Boston College and NBER. van Wincoop: University of Virginia and NBER. We are grateful to John McMillan, two referees, and Jeff Bergstrand for exceptionally generous and detailed comments and suggestions. We have also benefitted from many other comments by colleagues too numerous to single out, through e-mail and in presentations at conferences and at university workshops.

enforcement costs, costs associated with the use of different currencies, legal and regulatory costs, and local distribution costs (wholesale and retail). Trade costs are reported in terms of their ad-valorem tax equivalent. The 170-percent headline number breaks down into 55-percent local distribution costs and 74-percent international trade costs (1.7=1.55*1.74-1).

Both domestic and international trade costs are included because it is arbitrary to stop counting trade costs once goods cross a border. It is not even obvious when goods cross a border in the economic sense: Is it when they arrive on the dock, leave the dock, arrive at the importer?

Costs broadly defined and reported below may include some rent. Exclusion of rent requires a markup theory and its application.

Three Sources. We report on trade costs from three broad sources. Direct measures of trade costs are discussed in section 2. Direct measures are remarkably sparse and inaccurate. Two types of indirect measures complement the incomplete and inadequate direct measures: inference from quantities (trade volumes), discussed in section 3, and inference from prices, discussed in section 4.

Theory Looms Large. In our survey, theory looms large. A theoretical approach is inevitable to infer the large portion of trade costs that cannot be directly measured in the data. The literature on inference about trade barriers from final goods prices remains largely devoid of theory. We point to ways in which trade theory can be effectively used to fill this gap and learn more about trade barriers from evidence on prices. Recent developments have bridged the gap between practice and theory in the inference of trade costs from trade flows. Readers who pause with us on the bridge will produce better work in the future.

The gravity model provides the main link between trade barriers and trade flows. Gravity is often taken to be rather atheoretic or justified only under highly restrictive assumptions. We place gravity in a wide class of trade separable general equilibrium models. Trade separability obtains when the allocation of trade across countries is separable from the allocation of production and spending within countries. Gravity links the cross-country general equilibrium trade allocation to the cross-country trade barriers, all conditional on the observed consumption and production allocations. Inferences about trade costs therefore do not depend on the general equilibrium structure that lies beneath the observed consumption and production allocations within countries.

Appropriate aggregation of trade costs is a key concern. Aggregation of some sort is inevitable due both to the coarseness of observations of complex underlying phenomena and the desirability of simple measures of very high dimensional information. We show how theory can be used to replace common atheoretic aggregation methods with ideal aggregation. We also show how theory can shed light on aggregation bias and what can be done to resolve it.

Trade Costs Are Large and Variable. Mattel's Barbie doll, discussed in Robert Feenstra (1998), illustrates large costs. The production costs for the doll are $1, while it sells for about $10 in the United States. The cost of transportation, marketing, wholesaling, and retailing have an ad-valorem tax equivalent of 900 percent.

A rough estimate of the tax equivalent of "representative" trade costs for industrialized countries is 170 percent. This number breaks down as follows: 21 percent transportation costs, 44 percent border-related trade barriers, and 55 percent retail and wholesale distribution costs (2.7=1.21*1.44*1.55). The 21-percent transport cost includes both directly measured freight costs and a 9-percent tax equivalent of the time value of goods in transit. Both are based on estimates for U.S. data. The 44-percent border-related barrier is a combination of direct observation and inferred costs. Total international trade costs are then about 74 percent (0.74=1.21*1.44-1). Representative retail and wholesale distribution costs are set at 55 percent, close to the average for industrialized countries.

Direct evidence on border costs shows that tariff barriers are now low in most countries, on average (trade-weighted or arithmetic) less than 5 percent for rich countries, and with a few exceptions are on average between 10 percent and 20 percent for developing countries. Our overall representative estimate of policy barriers for industrialized countries (including nontariff barriers) is about 8 percent. Inferred border costs appear on average to dwarf the effect of tariff and nontariff policy barriers. An extremely rough breakdown of the 44-percent number reported above is as follows: an 8-percent policy barrier, a 7-percent language barrier, a 14-percent currency barrier (from the use of different currencies), a 6-percent information cost barrier, and a 3-percent security barrier for rich countries.

Trade costs are also highly variable across both goods and countries. Trade barriers in developing countries are higher than those reported above for industrialized countries. High value-to-weight goods are less penalized by transport costs. The value of timeliness varies across goods, explaining modal choice. Poor institutions and poor infrastructure penalize trade differentially across countries. Sectoral trade barriers appear to vary inversely with elasticities of demand. Policy barriers, especially nontariff barriers (NTBs), also vary significantly across goods. NTBs are highly concentrated: in many sectors they are close to zero, but U.S. textiles and apparel have 71 percent of products covered, with tariff equivalents ranging from 5 percent to 33 percent.

2. *Direct Evidence*

Direct evidence on trade costs comes in two major categories, costs imposed by policy (tariffs, quotas and the like) and costs imposed by the environment (transportation, insurance against various hazards, time costs). We review evidence on international policy barriers, transport costs and wholesale and retail distribution costs. We focus on current and recent trade costs in this survey; see the work of Williamson and co-authors (e.g., Kevin O'Rourke and Jeffrey Williamson 1999) for historical evidence.

An important theme is the many difficulties faced in obtaining accurate measures of trade costs. Particularly egregious is the paucity of good data on policy barriers.[2] Transport-cost data is limited in part by its private nature. Many other trade costs, such as those associated with information barriers and contract enforcement, cannot be directly measured at all. Better data on trade costs are feasible with institutional resources and would yield a high payoff.

2.1 *Policy Barriers*

2.1.1 *Measurement Problems and Limitations*

How high are policy barriers to trade? This seemingly simple question cannot usually be answered with accuracy for most goods in most countries at most dates. The inaccuracy arises from three sources: absence of data, data that are useful only in combination with other missing or fragmentary data, and aggregation bias. Each of the difficulties is discussed further below.

The key open[3] (to the research community) source for panel data on policy barriers

[2] The grossly incomplete and inaccurate information on policy barriers available to researchers is a scandal and a puzzle. It is natural to assume that trade policy is well-documented since theory and politics have emphasized it for hundreds of years.

[3] UNCTAD sells the TRAINS data each year to commercial customers who use it to provide current information on trade costs to potential traders. It comes with a front end designed for convenience in pulling a maximum of 200 lines of data while preventing a user from gaining access to the whole database and using it to compete with UNCTAD's potential sales to other customers.

to trade is the United Nations Conference on Trade and Development's Trade Analysis & Information System, TRAINS. It contains information on trade control measures (tariff, para-tariff[4] and nontariff measures) at tariff line level for a maximum of 137 countries beginning in the late 1980s. TRAINS reports all data on bilateral tariffs, nontariff barriers, and bilateral trade flows at the six-digit level of the Harmonized System (HS) product classification for roughly 5000 "products." Countries use a finer product classification when reporting tariffs to UNCTAD. Thus, multiple tariff "lines" underlie each six-digit aggregate. It is in some cases possible to drill down to the national tariff line level. Some organizations have obtained the full TRAINS database (without the front end software) and provide access to a limited set of users inside the organization.[5]

Table 1 gives a sense of the substantial incompleteness of TRAINS. For each year from 1989 to 2000, the table reports the fraction of countries that report some lines (though possibly a very limited number) for tariffs, NTBs, and trade flows. Of 121 reporting countries in 1999, 43 percent report tariffs, 30 percent report NTBs, 55 percent report trade flows, and 17.4 percent have data for all three. The other countries report no data at all for any good. Coverage is not much better in other years. Coverage is better for OECD countries—over 50 percent have tariff and NTB information recorded in 1999, with considerable variability in coverage across the years.

The World Bank has very recently put together from these elements the most comprehensive system for researchers, and in principle provides public access.[6] Its World Integrated Trade System (WITS) software is coupled to TRAINS and to the World Trade Organization (WTO) Integrated Data Base and Consolidated Tariff Schedules[7] along with the UN Statistical Division's COMTRADE trade flow data. In principle it allows users to drill down and select data according to their own criteria, to track the complexities of trade policy as finely as the primary inputs allow.[8] WITS has some other data handling and modeling functions as well.

National sources in combination with the above allow better measurement of a single country's trade policy. A series from the Institute for International Economics provides measures of a few national trade policies

[4] Para-tariff measures include customs surcharges such as import license fees, foreign exchange taxes, stamps, etc.

[5] For example, Boston College has purchased disaggregated tariff information from UNCTAD's TRAINS database for the years 1988 through 2001, inclusive. We have data for 137 countries for at least one year, counting the European Union as a single country, but far less than the maximum amount suggested by fourteen years times 137 countries.

[6] See http://wits.worldbank.org. At this writing, there are still technical glitches facing a user trying to gain access. WITS only runs in late model Windows machines, users may need some IT support to install the software; and a user must to pay fees to UNCTAD for use of COMTRADE and TRAINS. Email queries are not answered, in our experience, without using friends at the Bank as intermediaries.

[7] The WTO Consolidated Tariff Schedule database lists the Most Favored Nation (MFN) bound tariffs at the tariff line level. The bound tariffs are the upper limits under the member countries' WTO obligation for actual tariffs charged to member countries not associated with the importer in a free trade agreement or a customs union, and often exceed the actual duties charged. The WTO's Integrated Data Base contains information on the applied rates at the national tariff line level. The data is closed to researchers. The WTO periodically reports on individual member-country trade policy with a published trade policy review based on its data.

[8] The World Bank Trade and Production database produces from its resources a set of three- and four-digit aggregates of trade, production, and tariff data. It is published on the Bank website and presumably will be regularly updated. The Trade and Production database covers 67 developing and developed countries over the period 1976–99. Again, the description given misleadingly suggests a useful panel; the actual data is full of missing observations due to the underlying limitations of TRAINS. The sector disaggregation in the database follows the International Standard Industrial Classification (ISIC) and is provided at the three-digit level (28 industries) for 67 countries and at the four-digit level (81 industries) for 24 of these countries.

TABLE 1
PERCENTAGE OF COUNTRIES WITH DATA IN TRAINS

	All Countries				OECD Countries			
Year	Tariff	NTB	Trade	All	Tariff	NTB	Trade	All
1989	0.8	1.7	1.7	0.8	0.0	0.0	5.6	0.0
1990	1.7	0.8	2.5	0.0	0.0	0.0	5.6	0.0
1991	6.6	5.8	11.6	0.8	11.1	16.7	27.8	0.0
1992	10.7	10.7	33.9	5.0	16.7	16.7	61.1	5.6
1993	30.6	23.1	35.5	13.2	83.3	72.2	72.2	50.0
1994	10.7	20.7	57.9	5.0	0.0	27.8	83.3	0.0
1995	33.1	17.4	63.6	14.0	50.0	22.2	83.3	22.2
1996	18.2	13.2	62.8	6.6	61.1	55.6	83.3	44.4
1997	29.8	15.7	58.7	8.3	27.8	11.1	83.3	11.1
1998	36.4	17.4	47.1	9.1	50.0	5.6	83.3	5.6
1999	43.0	29.8	55.4	17.4	55.6	50.0	83.3	38.9
2000	36.4	2.5	0.0	0.0	55.6	55.6	0.0	0.0

Notes: The data are from UNCTAD's TRAINS database (Haveman Repackaging). The table reports the percentage of all countries, based on total of 121 reporting countries, that have at least one type of data for one year available through TRAINS. For OECD countries the percentages are based on 19 countries. "All" indicates that a country has reported all three types of data for that year.

for single years.[9] See also the WTO's Trade Policy Reviews series.

Data limitations can make even the available trade-barrier information difficult to interpret. TRAINS does not routinely report ad-valorem equivalents of specific tariffs (on quantities rather than values). Where specific tariffs are prevalent, such as in agriculture in many countries, the omission of specific tariffs is highly misleading. Many tariffs are in the form of specific taxes (on quantity), and must be converted to ad valorem equivalents using price information that must be matched up to the tariffs. Imperfections of classification or other information introduce potential measurement error. TRAINS reports the percentage of underlying lines that have specific tariffs in order to provide some information about how widespread the measurement problem may be. The World Bank's trade-barrier database substantially supplements TRAINS by supplying missing specific tariff information. The match of the tariff line classifications with the commodity classifications for trade flows is imperfect as well, introducing measurement error when converting all data to the Harmonized System.

[9] See Gary Hufbauer and Kimberly Ann Elliott (1994) on the United States; Patrick Messerlin (2001) on the European Union; Yoko Sazanami, Shujiro Urate, and Hiroki Kawai (1995) on Japan; Namdoo Kim (1996) on South Korea; and Shuguang Zhang, Zhang Yansheng, and Wan Zhongxin (1998) on China.

Nontariff barriers (NTBs) are much more problematic than tariff barriers. The World Bank database unfortunately does not provide NTB data. A user must use TRAINS or more specialized databases directed at particular NTBs. The TRAINS database records the presence or absence of a nontariff barrier (NTB) on each six-digit line. Many differing types of nontariff barriers are recorded in TRAINS (a total of eighteen types). The NTB data requires concordance between the differing NTB, tariffs, and trade classification systems at the national level, converting to the common HS system.

Jon Haveman's extensive work with TRAINS has produced a usable NTB database.[10] Haveman follows what has become a customary grouping of NTBs into hard barriers (price and quantity measures), threat measures (antidumping and countervailing duty investigations and measures), and quality measures (standards, licensing requirements, etc.). A fourth category is embargoes and prohibitions. A common use of the NTB data is to construct a measure of the prevalence of nontariff barriers, such as the percentage of HS lines in a given aggregate that are covered by NTBs. Nontariff barrier information in TRAINS is particularly prone to incompleteness and poor-quality problems, so analysts seeking to study particular sectors such as the Multi-Fibre Arrangement (MFA) will do better to access specialized databases such as the World Bank's MFA data.[11]

No information about NTB restrictiveness is provided in TRAINS, since measuring the restrictiveness of each type of nontariff barrier requires an economic model. In some important cases, individual analysts have developed direct measures of the restrictiveness of NTBs based on quota license prices where these are available.

Indirect methods of measuring the restrictiveness of NTBs are important because of the paucity of direct measures. One method is to infer the restrictiveness of nontariff barriers through the comparison of prices. Some important trade lines are well-suited for price comparisons (homogeneous products sold on well-organized exchanges, for example), but even here there are important issues with domestic transport and intermediary margins and the location of wholesale markets relative to import points of entry. Evidence from price comparisons is discussed in section 4. The restrictiveness of nontariff barriers can also be inferred from trade quantities in the context of a well-specified model of trade flows. Inference about nontariff barriers from trade flows is discussed in section 3. Alan Deardorff and Robert Stern (1998) and Sam Laird and Alexander Yeats (1990) provide other detailed discussions of inference about the restrictiveness of NTBs.

Aggregation is an important problem in the use and analysis of trade barriers. Tariffs and NTBs comprise some 10,000 lines, with large variation across the lines. The national customs authorities are the primary sources of trade restrictions, and their classification systems do not match up internationally or even intranationally, as between trade flows on the one hand and tariff and nontariff-barrier classes on the other hand. Matching up the tariff, nontariff, and trade-flow data requires aggregation, guided by concordances that are imperfect and necessarily generate measurement error. Moreover, for many purposes of analysis, the comprehension of the analyst is overwhelmed by detail, and further aggregation is desirable. Atheoretic indices such as arithmetic (equally weighted) and trade-weighted average tariffs are commonly used, while production-weighted averages sometimes replace them. As for nontariff barriers, the binary indicator is aggregated into a nontariff barrier coverage ratio, the arithmetic or

[10] See the Ultimate Trade Barrier Catalog at http://www.eiit.org/Protection.

[11] National tariff line information is also very problematic when analyzing nontariff barriers. For example, matching up reported trade flows with annual quotas immediately runs into inconsistencies in reporting conventions.

trade-weighted percentage of component sectors with nontariff barriers.

Ideal aggregation is proposed by James Anderson and Peter Neary (1996, 2003) based on the idea of a uniform tariff equivalent of differentiated tariffs and NTBs. Theoretically consistent aggregation depends on the purpose of the analysis, so the analyst must specify tariff equivalence with reference to an objective that makes sense for the task at hand. Anderson and Neary develop and apply indices for the small-country case that are equivalent in terms of welfare and in terms of distorted aggregate trade volume,[12] and show that atheoretic aggregation can significantly bias the measurement of trade restrictiveness.[13] The theme of appropriate aggregation in the different setting of many countries in general equilibrium plays a prominent role in our discussion of indirect measurement of trade costs, so we defer a full treatment to that section. Ideal aggregation is informationally demanding, so for that reason and because of their familiarity and availability in the work of others, we report the standard trade-weighted and arithmetic averages of tariffs and of NTB coverage ratios below.

2.1.2 *Evidence on Policy Barriers*

Tariffs. Trade-weighted and arithmetic average tariffs are reported in table 2 for fifty countries for 1999, based on TRAINS data.[14] The relatively small number of countries reflects reporting difficulties typical of TRAINS; some earlier years contain data for more countries. The reported numbers aggregate the thousands of individual tariff lines in the underlying data. The table confirms that tariffs are low among most developed countries (under 5 percent), while developing countries continue to have higher tariff barriers (mostly over 10 percent). Dispersion across countries is wide: Hong Kong and Switzerland have 0 percent tariffs, the United States has a 1.9 percent simple average, and at the high end India has 30.1 percent and Bangladesh 22.7 percent.

The variation of tariffs across goods is quite large in all countries; typically only a few are large. Intuition suggests that the variation of tariffs adds to the welfare cost. Marginal deadweight loss is proportional to the tariff, hence the cumulated dead weight loss triangle varies with the square of the tariff. Coefficients of variation (the ratio of the standard deviation to the mean) of tariffs, either arithmetic or trade-weighted, thus sometimes supplement averages. Anderson and Neary (2003) report trade-weighted coefficients of variation of tariffs for 25 countries around the year 1990, ranging from 0.14 to 1.67, many being clustered around one. They show that a proper analysis qualifies the simple intuition considerably, with welfare cost increasing in an appropriately weighted coefficient of variation.

Bilateral variation of tariffs can also be large. Preferential trade is mainly responsible since insiders face a zero tariff while outsiders face the MFN tariff, but aggregation over goods induces further bilateral variation due to differing composition of trade across partners. James Harrigan (1993) reports bilateral production-weighted average tariffs in 28 product categories for OECD countries for

[12] See also Anderson (1998) for equivalence in terms of sector-specific factor income and its relationship to the commonly reported effective rate of protection.

[13] Anderson and Neary (2003) report results using a volume equivalent uniform tariff— replacement of the differentiated tariff structure with a uniform tariff such that the general equilibrium aggregate value of trade in distorted products (in terms of external prices) is held constant. The ideal index is usually larger than the trade-weighted average—an arithmetic average across countries in the study yields approximately 11 percent for the trade weighted average and 12 percent for the uniform volume equivalent, while the U.S. numbers in 1990 are 4 percent and 4.8 percent respectively. For purposes of comparison between the initial tariffs and free trade, the two indexes are quite highly correlated. For a smaller set of evaluations of year-on-year changes, in contrast, Anderson and Neary show that the ideal and trade-weighted average indexes are uncorrelated.

[14] Calculations were based on TRAINS annual databases purchased by Boston College without the frontend software and assembled into panel data.

TABLE 2
SIMPLE AND TRADE-WEIGHTED TARIFF AVERAGES—1999

Country	Simple Average	TW Average
Argentina	14.8	11.3
Australia	4.5	4.1
Bahamas	0.7	0.8
Bahrain	7.8	–
Bangladesh	22.7	21.8
Barbados	19.2	20.3
Belize	19.7	14.9
Bhutan	15.3	–
Bolivia	9.7	9.1
Brazil	15.5	12.3
Canada	4.5	1.3
Chile	10.0	10.0
Colombia	12.2	10.7
Costa Rica	6.5	4.0
Czech Republic	5.5	–
Dominica	18.5	15.8
Ecuador	13.8	11.1
European Union	3.4	2.7
Georgia	10.6	–
Grenada	18.9	15.7
Guyana	20.7	–
Honduras	7.5	7.8
Hong Kong	0.0	0.0
India	30.1	–
Indonesia	11.2	–
Jamaica	18.8	16.7
Japan	2.4	2.9
Korea	9.1	5.9
Mexico	17.5	6.6
Montserrat	18.0	–
New Zealand	2.4	3.0
Nicaragua	10.5	11.0
Paraguay	13.0	6.1
Peru	13.4	12.6
Philippines	9.7	–
Romania	15.9	8.3
Saudi Arabia	12.2	–
Singapore	0.0	0.0
Slovenia	9.8	11.4
South Africa	6.0	4.4
St. Kitts	18.7	–
St. Lucia	18.7	–
St. Vincent	18.3	–
Suriname	18.7	–
Switzerland	0.0	0.0
Taiwan	10.1	6.7
Trinidad	19.1	17.0
Uruguay	4.9	4.5
USA	2.9	1.9
Venezuela	12.4	13.0

Notes: The data are from UNCTAD's TRAINS database (Haveman repackaging).
A "–" indicates that trade data for 1999 are unavailable in TRAINS.

1983. For Canada, the reported range of bilateral averages runs from 1.2 percent to 3.2 percent and for Japan from 2.3 percent to 4.5 percent. The United States has more modest differences, from 1.6 percent to 2.3 percent.[15]

Nontariff Barriers. Table 3 reports the prevalence of nontariff barriers for 34 countries in 1999 based on data from TRAINS and Haveman's work (rendering it more usable). The smaller number of countries available than for tariffs reflects previously noted limitations of TRAINS. We report arithmetic and trade-weighted NTB coverage ratios; the percentage of tariff lines subject to NTBs. The trade-weighted NTB coverage ratios generally exceed the arithmetic average NTB coverage ratios, often considerably so. For example, on the narrow definition (defined below), the U.S. arithmetic NTB coverage is 1.5 percent, while the trade-weighted NTB coverage is 5.5 percent. This reflects the fact that NTBs tend to fall on important traded goods such as textiles and apparel.

Narrow NTB coverage is basically price and quantity control measures and quality control measures, while broad coverage is the narrow classification plus threat measures related to antidumping. Threats are significant, since Robert Staiger and Frank Wolak (1994) have shown that the threat of antidumping impedes trade considerably (on average for the United States they find a seventeen-percent reduction in trade due to an ongoing investigation). Narrow NTBs cover less than 10 percent of trade for the rich countries and modest amounts of trade in most except for Argentina and Brazil. Table 3 implies that antidumping is quite common. For example, based on the TRAINS data, the United States in 1999 had 1.5 percent of tariff lines subject to NTB when narrowly defined but 27.2 percent of tariff lines subject to NTB when broadly defined. These measures do not include withdrawn suits, which are presumptively restrictive since they facilitate collusion. Table 3 also shows wide variability over countries in the use of NTBs. The broad definition with arithmetic averages has South Africa at 11.3 percent and Argentina at 71.8 percent.

The use of NTBs is concentrated in a few sectors in most economies. Table 4 reports sectoral NTB coverage ratios for the United States, European Union, Japan, and Canada for 1999. NTBs are widely used by developed countries in food products (for example, the trade-weighted NTB coverage of agriculture, forestry and fishery products is 74 percent for the United States and 24 percent for the European Union), textiles/apparel (71 percent for the United States and 42 percent for the European Union), wood and wood products (39 percent for the United States and 26 percent for the European Union) and in some other areas of manufacturing. The products involved are quite significant in the trade of developing countries but also somewhat significant in the trade of developed countries with each other.

In comparison to tariffs, NTBs are concentrated in a smaller number of sectors and in those sectors they are much more restrictive. Deardorff and Stern (1998) survey most of what limited data is available on quota license prices. Table 5 (based on table 3.6 of Deardorff and Stern 1998) gives estimates of tax equivalents based on annual averages of weekly Hong Kong MFA quota license prices[16] (which are themselves averages of

[15] Harrigan's data is available to the public. Huiwen Lai and Daniel Trefler (2002) are notable for compiling three-digit bilateral tariffs for fourteen importers and 36 exporters for 1972, 1977, 1982, 1987, and 1992. They emphasize the hard work involved and have not yet made the data public.

[16] Where markets subject to quotas are thick and well-organized and behavior of all agents is competitive, quota license prices provide the best evidence of tariff equivalents. Using license price data even under these assumptions forces the analyst to face the many dimensions on which the quota is not equivalent to a tariff—daily price quotes exhibit within-year variation with economically significant patterns (seasonality, year-end jumps and drops)—such that no single index of them can generally be equivalent to a tariff. Nevertheless, average license prices in combination with the substantial rent-retaining tariffs found on most quota-constrained products provide a useful measure of the restrictiveness of quotas.

TABLE 3
NON-TARIFF BARRIERS—1999

Country	NTB ratio (narrow)	TW NTB ratio (narrow)	NTB ratio (broad)	TW NTB ratio (broad)
Algeria	.001	.000	.183	.388
Argentina	.260	.441	.718	.756
Australia	.014	.006	.225	.351
Bahrain	.009	–	.045	–
Bhutan	.041	–	.045	–
Bolivia	.014	.049	.179	.206
Brazil	.108	.299	.440	.603
Canada	.151	.039	.307	.198
Chile	.029	.098	.331	.375
Colombia	.049	.144	.544	.627
Czech Republic	.001	–	.117	–
Ecuador	.065	.201	.278	.476
European Union	.008	.041	.095	.106
Guatemala	.000	.000	.348	.393
Hungary	.013	.034	.231	.161
Indonesia	.001	–	.118	–
Lebanon	.000	–	.000	–
Lithuania	.000	.000	.191	.196
Mexico	.002	.000	.580	.533
Morocco	.001	–	.066	–
New Zealand	.000	.004	.391	.479
Oman	.006	.035	.134	.162
Paraguay	.018	.108	.256	.385
Peru	.021	.094	.377	.424
Poland	.001	.050	.133	.235
Romania	.001	.000	.207	.185
Saudi Arabia	.014	–	.156	–
Slovenia	.030	.019	.393	.408
South Africa	.000	.002	.113	.161
Taiwan	.057	.074	.138	.207
Tunisia	.000	.000	.317	.598
Uruguay	.052	.098	.354	.470
USA	.015	.055	.272	.389
Venezuela	.131	.196	.382	.333

Notes: The data are from UNCTAD's TRAINS database (Haveman repackaging). The "narrow" category includes, quantity, price, quality and advance payment NTBs, but does not include threat measures such as antidumping investigations and duties. The "broad" category includes quantity, price, quality, advance payment and threat measures. The ratios are calculated based on six-digit HS categories.
A "–" indicates that trade data for 1999 are not available.

TABLE 4
NTB COVERAGE RATIOS BY SECTOR—1999

		United States 1999				EU-12 1999				Japan 1996				Canada 1999			
		Narrow NTB-ratio		Broad NTB-ratio		Narrow NTB-ratio		Broad NTB-ratio		Narrow NTB-ratio		Broad NTB-ratio		Narrow NTB-ratio		Broad NTB-ratio	
		S	TW	S	TW	S	TW	S	TW	S	TW	S	TW	S	TW	S	TW
ISIC	Description																
1	Agric., Forestry, Fish.	.011	.052	.719	.743	.001	.001	.229	.241	.153	.227	.897	.962	.028	.022	.878	.938
2	Mining, Quarrying	.000	.000	.018	.099	.001	.055	.001	.055	.028	.008	.193	.706	.000	.000	.027	.014
21	Coal Mining	.000	.000	.000	.000	.004	.000	.004	.000	.000	.000	.667	1.000	.000	.000	.000	.000
22	Crude Petroleum	.000	.000	.250	.105	.004	.067	.004	.067	.000	.000	1.000	1.000	.000	.000	.375	.019
23	Metal Ore Mining	.000	.000	.000	.000	.000	.000	.000	.000	.087	.000	.087	.000	.000	.000	.000	.000
29	Other Mining	.000	.000	.000	.000	.001	.038	.001	.038	.014	.129	.120	.184	.000	.000	.000	.000
3	Manufacturing	.015	.047	.245	.423	.007	.042	.083	.107	.044	.102	.322	.366	.171	.044	.261	.196
31	Food, Bev., Tobacco	.072	.120	.644	.809	.004	.011	.489	.474	.185	.329	.925	.893	.185	.348	.456	.453
32	Textiles, Apparel	.000	.002	.509	.708	.030	.255	.102	.420	.022	.050	.163	.120	.762	.681	.816	.784
33	Wood, Wood Prod.	.000	.000	.459	.389	.000	.007	.197	.263	.000	.000	.098	.025	.016	.015	.262	.252
34	Paper, Paper Prod.	.000	.000	.053	.023	.000	.000	.000	.000	.000	.000	.133	.036	.000	.000	.000	.000
35	Chem., Petrol Prod.	.036	.149	.114	.322	.003	.011	.032	.033	.048	.169	.635	.750	.013	.007	.047	.073
36	Non-Metal Min. Prod.	.000	.006	.014	.029	.000	.016	.000	.043	.000	.000	.073	.160	.000	.000	.000	.000
37	Basic Metal Ind.	.003	.044	.006	.044	.002	.010	.012	.016	.051	.086	.375	.139	.000	.000	.381	.362
38	Fab. Metal Prod.	.002	.039	.166	.450	.000	.010	.005	.012	.032	.057	.095	.266	.000	.000	.048	.179
39	Other Manuf.	.000	.002	.122	.199	.000	.017	.238	.222	.000	.000	.134	.112	.000	.000	.073	.012
Total All Products		.015	.055	.272	.389	.008	.041	.095	.106	.055	.098	.369	.442	.151	.039	.307	.198

Notes: "S" indicates "simple" and "TW" indicates "trade-weighted." Data are from UNCTAD's TRAINS database (Haveman repackaging). The "Narrow" category includes, quantity, price, quality and advance payment NTBs, but does not include threat measures such as antidumping investigations and duties. The "Broad" category includes quantity, price, quality, advance payment and threat measures. The ratios are calculated for two-digit ISIC categories based on the six-digit HS classifications used by TRAINS, using HS to ISIC concordances published by the World Bank.

transactions within the week) for textiles and apparel subject to quota between controlled exporters and the United States in 1991 and 1993. The license prices imputed for other suppliers depend on arbitrage assumptions and especially on relative labor productivity assumptions which may not be met. The prices are expressed as ad-valorem tax equivalents using Hong Kong export prices for the underlying textile and apparel items, trade-weighted across suppliers.[17] We add the U.S. tariffs on the corresponding items to the license prices to form the full tax equivalent. The table shows fairly high tax equivalents, especially in the largest trade categories (23 percent for products of broad-woven fabric mills, 33 percent for apparel made from purchased material). There is also high variability of license prices and tax equivalents across commodities (from 5 percent to 23 percent for textiles, from 5 percent to 33 percent for apparel). Earlier years reveal higher tax equivalents.[18] Thus there is

[17] Because the license prices are for transfer between holders and users and are effectively subject to penalty for the holder, the implied tariff equivalents are lower bounds to true measures of restrictiveness; this bias direction probably also applies to the intertemporal averaging.

[18] Anderson and Neary (1994) report trade-weighted average tax equivalents across the MFA commodity groups for a set of exporters to the United States in the mid-1980's. The Hong Kong average exceeded 19 percent in each year and ranged to over 30 percent in some years, while tax equivalents (very likely biased upward) for other countries were much larger, some over 100 percent.

TABLE 5
TARIFF EQUIVALENTS OF U.S. MFA QUOTAS, 1991 AND 1993 (PERCENT)

Sector	1991	1993				
	Rent Tar Eq.	Rent Tar Eq.	S Tariff	TW Tariff	Rent + TW Tariff	%US Imports
Textiles:						
Broadwoven fabric mills	8.5	9.5	14.4	13.3	22.8	0.48
Narrow fabric mills	3.4	3.3	6.9	6.7	10.0	0.22
Yarn mills and textile finishing	5.1	3.1	10.0	8.5	11.6	0.06
Thread mills	4.6	2.2	9.5	11.8	14.0	0.01
Floor coverings	2.8	9.3	7.8	5.7	15.0	0.12
Felt and textile goods, n.e.c.	1.0	0.1	4.7	6.2	6.3	0.06
Lace and knit fabric goods	3.8	5.9	13.5	11.8	17.7	0.04
Coated fabrics, not rubberized	2.0	1.0	9.8	6.6	7.6	0.03
Tire cord and fabric	2.3	2.4	5.1	4.4	6.8	0.08
Cordage and twine	3.1	1.2	6.2	3.6	4.8	0.03
Nonwoven fabric	0.1	0.2	10.6	9.5	9.7	0.04
Apparel and fab. textile products:						
Women's hosiery, except socks	5.4	2.3				
Hosiery, n.e.c.	3.5	2.4	14.9	15.3	17.7	0.04
App'l made from purchased mat'l	16.8	19.9	13.2	12.6	32.5	5.71
Curtains and draperies	5.9	12.1	11.9	12.1	24.2	0.01
House furnishings, n.e.c.	8.3	13.9	9.3	8.2	22.1	0.27
Textile bags	5.9	9.0	6.4	6.6	15.6	0.01
Canvas and related products	6.3	5.2	6.9	6.4	11.6	0.03
Pleating, stitching, ... embroidery	5.2	7.6	8.0	8.1	15.7	0.02
Fabricated textile products, n.e.c.	9.2	0.6	5.2	4.8	5.4	0.37
Luggage	2.6	10.4	12.1	10.8	21.2	0.28
Women's handbags and purses	1.0	3.1	10.5	6.7	9.8	0.44

Notes: "S" indicates "simple" and "TW" indicates "trade-weighted." Rent equivalents for U.S. imports from Hong Kong were estimated on the basis of average weekly Hong Kong quota prices paid by brokers, using information from International Business and Economic Research Corporation. For countries that do not allocate quota rights in public auctions, export prices were estimated from Hong Kong export prices, with adjustments for differences in labor costs and productivity. Sectors and their corresponding SIC classifications are detailed in USITC (1995) Table D-1. Quota tariff equivalents are reproduced from Deardorff and Stern (1998), Table 3.6 (Source USITC 1993,1995). Tariff averages, trade-weighted tariff averages and U.S. import percentages are calculated using data from the UNCTAD TRAINS dataset. SIC to HS concordances from the U.S. Census Bureau are used.

(i) substantial restrictiveness of MFA quotas and (ii) very large differentials in quota premia across commodity lines and across exporters.

Price comparison measures confirm this picture of the high and highly concentrated nature of NTBs with data from the agricultural sector. European and Japanese agriculture is even more highly protected than U.S. and Canadian agriculture. Details are discussed in section 4.

Using a variety of methods, Messerlin (2001) makes a notably ambitious attempt to assemble tariff equivalents of all trade policy barriers for the European Union. He combines the NTB tariff equivalents with the MFN tariffs. For 1999 the tariff equivalent of policy barriers were 5 percent for cereals, 64.8 percent for meat, 100.3 percent for dairy and 125 percent for sugar. In mining, the tariff equivalent was 71.3 percent. The combined arithmetic average protection rate was 31.7 percent in agriculture, 22.1 percent in textiles, 30.6 percent in apparel, and much less in other industrial goods. The combined arithmetic average protection rate in industrial goods was 7.7 percent. We use 7.7 percent for our representative trade policy cost.

2.2 *Transport Costs*

Direct transport costs include freight charges and insurance which is customarily added to the freight charge. Indirect transport user costs include holding cost for the goods in transit, inventory cost due to buffering the variability of delivery dates, preparation costs associated with shipment size (full container load vs. partial loads), and the like. Indirect costs must be inferred.

2.2.1 *Measurement Problems and Limitations*

There are three main sources of data for transport costs. The most direct is industry or shipping firm information. Nuno Limao and Anthony Venables (2001) obtain quotes from shipping firms for a standard container shipped from Baltimore to various destinations. David Hummels (2001a) obtains indices of ocean shipping and air freight rates from trade journals which presumably are averages of such quotes. Direct methods are best but not always feasible due to data limitations and the very large size of the resulting datasets.

Alternatively, there are two sources of information on average transport costs. National customs data in some cases allow fine detail. For example, the U.S. Census provides data on U.S. imports at the ten-digit Harmonized System level by exporter country, mode of transport and entry port, by weight (where available) and valued at f.o.b. and c.i.f. bases. Dividing the former value into the latter yields an ad-valorem estimate of bilateral transport cost. Hummels (2001a) makes use of this source for the United States and several other countries. The most widely available (many countries and years are covered) but least satisfactory average ad valorem transport costs are the aggregate bilateral c.i.f./f.o.b ratios produced by the IMF from matching export data (reported f.o.b.) to import data (reported c.i.f.).[19] The IMF uses the UN's COMTRADE database, supplemented in some cases with national data sources. Hummels (2001a) points out that a high proportion of observations are imputed; this and the compositional shifts in aggregate trade flows which occur over time lead him to conclude that "quality problems should disqualify these data from use as a measure of transportation costs in even semi-careful studies."[20]

2.2.2 *Evidence on Transport Costs*

Hummels (2001a) shows the wide dispersion in freight rates over commodities and across countries in 1994. The all-commodities trade-weighted average transport cost from national customs data ranges from 3.8 percent of the f.o.b. price for the United States to 13.3 percent for Paraguay. The all-commodities arithmetic average ranges from 7.3 percent for Uruguay to 17.5 percent for Brazil. The U.S. average is 10.7 percent. Across commodities for the United States the range of trade-weighted averages is from less than 1 percent (for transport equipment) to 27 percent for crude fertilizer. The arithmetic averages range from 5.7 percent for machinery and transport equipment to 15.7 percent for mineral fuels.

Hummels (1999) considers variation over time. The overall trade-weighted average transport cost for the United States declined over the last thirty years, from 6 percent to 4

[19] See the *Direction of Trade Statistics* and the *International Financial Statistics.*

[20] Nevertheless, because of their availability and the difficulty of obtaining better estimates for a wide range of countries and years, apparently careful work such as Harrigan (1993) and Scott Baier and Jeffrey Bergstrand (2001) uses the IMF data.

percent. Composition problems are acute because world trade in high-value-to-weight manufactures has grown much faster than trade in low-value-to-weight primary products. Hummels shows that air freight cost has fallen dramatically, while ocean shipping cost has risen (along with the shift to containerization, which improves the quality of the shipping service). He also documents the wide dispersion in the rate of change of air-freight rates across country pairs over the past forty years.

Notice that alongside tariffs and NTBs, transport costs look to be comparable in average magnitude and in variability across countries, commodities, and time. Transport costs tend to be higher in bulky agricultural products where protection in OECD countries is also high. Policy protection tends to complement natural protection, amplifying the variability of total trade costs.

Limao and Venables (2001) emphasize the dependence of trade costs on infrastructure. They gather price quotes for shipment of a standardized container from Baltimore to various points in the world. Infrastructure is measured as an average of the density of the road network, the paved road network, the rail network, and the number of telephone main lines per person. A deterioration of infrastructure from the median to the 75th percentile of destinations raises transport costs by 12 percent. The median landlocked country has transport costs that are 55 percent higher than the median coastal economy.[21] The infrastructure variables also have explanatory power in predicting trade volume. Inescapably, understanding trade costs and their role in determining international trade volumes must incorporate the internal geography of countries and the associated interior trade costs.

As for indirect costs, Hummels (2001b) imputes a willingness-to-pay for saved time. Each day in travel is on average worth 0.8 percent of the value of manufactured goods, equivalent to a 16-percent ad-valorem tariff for the average-length ocean shipment. Shippers switch from ocean to air in his mode choice model when the full (shipping plus time) cost of ocean exceeds that for air.[22] The use of averages here masks a lot of variation in the estimated value of time across two-digit manufacturing sectors, and is subject to upward aggregation bias due to larger growth in trade where savings are greatest due to the substitution effect. Infrastructure is likely to have a considerable effect on the time costs of trade.

In 1998, half of U.S. shipments was by air and half by boat.[23] Assigning one day to shipment by air anywhere in the world, as Hummels does, and using the twenty-day average for ocean shipping, leads to an average 9-percent tax equivalent of time costs. Hummels argues that faster transport (shifting from shipping to air, and faster ships) has reduced the tax equivalent of time costs for the United States from 32 percent to 9 percent over the period 1950–98.[24]

We combine 9-percent time costs and 10.7 percent U.S. average direct transport costs for our representative full transport cost of 21 percent (1.107*1.09-1).

[21] Limao and Venables also report similar results using the c.i.f./f.o.b. ratios of the IMF.

[22] Linking port of entry for U.S. imports with the travel time to the exporter (a country-average of times to the exporter's ports), he creates a matrix of ocean shipping times. Air freight is assumed to take one day for points anywhere in the world.

[23] This ignores trucking and rail modes, which are important to trade with Canada and Mexico, the two largest trade partners of the United States.

[24] The calculation is based on observing that U.S. imports, excluding Canada and Mexico, had 0-percent air shipment in 1950 and 50-percent air shipment in 1998. The average ocean shipment time was halved from forty days to twenty days over the same time. The net effect is a saving of 29.5 days, equal to forty days for 1950 minus 10.5 days for 1998. The latter is equal to .5 times twenty days for ocean shipping plus .5 times one day for air freight. For manufacturing, at 0.8 per cent ad valorem per day for the value of time in shipping, the saving of 29.5 days is worth a fall from 32 percent to 9 percent ad valorem.

TABLE 6
DISTRIBUTION MARGINS FOR HOUSEHOLD CONSUMPTION AND CAPITAL GOODS

Select Product Categories	Aus. 95	Bel. 90	Can. 90	Ger. 93	Ita. 92	Jap. 95	Net. 90	UK 90	US 92
Rice	1.239	1.237	1.867	1.423	1.549	1.335	1.434	1.511	1.435
Fresh, frozen beef	1.485	1.626	1.544	1.423	1.605	1.681	1.640	1.390	1.534
Beer	1.185	1.435	1.213	1.423	1.240	1.710	1.373	2.210	1.863
Cigarettes	1.191	1.133	1.505	1.423	1.240	1.398	1.230	1.129	1.582
Ladies' clothing	1.858	1.845	1.826	2.039	1.562	2.295	1.855	2.005	2.159
Refrigerators, freezers	1.236	1.586	1.744	1.826	1.783	1.638	1.661	2.080	1.682
Passenger vehicles	1.585	1.198	1.227	1.374	1.457	1.760	1.247	1.216	1.203
Books	1.882	1.452	1.294	2.039	1.778	1.665	1.680	1.625	1.751
Office, data proc. mach.	1.715	1.072	1.035	1.153	1.603	1.389	1.217*	1.040	1.228
Electronic equip., etc.	1.715	1.080	1.198	1.160	1.576	1.432	1.224*	1.080	1.139
Simple Average (125 categories)	1.574	1.420	1.571	1.535	1.577	1.703	1.502	1.562	1.681

Notes: The table is reproduced from Bradford and Lawrence, "Paying the Price: The Cost of Fragmented International Markets", Institute of International Economics, forthcoming (2003). Margins represent the ratio of purchaser price to producer price. Margins data on capital goods are not available for the Netherlands, so an average of the four European countries' margins is used.

2.3 *Wholesale and Retail Distribution Costs*

Wholesale and retail distribution costs enter retail prices in each country. Since wholesale and retail costs vary widely by country, this would appear to affect exporters' decisions. Local trade costs apply to both imported and domestic goods, however, so relative prices to buyers don't change and neither does the pattern of trade. Section 3 gives a formal argument. Section 4 discusses the effect of distribution margins on inference about international trade costs from retail prices.

Ariel Burstein, Joao Neves, and Sergio Rebelo (2003) construct domestic distribution costs from national input-output data for tradable consumption goods (which correspond most closely to the goods for which narrowly defined trade costs are relevant). They report a weighted average of 41.9 percent for the United States in 1992 as a fraction of the retail price. They also show that their input-output estimates of U.S. distribution costs are roughly consistent with survey data from the U.S. Department of Agriculture for agricultural goods and from the 1992 Census of Wholesale and Retail Trade. For other G-7 countries they report distribution costs in the range of 35–50 percent.

Scott Bradford and Robert Lawrence (2003) use the same input-output sources to measure distribution costs for the United States and eight other industrialized countries, but instead divide by the producer price, consistent with the approach in this survey of reporting trade barriers in terms of ad valorem tax equivalents. Table 6 reports distribution costs for selected tradable household consumption goods and an arithmetic average for 125 goods. The averages range over countries from 42 percent in Belgium to 70 percent in Japan. Average U.S. distribution costs are 68 percent of producer prices. The range of distribution costs is much larger across goods than across countries, for example running from 14 percent on electronic equipment to 216 percent on ladies clothing in the United States.[25] We take 55 percent as our representative domestic distribution cost, close to the tradable consumption goods average of OECD countries.

3. *Inference of Trade Costs from Trade Flows*

Trade costs can be inferred from an economic model linking trade flows to observable variables and unobservable trade costs. Inference has mainly used the gravity model.[26] The economic theory of gravity is developed here extensively, revealing new properties that clarify procedures for good empirical work, reporting results, and doing sensible comparative statics. Our development embeds gravity within the classic concerns of trade economists with the equilibrium allocation of production and expenditure within nations.

A variety of ad hoc trade cost functions have been used to relate the unobservable cost to observable variables. Plausibly, cost falls with common language and customs, better information, better enforcement and so forth. Many implementations impose restrictions that seem implausible. Some proxies may not be exogenous, such as membership in a currency union or regional trade agreement; their effects will not be uniform; functional form is often too simple and so forth. Further economic theory is needed to identify the underlying structure of trade costs.

3.1 *Traditional Gravity*

Most estimated gravity equations take the form

$$x_{ij} = \alpha_1 y_i + \alpha_2 y_i + \sum_{m=1}^{M} \beta_m \, ln(z_{ij}^m) + \varepsilon_{ij} \qquad (1)$$

where x_{ij} is the log of exports from i to j, y_i and y_j are the log of GDP of the exporter and importer, and z_{ij}^m $(m=1,\ldots,M)$ is a set of observables to which bilateral trade barriers are related. The disturbance term is ε_{ij}. A large recent literature developed estimating this type of gravity equation after a surprising finding by McCallum (1995) that the U.S.-Canada border has a big impact on trade.[27] McCallum estimated a version of (1) for U.S. states and provinces with two z variables: bilateral distance and an indicator variable that is equal to one if the two regions are located in the same country and equal to zero otherwise. He found that trade between provinces is more than twenty times trade between states and provinces after controlling for distance and size. The subsequent literature has often added so-called remoteness variables, which are intended to capture the average distance of countries or regions from their trading partners.

Gravity equations can be derived from a variety of different theories. None lead to traditional gravity, despite this literature's use of references such as Anderson (1979) and Deardorff (1998) to justify estimation of (1).[28]

3.2 *Theory-Based Gravity*

Gravity-like structure obtains in a wide class of models, those where the allocation of trade across countries can be analyzed

[25] We should warn that some of these distribution cost estimates include rents in the form of monopolistic markups rather than actual costs incurred by the local distribution sector. Although it would be desirable to take out these rents, no data are available to do so.

[26] A subset of the gravity literature uses it to discriminate among theories of the determinants of trade. It is not very well-suited for this purpose, since many trade models will lead to gravity (Deardorff 1998).

[27] See Michael Anderson and Stephen Smith (1999a,b), Natalie Chen (2002), Carolyn Evans (2003a,b,c), John Helliwell (1996, 1997, 1998), Helliwell and John McCallum (1995), Helliwell and Genevieve Verdier (2001), Russell Hillberry (1998, 2002), Volker Nitsh (2000), Shang-Jin Wei (1996), and Holger Wolf (2000a,b).

[28] For recent surveys of gravity theory, see Harrigan (2002), and Feenstra (2002, 2003).

separately from the allocation of production and consumption within countries.[29] Let $\{Y_i^k, E_i^k\}$ be the value of production and expenditure in country i for product class k. A product class can be either a final or an intermediate good. A model is *trade separable* if the allocation of $\{Y_i^k, E_i^k\}$ for each country i is separable from the bilateral allocation of trade across countries.

Trade separability obtains under the assumption of separable preferences and technology. Each product class has a distinct natural aggregator of varieties of goods distinguished by country of origin. This assumption allows the two-stage budgeting needed to separate the allocation of expenditure across product classes from the allocation of expenditure within a product class across countries of origin.[30] What exactly is implied by product differentiation within a product class will differ depending on the product class and its myriad characteristics. For inputs of physically standardized products (grains, petroleum, steel plate), differentiation may be on terms of delivery or subtle qualities that affect productivity while branded final goods suggest hedonic differentiation over varieties of clothing or perfume. Imposing some form of differentiation is unavoidable if empirical trade models are to fit the bilateral trade data, as Armington noted long ago.

The class of trade separable models yields bilateral trade without having to make any assumptions about what specific model accounts for the observed output structure $\{Y_i^k\}$ and expenditure allocations $\{E_i^k\}$.[31] Bilateral trade is determined in *conditional general equilibrium* whereby product markets for each good (each brand) produced in each country clear conditional on the allocations $\{Y_i^k, E_i^k\}$. Inference about trade costs is based on this conditional general equilibrium. Comparative static analysis, in contrast, requires consideration of the full general equilibrium. A change in trade barriers, for example, will generally affect the allocations $\{Y_i^k, E_i^k\}$.

Two additional restrictions to the class of trade separable models yield gravity. These are: the aggregator of varieties is identical across countries and CES; and ad-valorem tax equivalents of trade costs do not depend on the quantity of trade. The CES form imposes homothetic preferences and the homogeneity equivalent for intermediate input demand. These assumptions simplify the demand equations and market clearing equations. These two restrictions can be relaxed in various useful ways discussed below. Our derivation provides much more context to Deardorff's (1998) remark, "I suspect that just about any plausible model of trade would yield something very like the gravity equation …".

If X_{ij}^k is defined as exports from i to j in product class k, the CES demand structure implies (under the expositional simplification of equal weights for each country of origin)

$$X_{ij}^k = \left(\frac{p_{ij}^k}{P_j^k} \right)^{1-\sigma_k} E_j^k \qquad (2)$$

where σ_k is the elasticity of substitution among brands, p_{ij}^k is the price charged by i for exports to j and P_j^k is the CES price index:

[29] The general equilibrium of trade in a many-country world is enormously simplified by this natural assumption.

[30] Explicitly, separability restricts the dual cost function $c(p,w,y)$ where p is a vector of traded input prices and w is a vector of non-traded primary input prices while y is the level of output. The separability restriction is imposed in $c(p,w,y)=f[g(p),w,y]$ where $g(p)$ is a homogenous of degree one and concave function of p By Shephard's Lemma, f_g is the aggregate demand for the traded input class while $f_g g_p$ is the demand vector for the individual products. A similar structure characterizes the assumption of separability in final demand.

[31] Specifically, it does not matter what one assumes about production functions, technology, the degree of competition, or specialization patterns. The nature of preferences and technology that gives rise to the observed expenditure allocations E_i^k also does not matter.

$$P_j^k = \left[\sum_i (p_{ij}^k)^{1-\sigma_k} \right]^{1/(1-\sigma_k)}. \quad (3)$$

The assumption that trade costs are proportional to trade implies that the price p_{ij}^k can be written as $p_i^k t_{ij}^k$, where p_i^k is the "supply price" received by producers in country k and t_{ij}^k-1 is the ad-valorem tax equivalent of trade costs, independent of volume.

If supply is monopolistic, the "supply price" is the product of marginal production cost and the markup. So long as the markup is invariant over destinations,[32] t_{ij}^k contains only trade costs. Otherwise the tax equivalent must be interpreted to contain markups. With competitive supply, this issue does not arise. Markups arising from monopoly power in the distribution sector itself are a more important issue with interpreting t_{ij}^k-1 as a cost.

Imposing the market-clearing conditions

$$Y_i^k = \sum_j X_{ij}^k \quad (4)$$

for all i and k yields gravity. Solve for the supply prices p_i^k from the market-clearing conditions and substitute the result in (2) and (3). This yields the system

$$X_{ij}^k = \frac{E_j^k Y_i^k}{Y^k} \left(\frac{t_{ij}^k}{P_j^k \Pi_i^k} \right)^{1-\sigma_k} \quad (5)$$

$$(\Pi_i^k)^{1-\sigma_k} = \sum_j \left(\frac{t_{ij}^k}{P_j^k} \right)^{1-\sigma_k} \frac{E_j^k}{Y^k} \quad (6)$$

$$(P_j^k)^{1-\sigma_k} = \sum_i \left(\frac{t_{ij}^k}{\Pi_i^k} \right)^{1-\sigma_k} \frac{Y_i^k}{Y^k} \quad (7)$$

where Y^k is world output in sector k. The indices P_j^k and Π_i^k can be solved as a function of trade barriers $\{t_{ij}^k\}$ and the entire set $\{Y_i^k, E_i^k\}$. Trade flows therefore also depend on trade barriers and the set $\{Y_i^k, E_i^k\}$.[33]

A wide variety of production and expenditure models may lie behind $\{E_j^k, Y_i^k\}$. In particular, some of the E's and Y's may be equal to zero. Previous derivations of gravity have usually made much more restrictive assumptions.[34]

The main insight from the theory is that bilateral trade depends on *relative* trade barriers. The key variables Π_i^k and P_j^k are outward and inward multilateral resistance, respectively. They summarize the average trade resistance between a country and its trading partners in an ideal aggregation sense, which we develop below. Basic demand theory suggests that the flow of good i into j is increased (given $\sigma>1$) by high trade costs from other suppliers to j as captured by inward multilateral resistance P_j^k. But, less obviously, high resistance to shipments from i to its other markets, captured in outward multilateral resistance Π_i^k, tips more trade back into i's market in j. Gravity gives these insights an elegantly simple form in (5).

The trade cost t_{ij}^k may include local distribution costs in the destination market, but those domestic costs do not affect trade flows. This rather surprising and important result follows from basic gravity theory. Suppose for any goods class k that each destination j has its own domestic margin m_j^k. If we multiply all

[32] With Cournot competition, the markup is invariant over destinations in symmetric monopolistic competition. Generally it is equal to $1/[1-1/\sigma(1-s_{ij})]$ where s_{ij} is the market share of i in j.

[33] Theoretical gravity equations in Anderson (1979), Jeffrey Bergstrand (1985, 1989, 1990) and Scott Baier and Bergstrand (2001) look far more complicated than (5), with a large number of prices and price indices. Anderson and van Wincoop (2003) show that all these prices can be summarized by just two price indices, one for the importer and one for the exporter, which are solved as a function of trade barriers and total supply and demand in each location.

[34] The first paper to formally derive a gravity equation from a general equilibrium model with trade costs is Anderson (1979), who assumes that every country produces a particular variety. Bergstrand (1989, 1990) and Baier and Bergstrand (2001) derive gravity equations in models with monopolistic competition, endogenizing variety. Anderson and van Wincoop (2003) assume each country has an endowment of its good.

narrowly defined trade barriers t_{ij}^k by domestic trade costs m_j^k in the destination market, it is easily verified from (6)–(7) that the P_j^k are multiplied by m_j^k, the Π_i^k are unchanged, and therefore trade flows are also unchanged.

The invariance of trade patterns to domestic distribution costs that apply to all goods has another important practical implication. We can only identify relative costs with the gravity model. One way to interpret an inferred system of trade costs $\{t_{ij}^k\}$ is to pick some region i and normalize $t_{ii}^k=1$. Essentially this procedure treats the trade cost of i with itself as a pure local distribution cost and divides all other trade costs by the local distribution cost in region i.

As in most gravity papers, Anderson and van Wincoop (2003) consider a one-sector economy. They show that when consumers have CES preferences with common elasticity σ among all goods, the gravity equation can be written, omitting the superscripts k, as

$$X_{ij} = \frac{Y_i Y_j}{Y_w}\left(\frac{t_{ij}}{\Pi_i P_j}\right)^{1-\sigma} \tag{8}$$

$$P_j^{1-\sigma} = \sum_i \Pi_i^{\sigma-1}\theta_i t_{ij}^{1-\sigma} \forall j \tag{9}$$

$$\Pi_i^{1-\sigma} = \sum_j P_j^{\sigma-1}\theta_j t_{ij}^{1-\sigma}, \forall i. \tag{10}$$

where Y_i and Y_j are levels of GDP, Y_w is world GDP, and θ_i is the income share of country i. This is a special case of (5) with expenditure equal to output because it is a one-sector economy ($E_i=Y_i$). With symmetry of trade costs ($t_{ij}=t_{ji}$), $\Pi_i=P_i$.

A couple of comments are in order about homogeneous goods trade. When we let the elasticity of substitution σ_k in (5)–(7) go to infinity, trade converges to that in a homogeneous goods model. However, no information about trade barriers can be inferred. As an example, consider a two-country model with trade in a homogeneous good k. If country 1 exports k to country 2, its export is equal to $E_2^k–Y_2^k$. Gravity equation (5), accounting for multilateral resistance, indeed converges to this for the two-country case as σ_k approaches infinity, assuming any non-zero international trade barrier (and normalizing domestic barriers to zero). Since the bilateral trade flow in the conditional general equilibrium does not depend on the magnitude of trade barriers, nothing can be learned about trade barriers. More generally, it is difficult to learn much about trade barriers from a gravity equation for sectors where the elasticity of substitution is extremely high. The expressions $t_{ij}^{1-\sigma_k}$ on which trade flows (directly and through multilateral resistance) depend are virtually zero when σ_k is very high as long as the trade barrier is positive ($t_{ij}^k>1$). Conditional on a trade barrier being positive, the size of the trade barrier does not matter much to the pattern of trade flows in the conditional general equilibrium, hence we cannot learn much about their size.

Several authors have derived gravity equations for homogeneous goods trade when trade is an aggregate of a variety of homogeneous goods. Deardorff (1998) derived a gravity equation in the Heckscher Ohlin model with complete specialization. This is essentially a differentiated goods model though, with each country producing a different brand. It does not mean much to say that a good is homogeneous when there is only one producer.

A real homogeneous goods model, with multiple producers of the same homogeneous good, is the Ricardian model of Jonathan Eaton and Samual Kortum (2002). Their model leads to a gravity equation for an aggregate of homogeneous goods. It is also a model with trade separability, although the rationale is somewhat different. Production is Ricardian, with the cost of production in country i in good k given by $c_i/z(k)$ where $z(k)$ is the realization of technology in good k, an element in a continuum of goods. Productivity is drawn from a Fréchet distribution. The distribution has two parameters. The first is T_i, with higher T_i meaning a higher average realization for country i. The second is θ, with a larger value implying lower productivity differences across countries. For a particular good, users always buy from the cheapest source. The price is the production cost times the trade cost t_{ij}. Each good is produced with both labor and a bundle of intermediate goods that consists of the same CES index of all final goods as the utility function over final goods.

Since there is a continuum of goods and the setup is the same for all goods (same production function, same productivity distribution, same trade cost), the fraction that country j spends on goods from i is equal to the probability for any particular good that j sources from i. With the assumed Fréchet distribution this is equal to

$$\frac{X_{ij}}{E_j} = \frac{T_i(c_i t_{ij})^{-\theta}}{\sum_i T_i(c_i t_{ij})^{-\theta}}.$$

The probability of shipment from country i is lowered by the trade cost of getting the good to country j, relative to the average trade cost of shipping from all other destinations, and lowered by a higher cost of labor. The same mathematical representation of the allocation of trade arises as with the CES structure of demand for goods differentiated by place of origin. This equation is the same as (2), with $\sigma=\theta+1$ and $p_i^{1-\sigma}$ replaced by $T_i c_i^{1-\sigma}$. The p_i is essentially replaced by c_i, which can be solved in the same way from the conditional general equilibrium. This gives rise to the same gravity equation as before.[35]

It is worth noting that gross output is now larger than net output due to the input of intermediates. The output in the gravity equation (8) is gross output. Since Eaton and Kortum assume that intermediates are a fraction $1-\beta$ of the production cost c_i, with labor a fraction β, gross output is $1/\beta$ times value added. If we interpret Y_i in (8) as value added, the gravity equation must be multiplied by $1/\beta$.

3.3 *The Trade Cost Function*

The theoretical gravity model allows inference about unobservable trade costs by (i) linking trade costs to observable cost proxies and (ii) making an assumption about error terms which link observable trade flows to theoretically predicted values. Here we focus on (i), the next section deals with (ii). For now we will focus on inference about trade barriers from the aggregate gravity equation (8). In a section about aggregation below we will return to the disaggregated gravity equation (5).

Bilateral trade barriers are assumed to be a function of observables z_{ij}^m, commonly loglinear:

$$t_{ij} = \prod_{m=1}^{M} (z_{ij}^m)^{\gamma_m} \qquad (11)$$

Normalizing such that $z_{ij}^m=1$ measures zero trade barriers associated with this variable, $(z_{ij}^m)^{\gamma_m}$ is equal to one plus the tax equivalent of trade barriers associated with variable m. The list of observable arguments z_{ij}^m which have been used in the trade cost function in the literature includes directly measured trade costs, distance, adjacency, preferential trade membership, common language, and a host of others. Gravity theory has used arbitrary assumptions regarding functional form of the trade cost function, the list of variables, and regularity conditions.

As an illustration of the functional form problem, consider distance. By far the most common assumption is that $t_{ij}=d_{ij}^{\rho}$. Gene Grossman (1998) argues that a more reasonable assumption is that $\tau_{ij}=t_{ij}-1=d_{ij}^{\rho}$ since one can think of τ_{ij} as transport costs per dollar of shipments. Hummels (2001a) estimates the ρ in the second specification by using data on ad-valorem freight rates and finds a value of about 0.3. Limao and Venables (2001) estimate the first specification using c.i.f./f.o.b.

[35] Eaton and Kortum only derive a gravity specification for X_{ij}/X_{ii}.

data and also find an estimate of ρ of about 0.3. Although these numbers are the same, they are inconsistent with each other. If the Grossman specification is correct with ρ=0.3, one would expect a distance elasticity of t_{ij} of $0.3\tau_{ij}/(1+\tau_{ij})$, evaluated at some average τ, which is much less than 0.3. Highly misleading results for trade barrier estimates arise when the wrong functional form is adopted.

Eaton and Kortum (2002) generalize the treatment of distance with a flexible form which can approximate both of the preceding specifications. They assume that there are different trade barriers for six different distance intervals. While implicitly they still assume a particular functional form, in the form of a step function, this spline approach is likely to be more robust to specification error.

Another functional form issue is that the most common setup (11) is multiplicative in the various cost factors. Hummels (2001a) argues that an additive specification is more sensible. A multi-dimensional generalization of the approach by Eaton and Kortum (2002) may be applied, although there is a tradeoff between degrees of freedom and generality of the specification. To the extent that theory has something to say about the functional form, it is preferable to use this information over econometric solutions that waste degrees of freedom.

The second problem is which observables to include. Empirical practice can improve with a more theoretical approach to the z's. Especially for abstract trade barriers such as information costs, it is often unclear what specific variables are meant to capture. Even in the absence of a specific theory, it is useful to ponder the relationship between trade barriers and observed variables. For example, common empirical practice uses a language variable that is one if two countries speak the same language and zero otherwise. Jacques Melitz (2003) considers ways in which language differences affect trade and develops several variables that each capture different aspects of communication. One such variable is "direct communication," which depends on the percentages of people in two countries that can speak the same language. Another is the binary variable "open-circuit communication," which is one if two countries have the same official language or the same language is spoken by at least 20 percent of the populations of both countries. The first variable reflects that trade requires direct communication, while the second variable is meant to capture an established network of translation. Another example is distance. It is common to model distance as the Great Circle distance between capitals. Where these differ from commercial centers it is sometimes taken to be superior to use distance between commercial centers. But then what of countries with more than one commercial center, of interior infrastructure?[36]

Implausibly strong regularity (common coefficients) conditions are often implicitly imposed on the trade cost function. For example, the effect of membership in a customs union or a monetary union on trade costs is often assumed to be uniform for all members. As for customs unions, a uniform external tariff is indeed approximately the trade policy (though NTBs remain inherently discriminatory), while free trade agreements continue to have different national external tariffs and thus different effects. As for monetary unions, the effect of switching from national to common currencies is likely to be quite different depending on the national currency. Similar objections can be raised to a number of the other commonly used proxies z_{ij}^m such as common language or adjacency dummies. NTBs present an acute form of this problem. The effect on trade barriers of NTBs in a country i will generally vary across trading partners j, goods k, and time t. Regression residuals reflect the NTBs but also random error. To identify the tariff equivalent of NTBs Harrigan (1993)

[36] Some investigators (e.g., Bergstrand) measure bilateral distance between ports, supplemented by twice the land distance between ports and commercial centers, reflecting the rough difference in cost between water and land shipment.

assumes, not very plausibly, that the importing country's NTB has the same trade displacement effect for each exporter i that it buys good k from. Trefler (1993) assumes even less plausibly that U.S. NTBs have the same trade-reducing effect for all goods k that it imports from the rest of the world.[37]

We sympathize with efforts to identify trade costs with simple forms of (11). Our criticism of the ad hoc functional form and the regularity assumptions aims to stimulate improvement in estimation and useful comparative statics. Unpacking the reduced form to its plausible structural elements will aid both.

3.4 *Estimation of Trade Barriers*

Given the trade cost function, the logarithmic form of the empirical gravity equation becomes (dropping the constant term)

$$x_{ij} = y_i + y_j + \sum_{m=1}^{M} \lambda_m \, ln(z_{ij}^m) - (1-\sigma) ln(\Pi_i) - (1-\sigma) ln(P_j) + \varepsilon_{ij} \qquad (12)$$

where $x_{ij}=ln(X_{ij})$, $y_i=ln(Y_i)$, and $\lambda_m=(1-\sigma)\gamma_m$, x_{ij} and y_i are observables, and ε_{ij} is the error term.

The theoretical gravity equation can be estimated in three different ways.[38] Anderson and van Wincoop (2003) estimate the structural equation with nonlinear least squares after solving for the multilateral resistance indices as a function of the observables z_{ij}^m (bilateral distances and a dummy variable for international borders) and the parameters λ_m. Another approach, which also gives an unbiased estimate of the parameters λ_m, is to replace the inward and outward multilateral resistance indices and production variables, $y_j-(1-\sigma)\ln P_j$ and $y_j-(1-\sigma)ln(\Pi_i)$, with inward and outward region-specific dummies. With symmetry, a single set of region-specific dummies suffice. This approach is adopted by Anderson and van Wincoop (2003), Eaton and Kortum (2002), Asier Minondo (2002), Andrew Rose and van Wincoop (2002), and Hummels (2001a). Keith Head and Thierry Mayer (2001) and Head and John Ries (2001) follow an estimation approach that is identical to replacing multilateral resistance variables with country dummies in the case where internal trade data X_{ii} exist for all regions or countries. Assuming that $z_{ij}=z_{ji}$, it follows from (12) that

$$ln\left(\frac{X_{ii}X_{jj}}{X_{ij}X_{ji}}\right)^{0.5} = \sum_{m=1}^{M} \lambda_m \, ln\left(\frac{(z_{ii}^m z_{jj}^m)^{0.5}}{z_{ij}^m}\right). \qquad (13)$$

The parameters λ_k can then be estimated through a linear regression.

A third method is to use data for the price indices and estimate with OLS. This requires data on price levels for a cross-section regression or changes in price indices when there are at least two years of data. The latter is the approach taken by Bergstrand (1985, 1989, 1990), Baier and Bergstrand (2001), and Head and Mayer (2000). As discussed in Baier and Berstrand (2001), it is often hard or impossible to measure the theoretical price indices in the data. Price indices, such as the consumer price index, also include nontradables and are affected by local taxes and subsidies. Nominal rigidities also affect observed prices, and have a big impact on relative prices when combined with nominal exchange rate fluctuations. Anderson and van Wincoop (2003) also argue that certain trade barriers, such as a home bias in preferences, do not show up in prices. Similarly, Deardorff and Stern (1998) explain why certain NTBs affect trade but not prices.

[37] Both authors implicitly impose a further regularity condition. NTB coverage ratios for each good are the explanatory variable, so all changes in this ratio are assumed to be equally important.

[38] A potential fourth method is to infer all bilateral trade barriers for a group of countries or regions from the residuals of the bilateral trade flows from the prediction of the frictionless gravity model. The information in this measure would be drowned in random error, however: there is an unboundedly large confidence interval about the point estimates because all degrees of freedom are used up.

Feenstra (2003) sums it up by writing that "the myriad of costs ... involved in making transactions across the border are probably not reflected in aggregate price indices ...". This does not mean that prices of individual tradable goods are entirely uninformative about trade costs. We turn to that topic in section 4.

The tax equivalent of trade barriers between i and j associated with variable m is estimated as

$$\left(z_{ij}^{m}\right)^{\lambda_m/(1-\sigma)} - 1 \approx \lambda_m (z_{ij}^{m} - 1)/(1-\sigma). \quad (14)$$

This shows that we need an estimate of σ in order to obtain an estimate of trade barriers. Assumptions about σ can make quite a difference. For example, the estimated tax equivalent is approximately nine times larger when using σ=2 instead of σ=10.

Gravity can only measure trade barriers relative to some benchmark, as noted above. The literature tends to compare trade barriers between countries to barriers within countries, or barriers between regions to barriers within regions. This is problematic since different countries or regions have different barriers for internal trade. The results will also depend on the measure of barriers within a region or country, such as the treatment of distance within a country or region. This is essentially an aggregation problem since a country or region is itself an aggregate. Head and Mayer (2001), Helliwell and Verdier (2001), and Russell Hillberry and David Hummels (2002a) show that the different measures of internal distance can make a big difference for the results.

The error term in the empirical gravity literature is usually taken to reflect measurement error. Trade flow data are notoriously rife with measurement error,[39] which is taken to justify a normally distributed additive error term which is orthogonal to the regressors. More careful consideration of measurement error is likely to be productive. Some recent implementations recognize the panel nature of the data in constructing standard errors (all bilateral observations from i or into j may have common disturbance elements).[40] Improved econometric techniques based on careful consideration of the error structure are likely to pay off. Recent literature on spatial econometrics (e.g. H. H. Kelejian and Ingmar Prucha 1999) may be helpful.

The error term may also reflect unobservable variables in the trade cost function (11). If the trade cost t_{ij} is multiplied by $e^{\varepsilon_{ij}}$, there is again an additive error term in the empirical gravity equation (12). In this case the structural estimation technique discussed above would have to be modified since the multilateral resistance variables also depend on these error terms. But the second estimation technique, replacing the multilateral resistance variables with region or country-specific dummies, is still appropriate. Eaton and Kortum (2002) introduce the error term in this way and adopt the second estimation method.

Non-orthogonality of the error term has two sources, omitted variables and endogeneity. If the error term reflects omitted trade frictions which are correlated with the included z_{ij}^{m} variables, it causes estimation bias. Endogeneity problems arise in several ways. Concerned that trade can affect output, many papers use instrumental variables for output Y_i and Y_j. The most common instruments are population or factor endowments.[41] With coefficients on the y's constrained to one, the natural implementation is to make the dependent variable size-adjusted trade $x_{ij}-y_i-y_j$, as Anderson and van Wincoop (2003) do. Endogeneity issues also arise when the proxies for trade costs are

[39] For example, mirror-image trade flow data do not match.

[40] See for example James Anderson and Douglas Marcouiller (2002).

[41] See for example Evans (2003a), Harrigan (1993), and Wei (1996).

endogenous. Examples are membership in currency unions or free trade agreements. Section 3.7.2 below discusses how the literature has addressed endogeneity in these specific cases.

3.5 *Estimation Bias with Traditional Gravity Equations*

The traditional gravity equation (1) omits the multilateral resistance terms of the theoretically consistent model (12). In the absence of multilateral resistance the two equations are the same, with $\beta_m=\lambda_m$. In that case $(z_{ij}^m)^{\beta m/(1-\sigma)}$ is an estimate of one plus the tax equivalent of trade barriers associated with this variable. This interpretation of the results from estimating (1) is generally incorrect because the z_{ij}^m are correlated with the multilateral resistance indices, which are themselves a function of trade barriers.

In order to better illustrate this point, consider the following simplified environment. There are two countries, the United States and Canada, with respectively N states and M provinces. The only trade barrier is a border barrier between states and provinces (ignore distance and other factors generating trade barriers). In that case $t_{ij}=b^{\delta_{ij}}$, where b is one plus the tariff equivalent of trade barriers associated with the border, and δ_{ij} is equal to zero when two regions are located in the same country and equal to one otherwise. The gravity equation is then (dropping the constant term)

$$x_{ij} = y_i + y_j + (1-\sigma)ln(b)\delta_{ij} - (1-\sigma)ln(P_i) - (1-\sigma)ln(P_j) + \varepsilon_{ij}. \quad (15)$$

If we ignore multilateral resistance, the estimate of $(1-\sigma)ln(b)$ is equal to the average within-country size adjusted trade minus the average cross-country size-adjusted trade. When there are N observations of trade within the United States (between states) and M observations of trade within Canada (between provinces), it is easy to check that the estimate of $(1-\sigma)ln(b)$ when ignoring the multilateral resistance terms is equal to $(1-\sigma)ln(b)$ plus the bias

$$(1-\sigma)\frac{M-N}{N+M}(\ln P_{US} - \ln P_{CA}) \quad (16)$$

where P_{US} and P_{CA} are the multilateral resistance indices for, respectively, U.S. states and Canadian provinces. $P_{CA}>P_{US}$ because Canada is smaller and provinces face border barriers with trade to all of the United States. The result is that for $\sigma>1$ the estimate of $(1-\sigma)ln(b)$ is biased upwards as long as $M>N$. This result is intuitive. If, for example, the only within-country trade in the sample is between provinces, the average size-adjusted within-country trade is very large. The reason is that relative trade barriers within Canada are lower than within the United States due to larger multilateral resistance, so that size-adjusted trade is larger between provinces than between states. This point was emphasized by Anderson and van Wincoop (2003).

The example above emphasizes size, but estimation bias holds more generally. Consider the role of distance. When $t_{ij}=d_{ij}^{\rho}b^{\delta_{ij}}$, the gravity equation becomes (4) with the term $(1-\sigma)\rho ln(d_{ij})$ added. There are two types of bias when attempting to estimate the border barrier b in a traditional gravity equation that omits multilateral resistance. First, the distance elasticity $(1-\sigma)\rho$ is generally incorrectly estimated since bilateral distance is correlated with the multilateral distance terms that are left in the error term. Second, even when $(1-\sigma)\rho$ is estimated correctly, we still obtain the same bias as in (16). The bias results as long as the multilateral resistance terms P_{CA} and P_{US} are different. Differences arise due to size, but also due to geography. For example, Canadian provinces are located on the North American periphery. As a result, their distances from main trading partners tend to be relatively large, so that $P_{CA}>P_{US}$. A McCallum type gravity equation with $N=0$ would then imply a positive U.S.-Canada border barrier even when none existed. If the geographic size of Canada were much smaller, so that trading distances between provinces are much smaller, P_{CA} would be smaller and the bias from estimating the border barrier b with McCallum's equation (N=0) would be smaller. Serge Coulombe (2002) emphasizes the role of these topological issues related to the special structure of the regions.[42]

3.6 *The Elasticity of Substitution*

Estimates of trade costs from trade flows are very sensitive to assumptions about the elasticity of substitution σ, so a look at the evidence is worthwhile. Although many papers have estimated this elasticity from bilateral trade flow models, only a few have used theory-based gravity.

One way to obtain an estimate of σ is to use information from directly observed trade barriers. Harrigan (1993), Hummels (2001a), Head and Ries (2001), and Baier and Bergstrand (2001) all combine estimation of theoretical gravity equations with information about tariffs and/or transport costs.

Hummels (2001a) assumes

$$t_{ij} = (f_{ij} + tar_{ij})\prod_{m=1}^{M}(z_{ij}^{m})^{\gamma_m} \qquad (17)$$

where tar_{ij} is the tariff rate and f_{ij} is a freight factor, equal to one plus the tax equivalent of freight costs. The gravity equation then becomes (dropping the constant term)

$$x_{ij} = y_i + y_i + (1-\sigma)ln(f_{ij} + tar_{ij}) \\ +\sum_{m=1}^{M}(1\text{-}\sigma)\gamma_m\, ln(z_{ij}^{m}) - (1\text{-}\sigma)\, ln(P_i) - (1\text{-}\sigma)\, ln(P_j). \qquad (18)$$

The elasticity of substitution can now be estimated through the coefficient on the log of directly observed trade costs. Hummels estimates (18) for 1992 data on sectoral imports of six countries from a large number of other countries. Multilateral resistance terms are replaced by country dummies for each sector. The estimated elasticity rises from 4.79 for one-digit SITC data to 8.26 for four-digit SITC data.

Head and Ries (2001) adopt a similar method. They consider two countries, the United States and Canada, and assume that the only trade barrier is a border-related barrier, which varies across three-digit industry data from 1990 to 1995. They estimate $(\sigma–1)ln(b_{it})$, with b_{it} equal to one plus the tariff equivalent of the border barrier in industry i at time t.[43] They then decompose the border barrier into tariff and nontariff components, $b_{it}=(1+tar_{it})(1+NTB_{it})$. It is assumed that $ln(1+NTB_{it})=K_t+\varepsilon_{it}$, where K_t is a time dummy and ε_{it} is a zero mean random disturbance. By regressing the estimate of $(1–\sigma)b_{it}$ on a time dummy and $ln(1+tar_{it})$ they obtain an estimate of both σ and average nontariff barriers across industries. This approach is similar to that in Hummels (2001a) in that it uses evidence on observed trade barriers to tie down the elasticity. Head and Ries obtain an estimate of σ of 11.4 when assuming that NTB_{it} is the same for all industries and 7.9 when allowing for industry fixed effects.

Baier and Bergstrand (2001) estimate a theoretical gravity equation where tariffs and transport costs are the only trade barriers. They use aggregate trade data for OECD countries and focus on changes in trade flows from the period 1958–60 to the period 1986–88. Their point estimate of σ is 6.4.

Harrigan (1993) implicitly models the impact on full trade costs of directly observed trade costs as $f_{ij}(1+tar_{ij})$ and finds significantly different coefficients on lnf_{ij} and $ln(1+tar_{ij})$ in the gravity regressions. His restricted regressions report most elasticity point estimates for

[42] While we have focused the discussion on U.S.-Canada, estimation bias when estimating the traditional gravity equation of course holds quite generally. Some authors have estimated both equations to allow for easy comparison. For example, Rose and van Wincoop (2001) find that the estimated trade barriers associated with the use of different currencies are much lower when estimating the theoretical gravity equation. Minondo (2001) finds that the estimate of border barriers for European trade is much lower when estimating the theoretical gravity equation.

[43] In this setup the estimate is obtained as a simple analytical function of the shares that both countries spend on their own goods.

28 sectors in the range from five to ten, with four above and one below.

An entirely different way to estimate σ is to simply estimate demand equations directly, using data on prices. But in general one estimates some combination of demand and supply relationships, the classic simultaneity problem. Feenstra (1994) is nonetheless able to obtain an estimate of the demand elasticity by using the fact that the second moments of demand and supply changes (their variances and covariance) have a linear relationship that depends on demand and supply elasticities. By assuming that supply elasticities are the same for all countries, a cross-section of the second moments allows for estimation of the elasticities. This estimation method therefore requires panel data and is applied by Feenstra to U.S. imports from 1967 to 1987 from various countries for six manufactured products. The products are highly disaggregated, finer than eight-digit SITC. The estimated elasticities range from three for typewriters to 8.4 for TV receivers.

Eaton and Kortum (2002) adopt yet another entirely different approach to obtain an estimate of σ. From (8) it follows that

$$\sigma - 1 = \frac{x_{ii} - x_{ij} - y_i + y_j}{ln(t_{ij}) - ln(t_{ii}) + ln(P_i) - ln(P_j)}. \quad (19)$$

The numerator consists of observables. The denominator is approximated as follows. Using data on retail price levels for fifty manufactured products in nineteen countries, they approximate $ln(P_j)$–$ln(P_i)$ as the arithmetic average of the log-price differentials between j and i for all 50 goods. Using the fact that log-price differentials between j and i are bounded above by $ln(t_{ij})$–$ln(t_{ii})$, they estimate $ln(t_{ij})$–$ln(t_{ii})$ as the maximum of log-price differentials between i and j. The parameter $\theta=\sigma-1$ can then be estimated as the average of the ratio on the right hand side of (19). They find $\theta=8.28$, so that $\sigma=9.28$.

Overall the literature leads us to conclude that σ is likely to be in the range of five to ten.

3.7 *The Size of International Trade Barriers*

In this section we report some results for international trade barriers, computed as the ratio of total trade barriers relative to domestic trade barriers: t_{ij}/t_{ii}. We first present summary measures of all international trade barriers and then discuss results which decompose border barriers into several likely sources. All our calculations are based on point estimates from gravity studies which differ somewhat in their list of trade cost proxies, in the data used, and in the assumptions used to construct standard errors.

3.7.1 *Summary Measures*

Part of the empirical gravity literature reports the tax equivalent of summary measures of trade barriers, those associated with distance and the presence of borders. Table 7 presents the results of a number of studies. The table indicates whether results are based on the traditional gravity equation (1) or the theory-based gravity equation (12). It also indicates whether numbers are based on aggregate trade data or disaggregate (sectoral) trade data. In the latter case table 7 reports the average trade barrier across sectors. Column four reports the tax equivalent of trade barriers reported by the authors, with the corresponding elasticity σ in brackets. In order to make results more comparable across papers, the final three columns re-compute the trade barriers for elasticities of five, eight, and ten. These are representative of the elasticities estimated in the literature. In some cases two numbers are shown, with the lower number applying to countries that share the same language and border.

The first three rows report results for total international trade barriers. The results are sensitive to the elasticity of substitution. For example, the U.S.-Canada trade barrier based on the Head and Ries (2001) study ranges from 35 percent for $\sigma=10$ to 97 percent for $\sigma=5$. From now on we will focus the discussion on the intermediate value of $\sigma=8$. In that case the findings by Head and Ries

TABLE 7
TARIFF EQUIVALENT OF TRADE COSTS

	method	data	reported by authors	$\sigma=5$	$\sigma=8$	$\sigma=10$
all trade barriers						
Head and Ries (2001) U.S.-Canada, 1990-1995	new	disaggr.	48 ($\sigma=7.9$)	97	47	35
Anderson and van Wincoop (2003) U.S.-Canada, 1993	new	aggr		91	46	35
Eaton and Kortum (2002) 19 OECD countries, 1990 750-1500 miles apart	new	aggr.	48–63 ($\sigma=9.28$)	123–174	58–78	43–57
national border barriers						
Wei (1996) 19 OECD countries, 1982-1994	trad.	aggr.	5 ($\sigma=20$)	26–76	14–38	11–29
Evans (2003a) 8 OECD countries, 1990	trad.	disaggr.	45 ($\sigma=5$)	45	30	23
Anderson and van Wincoop (2003) U.S.-Canada, 1993	new	aggr.	48 ($\sigma=5$)	48	26	19
Eaton and Kortum (2002) 19 OECD countries, 1990	new	aggr.	32–45 ($\sigma=9.28$)	77–116	39–55	29–41
language barrier						
Eaton and Kortum (2002) 19 OECD countries, 1990	new	aggr.	6 ($\sigma=9.28$)	12	7	5
Hummels (1999) 160 countries, 1994	new	disaggr.	11 ($\sigma=6.3$)	12	8	6
currency barrier						
Rose and van Wincoop (2001) 143 countries, 1980 and 1990	new	aggr.	26 ($\sigma=5$)	26	14	11

Notes: This table reports findings in the gravity literature on the tariff equivalent of a variety of factors that increase trade barriers. The second column indicates whether estimates are based on the traditional gravity equation —"trad."— or the theory-based gravity equation —"new". The third column indicates whether estimation is based on aggregate or disaggregate data. The numbers in the fourth column have been reported by the authors for various elasticities of substitution s that are shown in brackets. For results based on disaggregated trade data, the average trade barrier across sectors is reported (for Hummels (1999) only sectors with statistically significant estimates are used). The numbers in the last three columns re-compute these results for alternative values of σ. For results based on disaggregate data, the trade barriers are first re-computed for each sector and then averaged (with the exception of Head and Ries (2001), who only report average trade barriers across all sectors). When two numbers are reported, the lower number applies to countries that share a border and have a common language.

(2001) imply an average U.S.-Canada trade barrier of 47 percent based on average results from 1990 to 1995. Based on Anderson and van Wincoop's (2003) results, their estimated trade cost parameters with σ=8 imply a 46-percent U.S.-Canada trade barrier for 1993, virtually the same as Head and Ries (2001). This is calculated as the trade-weighted average barrier for trade between states and provinces, divided by the trade-weighted average barrier for trade within the United States and Canada. Eaton and Kortum (2002) report results for nineteen OECD countries for 1990. For countries that are 750–1500 miles apart the trade barrier is 58–78 percent. The lower 58 percent number, which applies to countries that share a border and language, is not much larger than the estimates of the U.S.-Canada barrier. Summarizing, international trade barriers are in the range of 40–80 percent for a representative elasticity estimate.

Edward Balistreri and Russell Hillberry (2002) argue that trade barrier estimates from these types of studies are implausibly large. Based on parameter estimates in Anderson and van Wincoop (2003), they report on trade barriers for Alberta and Alabama relative to the barrier of trade within Maryland, which has the lowest trade barrier. As an illustration, they report a trade barrier between Alabama and Quebec of 1684 percent for σ=3 and 322 percent for σ=5. These numbers are much larger than the 46 percent U.S.-Canada trade barrier reported above. There are several reasons for this discrepancy. First, elasticities of three and five are low relative to estimates in the literature reported above. If we set σ=8, the Alabama-Quebec barrier drops to 133 percent relative to the Maryland barrier. Second, Balistreri and Hillberry report trade barriers relative to Maryland instead of relative to average domestic trade barriers as in the papers discussed above. They therefore also capture trade barriers within the United States, which are known to be very large based on the direct evidence on distribution costs reported in section 2. The trade barrier for Alabama-Quebec trade relative to within-U.S. trade is 61 percent. The latter is still a bit above the 46-percent U.S.-Canada barrier reported above, since Alabama and Quebec are farther apart than the average state-province pair.

International trade barriers can be decomposed into barriers associated with national borders and barriers associated with geographic frictions such as distance. The next set of rows in table 7 report the magnitude of international trade barriers associated with borders. For σ=8, the estimated U.S.-Canada border barrier from Anderson and van Wincoop (2003) is 26 percent. With a total barrier of 46 percent, it implies a distance related barrier of 16 percent (1.16=1.46/1.26). This is not much different than direct estimates of trade costs reported in section II: the total transport cost estimate for the United States is 21 percent (1.21=1.107*1.09, with 10.7-percent freight cost and 9-percent time cost). In our representative total trade cost calculation, we assume all effects of distance are captured in the transport cost number (Hummels reports partial success in testing this hypothesis with his disaggregated data.)

Although based on traditional gravity equations, estimates of border barriers in Wei (1996) and Evans (2003a) are not too far off, respectively 14–38 percent and 30 percent. Both apply to OECD countries. The results from Eaton and Kortum (2002) are a bit higher, 39–55 percent. Eaton and Kortum do not actually report barriers associated with national borders. The 39–55 percent barrier applies to trade costs of countries 0–375 miles apart relative to domestic trade costs. Since even for those countries the average distance is larger than within countries, it overstates a bit the barrier associated with borders. But our own calculations based on estimates from Anderson and van Wincoop (2003) indicate that this overstatement is only about 6 percent. Summarizing, barriers associated with national borders are in the range of 25–50 percent for a representative elasticity estimate.[44]

All the trade costs estimates reported above are based on point estimates. Standard errors tend to be quite small, though, as the empirical fit of gravity equations is generally very strong. For example, a 95-percent confidence

[44] There is also a small literature that has documented border barriers at the level of states within the United States. See Wolf (2000a,b), Hillberry and Hummels (2002a), and Daniel Millimet and Thomas Osang (2001).

interval of U.S.-Canada border barriers based on Anderson and van Wincoop (2003) is 23 percent–28 percent. Standard errors of trade costs associated with distance are even smaller.

3.7.2 *Decomposition of Border Barriers*

It is of great interest to better understand what blocks build these border barriers. We saw in section 2 that policy barriers, in the form of tariffs and NTBs, amount to no more than about 8 percent for OECD countries,[45] suggesting important additional barriers associated with national borders. Estimates of five different types of trade frictions are associated with national borders.

Language Barriers. Table 7 reports estimates from Eaton and Kortum (2002) and Hummels (2001a) on language-related barriers. They both follow the common approach of introducing a dummy variable that is one if two countries have a common language and zero otherwise. Results from both papers imply a tax equivalent associated with speaking different languages of about 7 percent when $\sigma=8$.

Currency Barriers. The use of different currencies may pose a barrier to trade as well. Rose and van Wincoop (2001) introduce a dummy variable that is equal to one when two countries use the same currency (are part of a currency union) and equal to zero otherwise. Based on data for 143 countries, the estimated tariff equivalent associated with using different currencies is 14 percent when $\sigma=8$.

While Rose and van Wincoop (2001) is the only paper that computes the tariff equivalent of the reduction in trade barriers from joining a currency union, by now a large body of papers has documented a big positive impact of currency unions on trade. Rose (2000) estimates a traditional gravity equation using data for 186 countries from 1970 to 1990 and finds that countries in a currency union trade three times as much. The finding also applies to historical data. Using data from the nineteenth and early twentieth centuries, Antoni Estevadeordal, Brian Frantz, and Alan Taylor (2003) and Ernesto Lopez-Cordova and Christopher Meissner (2003) document a big impact on trade of belonging to the same commodity regime, such as the gold standard. Rose (2004) considers the evidence from nineteen studies on the effect of currency unions on trade and concludes that the combined evidence from all studies suggests a doubling of trade if countries belong to a currency union.

One problem with this evidence is that it is unclear why a currency union raises trade levels so much. There are small costs involved in exchanging currencies. There is substantial consensus that the impact of exchange rate volatility on trade is very small at best, with even the sign uncertain; hence the volatility-suppressing effect of currency unions should not raise trade much.

Another potential problem is endogeneity. Countries may join a currency union because they have close trade relationships rather than the other way around. The endogeneity problem is well-recognized in the literature. Anecdotal evidence on this issue is mixed. Rose (2000) discusses anecdotal evidence suggesting that trade does not play an important role in the decision to join a currency union, but Estevadeordal et.al. (2003) discuss an example where it does appear to have mattered. A second approach is to use instruments. Rose (2000) uses various instruments associated with inflation and continues to find that currency unions have a significant effect on trade. Alberto Alesina, Robert Barro, and Silvana Tenreyro (2002) and Barro and Tenreyro (2002) also adopt an IV approach. Using a probit analysis, they first compute the probability that a country adopts each of four potential anchor currencies.[46] For any two

[45] Recall that Messerlin (2001) reports a 7.7 percent arithmetic average protection rate in industrial goods for the European Union.

[46] The probability depends on exogenous factors, such as various size-related measures, distance from the anchor currency country, and various dummy variables such as common language with the anchor currency country.

countries i and j, the probability of belonging to the same currency union is then $\sum_{c=1}^{4} p_i^c p_j^c$, where p_i^c is the probability that country i adopts anchor currency c. This is used as an instrument for the currency union dummy. They find that the impact of currency unions on trade remains large and significant. Christian Broda and John Romalis (2003) also adopt an instrumental variables technique, but instead find that endogeneity is important. They argue that commodities trade should not be affected by currency unions. If this is the case, commodities trade can be used as an instrument in a regression of a currency union dummy on total trade. The resulting error term of this regression is used as an instrument for currency unions in a regression of manufacturing trade on a currency union dummy. They find that the coefficient on this currency union instrument is small and insignificant.

Another approach to deal with endogeneity uses time series evidence. If a country joins a currency union because of close trade ties, then trade is already high when the country joins the currency union and does not rise subsequent to that. Reuven Glick and Andrew Rose (2002) examine data for 200 countries from 1948 to 1997 and find that countries joining or exiting a currency union during this period experienced respectively a near-doubling or halving of bilateral trade. In contrast, Albrecht Ritschl and Nikolaus Wolf (2003) find evidence that endogeneity is important. After the Great Depression, most countries went off the gold standard and new regional currency blocs were formed. Ritschl and Wolf find that trade among members of these new currency blocs is two to three times as high as trade among countries that do not belong to the same currency bloc. But they show that trade among countries of these future currency areas was already high in the 1920s, a decade before the new currency blocs were formed, and that the actual formation of the currency areas did not raise bilateral trade among their members. However, their currency blocs have fixed exchange rate systems rather than a single currency, so that it is not evidence about currency unions.

Information Barriers. Border costs associated with information barriers are potentially important. This is documented by several authors. Richard Portes and Helene Rey (2002) first estimate a traditional gravity equation (1), regressing bilateral trade on GDPs, per capita GDPs, and distance. Then they add two information variables: a size-adjusted volume of telephone traffic and the size-adjusted number of branches of the importing country's banks located in the exporter's country. Both coefficients are positive and highly significant. Moreover, the coefficient on bilateral distance is reduced from -0.55 to -0.23.

James Rauch and Vitor Trindade (2002) have conducted the most careful work so far. They argue that information barriers to trade may be reduced when two countries have a substantial Chinese network, proxied as the product of the population percentages of Chinese ethnicity in the two countries. They estimate a traditional gravity equation for both "reference price" goods and differentiated goods which do not have reference prices.[47] The trade-increasing effect of the Chinese network variable is larger for the differentiated goods than for the reference price goods. The difference between them may be taken to represent the effects of information transfer using the network, with the trade increasing effect of networks on the reference price goods representing the value of the network to informal contract enforcement. Pushing inference based on their results to the limit, we calculate the information-cost-reducing value of strong Chinese immigrant links (where both partners have a larger than 1 percent Chinese population) to be worth as

[47] For the reference price goods they also distinguish between goods traded on organized exchanges and those not so traded.

much as a 47-percent increase in trade.[48] Earlier work done on the effect of Chinese immigrant links on the bilateral trade of the United States by Gould (1994) found larger effects, while Head and Ries (1998) found much more modest effects in the bilateral trade of Canada.

With an elasticity of substitution of eight, these Chinese networks save an information cost worth 6 percent. This may be both a lower bound and an upper bound to the tax equivalent of information barriers. It is a lower bound in that information barriers are likely to be important even if two countries have populations with the same ethnic background. It is an upper bound to the extent that one believes that other factors are picked up. For example, the results are largely driven by countries with very large Chinese populations, such as Taiwan, Hong Kong, China, Singapore, and Malaysia. Since Rauch and Trindade estimate a traditional gravity equation, they may not properly control for the large distance of these countries from the United States and Europe. It is also possible that strong historical trade ties drive the results, as would be the case in trade models where history matters. Evidence by Evans (2003a) also suggest that the 6-percent information barrier is an overstatement. She estimates a traditional gravity equation for OECD bilateral trade flow data for twelve industries, with trade dependent on GDPs, distance, remoteness, and a border dummy. She finds that the coefficient on the border dummy does not drop once a variety of industry-specific variables related to the importance and difficulty of information transfers are included (e.g. the frequency of technical service). More careful modeling of the underlying information costs in future work will probably be illuminating.

Contracting Costs and Insecurity. The costs of writing contracts and enforcing them or self-insuring the costs of default on unenforced contracts are all plausibly larger across borders. [49] Evans (2001) provides evidence that internal contracting costs within a firm are much lower than external contracting costs. Specifically, the tax equivalent of the trading costs of a foreign affiliate of a U.S. multinational with unaffiliated U.S. firms is on average 37 percent higher (for a demand elasticity of five) than the trading costs with its U.S. parent. She provides evidence that proprietary assets associated with transactions (e.g. the sale of a brand with a good reputation or a good that is technologically advanced) play an important role in this result. Since the goals of unaffiliated firms are different, there are greater risks involved in selling goods that involve substantial proprietary assets, which would involve larger contracting costs.

Rauch and Trindade (2002) can be interpreted to provide an inference about contracting and enforcement cost. Networks provide a kind of enforcement through sanctions which substitutes for weak international enforcement of formal contracts. Their finding of an 89 percent trade increasing effect of networks for reference price goods is hard to interpret as information costs and may be a result of contracting costs. But since these goods presumably have high elasticities of substitution, the tax equivalent is likely small (3 percent when σ=20).

[48] Table 9 of Rauch and Trindade (2002) reports the trade volume effect of changing the Chinese network variable from zero to its sample mean value for two subsets of trading partners, those with Chinese population shares greater than 1 percent and those with Chinese population shares less than 1 percent. For the smaller share group, the volume effects are modest—6.2 percent for the differentiated products and somewhat smaller for the reference priced products. For the subset of countries with larger Chinese population shares, the effect of switching on the sample mean of the Chinese network variable is a 178 percent rise in trade for differentiated goods, a 128 percent rise in trade in reference priced goods, and a 89 percent rise in trade in goods on organized exchanges. We attribute the difference between the impact on differentiated goods and goods traded in organized exchanges as a result of information costs: 47 percent=100*(2.78/1.89).

[49] See Anderson (2000) for a discussion of the literature on this topic.

Anderson and Marcouiller (2002) show that insecurity, associated both with contractual enforcement problems and with corruption, lowers trade substantially. Using survey data taken from businessmen by the World Economic Forum as an index of institutional quality and making institutional quality (both contractual enforcement and corruption) an argument of the trade cost function, they implement a variant of the theoretically consistent gravity model.[50] They report the effect of raising the quality of institutions from the Latin American average (for the seven Latin American countries in the sample—Argentina, Brazil, Chile, Colombia, Mexico, Peru, and Venezuela) to the E.U. average. Combined with their maximum likelihood point estimate of the elasticity of substitution equal to eight, the implied tax equivalent of relatively low quality institutions is 16 percent. Insecurity is therefore an important trade barrier for developing countries. It plays less of a role for industrialized countries on which trade barrier estimates in table 7 are based. An experiment based on their estimates that raises the U.S. security level to that of Singapore, the highest in the data, implies a tax equivalent of 3 percent for an elasticity of eight.

Nontariff Policy Barriers. Harrigan (1993), Head and Mayer (2000), and Chen (2002) consider the role of nontariff barriers in accounting for the impact of borders on trade levels. The last two papers do not find evidence of a positive relationship between sectoral estimates of the impact of borders on trade and sectoral data for nontariff barriers. Harrigan (1993) estimates the traditional gravity equation (1) for 1983 bilateral trade data for 28 industries among OECD countries and also finds that nontariff barriers have little effect on trade, with coefficients sometimes having the wrong sign. These insignificant results may be explained by failure to control for the endogeneity of NTBs suggested by political economy. Daniel Trefler (1993) shows that the effect of NTBs on U.S. trade with the rest of the world is increased in absolute value by an order of magnitude when controlling for the endogeneity of NTBs with instruments commonly used in the political economy literature. In contrast, in a related approach which uses a set of both rich and poor countries, Jong-Wha Lee and Phillip Swagel (1997) jointly estimate an equation relating sectoral imports to trade barriers and an equation relating sectoral NTBs to various driving factors from political economy. Once appropriately controlling for industry and country dummies in the trade regression, they find no evidence that NTBs affect trade flows. The difference between their results and Trefler's may partly reflect differences across countries which are not controlled for, and partly the richer set of political economy variables which Trefler is able to deploy for the United States alone.

There is extensive evidence that free trade agreements and customs unions increase trade and therefore reduce trade barriers (e.g. Jeffrey Frankel, Ernesto Stein, and Shang-Jin Wei 1998; Frankel and Wei 1997), but it is less clear what elements of these trade agreements play a role (tariffs, NTBs, or regulatory issues). As noted above, all gravity model analyses of NTBs and regional trade agreements impose very strong and presumptively implausible regularity restrictions on the effect of nontariff barriers and customs-union membership upon trade volume. Moreover, most of this evidence is cross-section evidence, which raises causality issues. Baier and Bergstrand (2003) control for endogenous regional trade agreements and greatly increase the implied magnitude of the membership effect. Time series evidence, in contrast, suggests that free trade agreements are less important than in cross-section findings. Although Helliwell (1998,

[50] They divide imports of j from i by U.S. imports from i, which cancels the exporter price index. The importer relative price index is approximated with a Törnqvist index.

1999) estimates a substantial drop in the border effect for U.S.-Canada trade since the 1988 free trade agreement, Coulombe (2002) shows that the same downward trend took place already before 1988 and applies to the entire sample 1981–2000. Similarly, Head and Mayer (2000) document a gradual drop in European border barriers from 1976 to 1995, with no significant drop after the implementation of the Single European Act of 1986.

Summary. Assuming an 8-percent policy related barrier (based on direct evidence from tariffs and NTBs), a 7-percent language barrier, 14-percent currency barrier, a 6-percent information cost barrier, and a 3-percent security barrier, overall border barriers are 44 percent. This falls within the estimated 25–50 percent range reported for OECD countries in table 7. The discussion above qualifies these magnitudes as quite rough.

3.8 *Aggregation Issues*

The gravity equations (5)–(7) are the basis for discussion of aggregation issues:

$$X_{ij}^k = \frac{E_j^k Y_i^k}{Y^k}\left(\frac{t_{ij}^k}{P_j^k \Pi_i^k}\right)^{1-\sigma_k}$$

$$(\Pi_i^k)^{1-\sigma_k} = \sum_j \left(\frac{t_{ij}^k}{P_j^k}\right)^{1-\sigma_k} \frac{E_j^k}{Y^k}$$

$$(P_i^k)^{1-\sigma_k} = \sum_i \left(\frac{t_{ij}^k}{\Pi_j^k}\right)^{1-\sigma_k} \frac{Y_i^k}{Y^k}.$$

There are two different aggregation issues. First, estimated trade costs at a disaggregate level overwhelm the comprehension of the analyst. Like others, we report atheoretic averages in this survey, but an ideal summary index is desirable. Second, in practice gravity equations are estimated from aggregate data; even sectoral data are not at the level of detail of reality. Aggregation bias results from estimating trade costs with aggregate data when trade costs vary at the disaggregate level.

3.8.1 *Ideal Aggregation*

Anderson and Neary (2003) develop an ideal summary index of trade costs, defined as the uniform trade cost that leads to the same aggregate trade level. The ideal index idea can be applied to aggregation in several dimensions, each providing answers to sensible questions. In the context of gravity models, it is natural to consider ideal aggregation over trading partners (e.g., aggregate t_{ij}^k over importers j for an export trade cost index) and aggregation over commodities (e.g., aggregate t_{ij}^k over k for a given link from i to j).

Summary Trade Costs for Each Region. The multilateral indices $\{P_j^k, \Pi_j^k\}$ are elegantly simple summary measures of trade costs for a particular region j with all its trading partners (including the region itself). P_j^k is an average import trade cost (including imports from oneself) and Π_j^k an average export trade cost (including exports to oneself). To see this, note that if, hypothetically, the actual bilateral cost set $\{t_{ij}^k\}$ is replaced with $\{P_j^k \Pi_i^k\}$ the equilibrium price indices themselves remain the same and aggregate trade flows $\sum_i X_{ij}^k$ and $\sum_j X_{ij}^k$ also remain unchanged.

To illustrate, table 8 reports multilateral resistance for U.S. states and Canadian provinces based on Anderson and van Wincoop (2003), assuming σ=8. In their model $P_i = \Pi_i$, so the import and export trade cost measures are the same. We use the approach of Balistreri and Hillberry (2002) by normalizing trade costs within Maryland as zero and therefore dividing all trade barriers by that of trade within Maryland. Note that Canadian provinces have systematically higher trade costs because they face the border cost on much more of their trade with the U.S. states and because they are more distant from more important sources of supply. Note also that populous northeastern U.S. states and California face the lowest trade costs, explained by the geographic advantage of

TABLE 8
TRADE COST INDICES FOR U.S. AND CANADIAN REGIONS

	Multilateral resistance
Provinces	
Alberta	1.65
British Columbia	1.55
Manitoba	1.65
New Brunswick	1.62
Newfoundland	1.74
Nova Scotia	1.63
Ontario	1.51
Prince Edward Island	1.67
Quebec	1.53
Saskatchewan	1.66
States	
Alabama	1.39
Arizona	1.49
California	1.31
Florida	1.38
Georgia	1.37
Idaho	1.49
Illinois	1.35
Indiana	1.36
Kentucky	1.37
Louisiana	1.41
Maine	1.43
Maryland	1.31
Massachusetts	1.33
Michigan	1.37
Minnesota	1.42
Missouri	1.38
Montana	1.50
New Hampshire	1.39
New Jersey	1.33
New York	1.30
North Carolina	1.37
North Dakota	1.47
Ohio	1.35
Pennsylvania	1.33
Tennessee	1.38
Texas	1.41
Vermont	1.40
Virginia	1.35
Washington	1.43
Wisconsin	1.39
Rest of USA	1.50

Notes: This table reports multilateral resistance indices for U.S. states and Canadian provinces, based on data and trade cost estimates from Anderson and van Wincoop (2003). The assumed elasticity of substitution is 8.

being close to a greater share of the output produced.[51]

Aggregating Across Trading Partners. Since bilateral trade barriers vary across countries, a summary measure of international trade barriers is useful. A natural aggregator for inward or outward shipments replaces all t_{ij} where $i \neq j$ with a uniform inward or outward international trade barrier that leaves aggregate international trade unchanged.

To illustrate this, consider a simple example. There are N countries, with N an odd integer. Each produce a fraction $1/N$ of the output of industry k. The countries are evenly spaced on a circle, with trade barriers between them proportional to their shortest distance on the circle. There are no trade barriers within countries. These simplifying assumptions imply that inward and outward summary barriers are the same and common across countries, and that $P_i = \Pi_i = P$. If b_i is the trade barrier for two countries that are i steps away from each other on the circle, it is easily checked that the uniform international trade cost index b that leads to the same level of international trade can be solved from (dropping industry subscripts)

$$b^{1-\sigma} = \frac{\sum_{i=1}^{0.5N-0.5} b_i^{1-\sigma}}{0.5N - 0.5}. \qquad (20)$$

Since $b^{1-\sigma}$ is convex in b, it follows that b is less than the average trade barrier. More (omitted) algebra, based on the substitution effect, shows that b is larger than the trade-weighted average barrier. The ideal index therefore lies in between the arithmetic average and trade-weighted average barrier in this example.

Aggregating Across Goods. Aggregation across goods is useful for summarizing detailed data. To illustrate, again consider a simple example. Assume that there are two equally sized countries ($Y_1 = Y_2$) and K sectors. Each country spends a fraction θ_k on sector k, $E_i^k = \theta_k Y_i$, and produces half of the output in each sector. The only trade barrier is a border barrier between the two countries that varies across industries: $t_{12}^k = b_k$. This example again has the advantage of closed form solutions for the multilateral resistance indices. It is easy to check that for each sector $P^k = \Pi^k$. When raised to be power $1-\sigma_k$ they are equal to $(0.5+0.5b_k^{1-\sigma_k})^{0.5}$. In our example the uniform barrier b that leads to the same aggregate trade between the two countries can be solved from

$$\sum_{k=1}^{K} \theta_k \frac{(b_k)^{1+\sigma_k}}{1-(b_k)^{1-\sigma_k}} = \sum_{k=1}^{K} \theta_k \frac{b^{1+\sigma_k}}{1-b^{1-\sigma_k}}. \qquad (21)$$

Starting from a benchmark where all barriers are equal to $\bar{b}$ and all elasticities are equal to $\bar{\sigma}$, we now introduce variation in the barriers and elasticities such that the industry size weighted averages

$$\sum_{k=1}^{N} \theta_k b_k \text{ and } \sum_{k=1}^{N} \theta_k \sigma_k$$

remain constant at respectively $\bar{b}$ and $\bar{\sigma}$. The function $B(b,\sigma) \equiv b^{1-\sigma}/(1+b^{1-\sigma})$ is decreasing and convex in b. Then Jensen's Inequality implies that for a mean-preserving spread in $\{b_k\}$, $b < \bar{b}$. Since the absolute effect on aggregate trade is larger when reducing a trade barrier than raising a trade barrier, the ideal index gives relatively more weight to low sectoral trade barriers. Variation in σ given a uniform b has no effect, from (21).

Simultaneous variation of $\{b_k, \sigma_k\}$ requires us to consider covariation. Consider marginal variation holding $\lambda = var(\sigma)/var(b)$ constant. We can use a second order Taylor expansion of (21) to show the following:

$$\partial b/\partial(var(b)) = -\alpha_1 + \alpha_2 \lambda corr(b,\sigma) \qquad (22)$$

where α_1 and α_2 are two positive constants. If the elasticity is low exactly in sectors with high barriers, the impact on aggregate trade of these high barriers is reduced, leading to

[51] A more subtle pattern may be discerned. Location near a border should, all else equal, raise multilateral resistance since the border cannot matter as much for a state located far from the boder where distance will naturally impede trade. (This is observable only in states, since all Canadian provinces are on the border.)

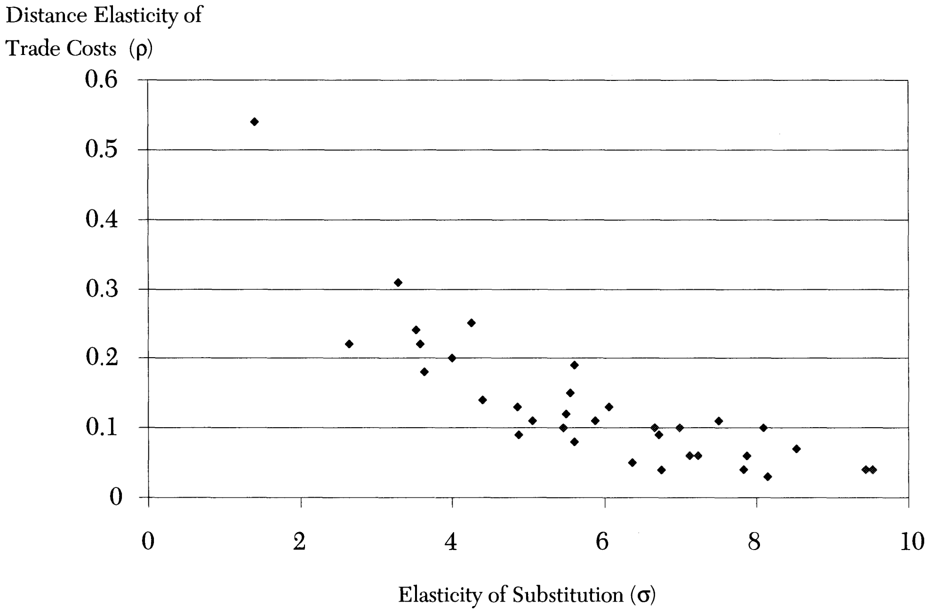


Figure 1. Elasticity of Substitution versus Distance Elasticity of Trade Costs

a smaller uniform barrier b. There is evidence that a negative correlation is realistic, which reinforces the conclusion that the uniform b is lower than $\bar{b}$.[52]

As an illustration, figure 1 graphs trade barriers against price elasticities for findings based on table 4 in Hummels (2001a). Hummels assumes $t_{ij}^k = d_{ij}^{\rho_k} m_{ij}^k$, where d_{ij} is distance and m_{ij}^k stands for trade barriers unrelated to distance. An increase in ρ_k implies higher distance related trade barriers. Figure 1 shows that sectors with a high distance elasticity ρ_k of trade costs tend to have low elasticities of substitution, which is consistent with $corr(b,\sigma)<0$. Similar conclusions can be drawn for language and adjacency as factors driving trade costs in Hummels (2001a). Evans (2003a) provides similar evidence for border related trade barriers. Using 1990 OECD country data for twelve industries, she estimates relatively high values of b^σ, with b the tariff equivalent of border barriers, for industries with a high degree of product differentiation (low elasticity). This can only be the case if industries with low elasticities tend to have relatively high border barriers.[53]

Finally, there is also reason to believe that the industry size weighted average $\bar{b} = \sum_{k=1}^{N} \theta_k b_k$ is lower than a simple average trade barrier. Head and Ries (2001) point out that U.S.-Canada trade barriers tend to be low in relatively large sectors such as motor vehicles. This will further reduce the uniform barrier below a simple average barrier across industries. Hillberry (2002) shows that for U.S.-Canada trade, a simple average border-related barrier across

[52] Simulation with discrete changes confirms the result from (22): negative correlation increases $\bar{b}-b$, positive correlation can change its sign, and the magnitude of $\bar{b}-b$ can be substantial.

[53] Chen (2002) finds that for European trade there is no relationship between b^σ and the Rauch and Trindade (2002) measure of product differentiation, again pointing to a negative relationship between b and σ across industries.

industries is twice as large as an output weighted average.

The example shows that there are several reasons to believe that an ideal index of trade barriers that aggregates over sectors will be lower than a simple average of sectoral trade barriers. This suggests that arithmetic average trade barriers reported in section 2, as well as the numbers in table 7 based on studies from disaggregate data, overstate an ideal index of trade barriers. How much they overstate the ideal index is unknown and is therefore an obvious area for future research.

3.8.2 *Estimation Bias and Aggregation*

Estimation based on aggregate data pretends that there is only one border barrier b and one elasticity σ while international trade barriers vary across goods. Consider aggregate estimation with multilateral resistance indices replaced by country dummies. The aggregate gravity equation can be written as

$$x_{ij}=\alpha_i+\alpha_j+(\sigma-1)ln(b)\delta_{ij}+\varepsilon_{ij} \qquad (23)$$

where α_i and α_j are country-specific constants that depend on the multilateral resistance variables, and δ_{ij} is one if i=j and zero otherwise.

First consider again the example above where the border barriers vary across country pairs. With N countries, the OLS estimate of $ln(b^{1-\sigma})$ is

$$ln(\hat{b}^{1-\sigma})=\frac{1}{N^2-N}\sum_{i\neq j}x_{ij}-\frac{1}{N}\sum_{i=1}^{N}x_{ii}$$

$$=\frac{\sum_{i=1}^{0.5N-0.5}ln(b_i^{1-\sigma})}{0.5N-0.5}. \qquad (24)$$

Together with (20) and the concavity of the log-transformation, it follows that the estimated border barrier is larger than the ideal index if we know the elasticity σ.

Next consider the two-country example with varying border barriers across industries. In this case the OLS estimate of the border barrier is:

$$(\hat{b})^{1-\sigma}=\frac{X_{12}}{X_{11}^{0.5}X_{22}^{0.5}}.$$

Substitution of the theoretical expressions for X_{11}, X_{22} and X_{12} yields

$$\frac{(\hat{b})^{1-\sigma}}{1+(\hat{b})^{1-\sigma}}=\sum_{k=1}^{K}\theta_k\frac{b_k^{1-\sigma_k}}{1+b_k^{1-\sigma_k}}. \qquad (25)$$

Together with (21) it follows that the estimate $\hat{b}$ based on the aggregate gravity equation is equal to the ideal index b when all price elasticities are identical and we set σ at that level. Although the assumption of identical elasticities across sectors is not realistic, this result nonetheless provides important guidance. There is confusion in the literature about whether one should use elasticities based on aggregate data or disaggregate data when interpreting estimation results based on aggregate data. It makes a big difference because price elasticities based on aggregate data are much smaller. The example shows that the elasticity of substitution at the more aggregate level, between sectors, is entirely irrelevant. One should choose elasticities at a sufficiently disaggregated level at which firms truly compete.

If the price elasticities differ across sectors, then we need to choose σ below the average across sectors in order for the estimate $\hat{b}$ to be equal to the uniform index b. If we choose σ to be equal to the average elasticity across sectors, then the estimate $\hat{b}$ would be lower than the uniform barrier b. The magnitude of this bias is unknown.

In the examples above the production and spending structure—the set $\{Y_i^k, E_i^k\}$—is unrelated to trade barriers. In general though the production and spending structure is endogenous and depends on trade costs. This can generate another important source of aggregation bias, as discussed in Hillberry (2002) and Hillberry and Hummels (2002b). They consider models with an endogenous production structure, whereby either (i) trade barriers vary across both industries and countries while the demand structure is the same across countries, or (ii) trade barriers only vary across countries while the (intermediate) demand structure varies across countries (due to varying gross output mix).[54]

Hillberry (2002) considers case (i) in the context of a monopolistic competition model with endogenous entry of firms. Countries will tend to specialize in sectors in which they have a comparative advantage reflected in trade costs. If for example trade costs take the form $t_{ij}^k = \tau_k + m_{ij}$, then a country i that faces relatively low trade barriers m_{ij} has comparative advantage in industries with relatively low industry-specific barriers τ_k.[55] This raises the level of trade among countries with low bilateral trade barriers. Trade becomes more sensitive to trade barriers: trade barriers reduce trade both through a standard substitution effect and because firms endogenously locate close to markets with which they have low trade barriers. Estimation based on aggregate data attempts to attribute the reduction in trade as a result of trade costs solely to the standard substitution effect. Estimated trade barriers will therefore be too high.

Hillberry (2002) considers another example applied to U.S. states and Canadian provinces. He assumes that trade costs take the form $t_{ij}^k = m_{ij} e^{\delta_{ij}\tau_k}$, where m_{ij} is a barrier related to distance and δ_{ij} is 0 if regions are located in the same country and 1 otherwise. τ_k is an industry-specific border barrier. This setup implies that regions located close to the border, which tend to trade a lot with the other country, have comparative advantage in the sectors with relatively low border costs. Since all state-province pairs get equal weight in gravity equation estimation, Hillberry argues that estimates from aggregate gravity equations overstate the impact of border barriers on international trade. He provides some evidence to confirm this, but because it is based on the traditional gravity equation, omitted variable bias may explain the finding.

Hillberry and Hummels (2002) consider case (ii) in which trade barriers are the same across industries but the demand structure varies across countries because firms use different bundles of intermediate goods in production. In that case countries tend to specialize in industries in which demand is relatively high from trading partners with whom they have relatively low trade barriers. This again raises the impact of trade barriers on aggregate trade flows, leading to an upward bias of trade cost estimates based on aggregate data. For trade within and between U.S. states, Hillberry and Hummels (2002b) regress the product of sectoral demand and supply, $E_j^k Y_i^k$, on distance, adjacency, and a within-state dummy. They find that the co-location of production and demand is mainly important at the very local level, within states or between adjacent states. They do not report results on the extent to which trade barrier estimates based on aggregate data overstate the ideal index.

Papers by Kei-Mu Yi (2003a,b) are also relevant in this context. These papers are based on evidence of an important role of vertical specialization in international trade, as documented in Hummels, Jun Ishii, and Yi (2001). Vertical specialization is defined as the use of imported intermediate goods in the production of goods that are exported. Yi (2003a) develops a two-country model (H and F) in which consumer goods are produced in three stages. The first stage produces an intermediate good, which is used in the second stage to produce another intermediate good. The final stage combines all intermediate goods from stage two to

[54] If neither (i) nor (ii) are satisfied, there is no aggregation bias. An example of this is the model by Eaton and Kortum (2002). All firms use the same CES bundle of all intermediate goods as inputs and all consumers have the same preferences. The demand structure is therefore the same across countries. Moreover, trade barriers are the same for all industries. As discussed previously, their model can be shown to lead to the standard aggregate gravity equation (8).

[55] If, on the other hand, trade barriers take a multiplicative form $t_{ij}^k = \tau_k m_{ij}$, the relative trade barrier in two different sectors is the same for all i, j and all countries will have the same production structure.

produce a nontraded final consumption good. These stages can take place in different locations, leading to "back and forth" trade. Yi (2003a) shows that trade barriers now have a magnified effect on world trade, which helps account for the growth in world trade over the past four decades. Yi (2003b) argues that taking into account this magnification factor should reduce estimates of trade barriers.

Two factors magnify the impact of trade barriers on trade. First, there are fewer imports of vertically specialized goods by the final goods sector. There is vertical specialization when for example stage 1 is produced in H, which is shipped to F for stage-2 production, which is shipped to H for final goods production. Stage 1 goods will then have crossed borders twice, magnifying the impact of border barriers.[56] Second, a rise in trade barriers affects the production structure in that it becomes more likely that the first two stages are produced in the same country (thus eliminating vertical specialization). If stage 1 is produced in H, higher trade barriers make it more attractive to produce stage 2 in H as well. Both magnification factors are again associated with aggregation bias. Standard gravity equations still hold for stage-2 sub-sectors that use the same production function (source stage-1 products from the same country).

While aggregation bias can theoretically arise in many ways, little is known about the empirical magnitude of aggregation bias. One obvious recommendation is to disaggregate. There is no bias in the estimation of trade costs for given expenditure and production at the disaggregated level. But disaggregation can never be as fine as reality, so some degree of aggregation bias is inevitable.

3.9 *Criticisms of the Gravity Approach*

The main criticisms against using gravity equations such as (8) and (5) to analyze trade volumes and trade costs are these:

1. Estimates of the distance elasticity of trade costs are unrealistically high (Grossman 1998) and have not dropped over time in the face of globalization ("the missing globalization puzzle"—David Coe, Arvind Subramanian, and Natalia Tumirisa 2002).
2. There are no import-competing sectors or non-tradables sectors that only supply to the domestic market (Charles Engel 2002).
3. One should allow the elasticity of substitution between domestic and foreign goods to be different from the elasticity of substitution among domestic goods (Engel 2002).
4. In contrast to the predictions of the model, the substantial increase in U.S.-Canada trade during the 1990s was not accompanied by a big drop in intra-provincial trade (Helliwell 2003).
5. The model implies trade among all countries for each sector, while the reality is dominated by zeros (Jon Haveman and Hummels 2004).
6. Estimated trade barriers are unrealistically high (Balistreri and Hillberry 2002).
7. Estimates of the gravity equation have the unrealistic implication that consumer prices are much higher in Canada than in the United States (Balistreri and Hillberry 2002).

Criticisms of the empirical validity of gravity equations are implicitly (sometimes explicitly) criticisms of the assumptions of theories underlying the gravity framework.

[56] The critical aspect of the Yi model that leads to this magnification effect is the asymmetry of production functions. This can be seen most clearly in comparison to the model of Eaton and Kortum (2002). As discussed above, the Eaton and Kortum model leads to a standard gravity equation, so there is no magnification effect. In the Eaton and Kortum model firms produce goods that are used both as intermediate goods and final consumption goods. Intermediate goods produced in one location are used as inputs in the production of intermediate goods in another location, which are used as intermediate inputs in another location, etc. Components will therefore have crossed borders infinitely many times. But in contrast to Yi (2003a,b), the same bundle of intermediate goods is used for the production of each good. This symmetry assumption regarding production functions kills the magnification effect.

Below we will therefore discuss how these criticisms can be addressed by generalizing assumptions of theories behind gravity equations (8) and (5). But before doing so, we will first discuss these criticisms in a bit more detail.

The puzzle of unrealistically high distance elasticities was first raised by Grossman (1998). Grossman pointed out that a distance coefficient of -1.42 in McCallum's gravity equation implies that regions that are 500 miles apart will trade 2.67 times more with each other than regions that are 1000 miles apart, which he considered implausibly large. The distance elasticity is much lower (-0.79 with a reported standard error of 0.03) in Anderson and van Wincoop (2003). This remains unreasonably large if distance is intended to capture transport costs. Grossman (1998) reasoned as follows. First write $t_{ij}=1+\tau_{ij}$, where τ_{ij} is the tax equivalent of transport costs. Defining $\alpha=\tau_{ij}/(1+\tau_{ij})$, one can write the elasticity of bilateral trade with respect to distance (holding constant multilateral resistance indices) as:

$$\alpha(1-\sigma)\partial ln(\tau_{ij})/\partial ln(d_{ij}). \qquad (4)$$

Transport costs are plausibly no more than proportional to distance, so that the elasticity $ln(\tau_{ij})/\partial ln(d_{ij})$ is less than one. Hummels (2001a) estimates it to be about 0.3.[57] As discussed in section 1, the tax equivalents for transport costs for the United States are on average about 11 percent. With the elasticity σ in the range of 5 to 10, the distance elasticity can be expected to be in the range of -0.12 to -0.26. If we include time costs, the tax equivalent of U.S. transport costs are on average 21 percent. Even then the implied distance elasticity is below available estimates, in the range of -0.21 to -0.46 for σ in the interval [5,10].

[57] Grossman (1998) assumes a constant distance elasticity of τ_{ij}. However, the estimated trade-distance elasticities from the gravity equations that Grossman refers to assume a constant distance elasticity of t_{ij}. If the Grossman specification is correct, it is possible that the high distance elasticities obtained from gravity equation estimation are a result of specification error of t_{ij}.

Apart from changing theoretical assumptions, to which we turn in a moment, there are at least two possible explanations in the literature for this distance elasticity puzzle. First, the distance may proxy for much more than trade costs. Portes and Rey (2002) find that the distance elasticity halves to -0.23 once information barriers are introduced separately. A second explanation is offered by Coe et al. (2002). They estimate the theoretical gravity equation (8) in levels rather than logs, which reduces the absolute value of the distance elasticity from -1.08 to -0.35 for 2000 trade data. A possible explanation for their finding is that the absolute value of the elasticity of trade with respect to distance is falling in distance, as Eaton and Kortum document. Cutting off the high-distance, low-trade volume observations by using logarithmic data with exclusion of zeroes excludes observations for which the elasticity is low in absolute value, hence pushing upward the estimated distance elasticity.

Apart from the level of the distance elasticity there is also the so-called "missing globalization puzzle," which is based on findings in a large number of studies that the distance elasticity has not declined or even risen over time (see survey of this literature in Coe et. al. 2002). This is often considered puzzling since transport costs have declined over time. Coe et al. (2002) find that when the gravity equation is estimated in levels rather than logs, the distance elasticity has fallen over the last two decades, from -0.51 in 1980 to -0.35 in 2000. It is not clear, though, why estimation in levels would resolve this puzzle. For industrialized countries, Brun et al. (2002) succeed in reversing a rising distance elasticity by including the cost of fuel in the trade cost function. A rise in the price of fuel acts like a negative productivity shock to transportation, so omitting the fuel cost variable in an era when it is rising will produce rising distance elasticities. Fuel costs both rose and fell in the period of their analysis, 1962–96, and it is not clear why the omitted fuel-cost variable is associated with rising distance elasticities in subperiods when fuel costs are falling. Moreover, fuel costs should affect the trade cost of developing countries symmetrically, so it is not clear why this effect only works with rich countries. The missing globalization puzzle may not be a real puzzle after all if one realizes that since 1913 technology growth in shipping has been relatively slow in comparison to the rest of the economy. Transport costs may therefore have increased as a fraction of average marginal production costs. This point is emphasized in Estevadeordal et. al. (2003).

The second criticism, raised by Engel (2002), is that there are no import-competing sectors or nontradables sectors that do not export. Engel makes this comment on Anderson and van Wincoop (2003) and argues that ignoring sectors that do not export can lead to an underestimation of border barriers. His reasoning is most clearly understood for nontradables. Assume that as a result of a rise in border barriers there is a shift in resources out of tradables into nontradables. The nontradables are only sold locally within a region and not between regions. As a result, a given 10-percent drop in exports of provinces will lead to a smaller increase in trade with other provinces than in the absence of nontradables. Engel argues that therefore a bigger border barrier is needed to account for the observed home bias. Under trade separability, however, this is not the case for the conditional general equilibrium model upon which estimation is based. Nontradables do not affect the gravity equation for tradables, and trade barrier estimates are therefore unaffected.[58]

To better understand this seemingly paradoxical result, consider the gravity equation (8). Let Y_{US} and Y_{CA} be the tradables output of respectively individual states and provinces, P_{US} and P_{CA} be the respective multilateral resistance variables for states and provinces and b the border barrier. Then trade between two provinces is equal to

$$\frac{Y_{CA}Y_{CA}}{Y_w}\frac{1}{P_{CA}^{1-\sigma}P_{CA}^{1-\sigma}}.$$

while trade between a province and state is equal to

$$\frac{Y_{US}Y_{CA}}{Y_w}\frac{b^{1-\sigma}}{P_{US}^{1-\sigma}P_{CA}^{1-\sigma}}.$$

For observed tradables outputs Y_{US} and Y_{CA}, the ratio of inter-provincial to state-province trade does not depend on nontradables and the estimated border barrier does not.

In contrast, full general equilibrium comparative statics is affected by the presence of nontradables. In the presence of nontradables a rise in border barriers will shift resources to the nontradables sector, so that tradables output of Canadian provinces, Y_{CA}, drops. Tradables output of U.S. states drops much less. As a result inter-provincial trade will rise less and state-province trade will drop more. For every 1-percent drop in international trade, the increase in interprovincial trade will be lower.

In short, the gravity equation and estimates of trade barriers are unaffected by non-tradables but comparative statics are. The same conclusion cannot be drawn for an import-competing sector. Many manufacturing firms do not export, a fact that cannot be accommodated by the gravity equation (5). Below we will consider an extension of fixed trade costs that modifies the gravity equation and allows for certain varieties to be only supplied to the domestic market.

The third criticism, also by Engel (2002), is that gravity theory does not allow the elasticity of substitution between domestic and foreign goods to be different from the elasticity of substitution among domestic goods. There is good reason to consider the case of a higher elasticity of substitution among domestic goods, commonly assumed

[58] Under trade separability, nontradables do not affect the marginal utility (or marginal productivity) of different types of tradable goods within a sector.

in the "new open economy macro" literature. In the context of Eaton and Kortum (2002) it is equivalent to the variation of productivity among regions within a country being lower than among regions of different countries, which is undoubtedly true. Extensions of gravity in this direction are discussed below.

The fourth criticism is raised by Helliwell (2003), based on a finding in another paper, Helliwell (1999). He documents that following the 1988 free trade agreement between the United States and Canada, trade between them rose rapidly while trade within Canada (interprovincial trade) did not fall much. This stands in sharp contrast to the finding by Anderson and van Wincoop (2003) that border barriers have increased interprovincial trade much more (factor 6) than they reduced state-province trade (about 44 percent).[59]

Helliwell (2003) criticizes the plausibility of the outcome from the comparative statics exercise in Anderson and van Wincoop (2003). Comparative statics exercises depend on the entire general equilibrium. This is therefore not necessarily a criticism of the gravity equation itself, which is based on a conditional general equilibrium. It may be that the modeling outside of the gravity structure, in particular the production structure, is incorrect.[60] This doesn't affect the estimation of the trade cost parameters, which only depends on the gravity structure.

One possible explanation for the Helliwell puzzle is a nontradables sector, as discussed above. But this would only work if trade barriers lead to a significant shift out of tradables into nontradables. Because of its much larger size, the share of the tradables sector would not be much affected in the United States and therefore be much bigger than in Canada. There is no evidence of this in the data. With tradables defined as the sum of mining, agriculture, and manufacturing, its share in total output is about the same in the two countries.

The criticism by Helliwell raises a more general point. An obvious direction for future research would be to evaluate the validity of theoretical gravity equations with respect to their time series implications. This is especially useful since most estimation is based on cross-section data. Baier and Bergstrand (2001) estimate a gravity equation with pooled data for two periods (1958–60 and 1986–88) to understand what factors drive the increase in world trade. However, their use of price data is unlikely to properly capture changes over time in multilateral resistance indices. Estevadeordal et al. (2003) estimate a gravity equation with pooled data for three years (1913, 1928, and 1938). While they include country-specific dummies to capture the multilateral resistance variables, the theory tells us that these will change over time. One should therefore include separate country dummies for each year. In order to exploit the time series properties of the data most effectively, it is even better to estimate gravity equations in their structural form.

The fifth criticism, launched by Haveman and Hummels (2004), is that gravity models imply that all countries purchase goods from all suppliers. Using 1990 bilateral trade data among 173 countries for four-digit SITC categories, Haveman and Hummels show that in 58 percent of cases importers buy from fewer than 10 percent of available suppliers. The Haveman and Hummels criticism is directed at models with complete specialization. However, Eaton and Kortum (2002) derived the same gravity equation in a model without complete specialization. In that model there are generally a limited number of suppliers, each of which sells to countries with whom they have relatively low trade barriers (in comparison to other suppliers). Countries therefore buy goods from only

[59] Possibly related to the Helliwell evidence is the point raised by W. Mark Brown (2003) that industries where border barriers have disappeared when comparing data on interstate and state-province trade tend to still have much more interprovincial trade than interstate trade.

[60] Anderson and van Wincoop (2003) adopt a simple endowment economy.

one supplier (the cheapest one), even though many suppliers may exist. It can therefore account for zero trade flows at a disaggregated level. At the aggregate level, since it is equivalent to the CES model, the Eaton and Kortum model cannot account for zero aggregate bilateral trade flows, which are seen for trade among small regions (certain states and provinces) or developing countries. Below we will also discuss an extension with fixed cost as an explanation for zero trade flows.

The sixth criticism, by Balistreri and Hillberry (2002), that estimated trade barriers are unrealistically high, has already been discussed above. The numbers would certainly be too high when they are interpreted as transport costs, as Balistreri and Hillberry do. In reality though, they reflect lots of trade barriers that cannot be directly measured. It is hard to dismiss estimates of large trade barriers without direct evidence to refute this. It is quite possible though that various extensions of existing gravity theory that we discuss below will eventually lead to a consensus of lower trade barriers than reported in table 7.

The final criticism, by Balistreri and Hillberry (2002), is that gravity equations have unrealistic implications for price differences. They argue that the Anderson and van Wincoop (2003) results imply that the consumer price index is 24 percent higher in Canada than in the United States (assuming σ=5). As discussed above, we think the focus on consumer prices indices in the data is a misuse of the model. They include things such as non-tradables, taxes and subsidies, and relative price fluctuations with exchange rates due to nominal rigidities of prices. We interpret multilateral resistance in the context of gravity theory as an ideal index of trade costs, and find the 24-percent greater trade cost index for Canada to be quite plausible given the revealed trade-cost impact of the border, Canada's much smaller size, the revealed trade-cost impact of distance, and the greater distance of Canadian provinces from U.S. centers of activity.

3.10 *Extensions of the Gravity Approach*

Simple extensions can meet criticisms and allow greater flexibility while retaining the essential simplicity of the gravity model (5)–(7). Three assumptions lead to gravity (5): (i) trade separability, (ii) the aggregator of varieties is identical across countries and CES, and (iii) ad-valorem tax equivalents of trade costs do not depend on the quantity of trade. A variety of extensions of (ii) and (iii) are discussed below. With relaxation of (i), two-stage budgeting is impossible. It is worthwhile to at least consider how estimates of trade barriers depend on this, but no work has yet been done in this direction.

Extensions of the CES Structure. Yuriy Tchamourliysky (2002) considers non-homothetic CES preferences. He modifies a standard CES consumption index by adding a constant δ_i to consumption of goods from country i by every country's consumers. He then derives a modified gravity equation and estimates δ_i to be negative, interpretable as a subsistence requirement. A subsistence requirement makes trade flows less sensitive to trade barriers. Incorrectly assuming homothetic CES preferences would make consumption excessively sensitive to trade barriers and the estimated trade barriers therefore too high. Tchamourliysky (2002) focuses particularly on distance as a trade barrier and finds that allowing for non-homothetic preferences reduces the impact of distance on trade barriers. This extension can potentially address several of the criticisms raised above. First, by reducing estimates of trade barriers, it addresses the concern by Hillberry and Balistreri (2002) that estimated trade barriers from theoretical gravity equations are unrealistically high. Second, a lower distance elasticity of trade addresses the concern of Grossman (1998).

The CES preference structure can also be generalized by assuming, using nested CES structures, that certain goods within a sector are better substitutes than other goods. This would for example allow domestic goods to be closer substitutes for each other than domestic goods are with foreign goods, addressing the criticism of Engel (2002). How much this matters to trade costs is unknown.[61]

Differences in Preferences and Technology. The gravity model assumes that preferences and technology are the same for all agents. This is a strong assumption and can be relaxed in a number of directions. First, it is possible that consumers in different regions or countries have different preferences. For example, consumers may be biased towards goods produced in their own country. Differences in preferences are, however, empirically indistinguishable from trade costs. Consider the following utility function for country j consumers:

$$\left(\sum_i \beta_i^{(1-\sigma)/\sigma} (c_{ij} / \gamma_{ij})^{(\sigma-1)/\sigma} \right)^{\sigma/(\sigma-1)} . \qquad (28)$$

Here c_{ij} is consumption of goods produced in country i by country j consumers. Utility generally differs across countries due to variation in the γ_{ij}. It is easily verified that in the absence of trade barriers the gravity equation is the same as in (8) with t_{ij} replaced by γ_{ij}. Under sufficiently strong restrictions on taste differences, it is possible to distinguish trade costs from taste differences.[62] Another promising route to distinguishing the two is to exploit time-series variation. Trade costs plausibly move over time while tastes presumptively are stable.

Evans (2003b) provides some evidence suggesting that estimates of large border-related trade barriers are not a result of a home bias for domestic goods. Information about the size of border-related barriers is usually obtained by comparing international trade to domestic trade within a gravity framework. Evans (2003b) estimates a traditional gravity equation to find that, after controlling for size, distance and remoteness, domestic sales within non-U.S. OECD countries is 4.36 times imports from the United States. But she finds that the home bias is virtually identical when comparing local sales in non-U.S. OECD countries of affiliates of U.S. multinationals to imports from the United States. This suggests that location—and not the nationality of the firm that sells the goods—is critical.

Demand can also be different when it comes from firms buying intermediate goods. Assume that for a specific industry the index of intermediates in the production function differs across firms. For example, let (28) represent the index of intermediates in the production functions of firms in country j. The same comment can now be made as for heterogeneous preferences of consumers. One cannot empirically distinguish trade barriers t_{ij} and the parameters γ_{ij} in production functions. If production functions of firms tend to give more weight to intermediates that are produced in their own country, it will be indistinguishable from a border barrier. One can therefore estimate positive border barriers even if none exist.

[61] Consider, for example, data on trade flows for U.S. states and Canadian provinces. Assume a one-sector world. Let σ_H be the elasticity of substitution between goods produced within a country and σ_F the elasticity of substitution between domestic and foreign goods. It is then easily verified that the gravity equation (8) is modified as follows:

$$X_{ij} = \frac{Y_i Y_j}{Y_w} \frac{t_{ij}^{1-\sigma_H}}{P_{j,c}^{\sigma_F-\sigma_H} P_j^{1-\sigma_F} \Pi_i^{1-\sigma_H}}$$

where $P_{j,c}$ is the price index of goods imported by region j from country c in which the exporting region i is located, P_j is the overall price index of importer j and Π_i is a price index specific to exporter i. In general the estimated trade barriers can now be different. The direction in which they will change is not obvious though and will depend on the details. It is easy, for example, to show that the estimate of the U.S.-Canada border barrier is the same as based on the standard gravity equation (8) when this is the only trade barrier.

[62] Bergstrand (1985) derives a gravity equation under a different type of heterogeneity in preferences. Consumers consider the elasticity of substitution among goods produced in foreign countries to be different from the elasticity of substitution between home and foreign goods, where "home" and "foreign" varies with the location of the consumer.

Fixed Costs of Trade. Bernard and Jensen (1997), Bernard and Wagner (1998) and Roberts and Tybout (1997) all report substantial evidence of fixed entry costs into foreign markets. They all show that having exported in the past significantly increases the probability of a firm exporting today. Introducing fixed costs can explain the many zeros in bilateral trade data. Coe et al. (2002) and Evenett and Venables (2002) also document that the number of zeros has substantially dropped over time. This suggests that a reduction in fixed costs or growth of income can play an important role in accounting for the growth of world trade. Evenett and Venables (2002) find that the removal of zeros accounts for one third of developing countries' export growth since 1970.

Evans (2003c) attempts to recompute the tax equivalent of proportional trade costs for U.S. exports after controlling for fixed costs. She assumes that fixed trade costs are borne by exporters. Once a firm has paid the sunk cost, it can export to all foreign markets. She finds that in 1992 only 25 percent of all U.S. firms exported abroad. The U.S. Census of Manufactures provides data for each industry on the proportion of U.S. output that is produced by firms that both export and sell domestically. The output Y_i^k of country i in industry k is then redefined as the output of firms that sell in both domestic and foreign markets. Evans (2003c) finds that estimates of proportional trade costs drop somewhat as a result of this. However, there are two problems with this approach. First, she estimates a traditional gravity equation. Second, one cannot ignore firms that only supply varieties to the domestic market. These are import-competing firms that affect the demand from firms in the industry that supply goods to both the domestic market and foreign markets.

In contrast to Evans (2003c), Peter Klenow and Andres Rodriguez-Clare (1997) develop a model in which fixed costs are borne by importers.[63] In that setup, it is possible that firms only export to a limited number of markets. The same would be the case if fixed costs are borne by exporters but vary across export markets. Based on data for Costa Rica from 1986 to 1993, Klenow and Rodriguez-Clare (1997) find that a drop in tariffs from 48.5 percent to 22.1 percent over this period led to an increase in the number of imported varieties of 30 percent for consumer goods and 20 percent for intermediate goods.

A simple generalization of (5)–(7) takes into account fixed costs. Assume that as a result of fixed costs in industry k country i exports to the limited set of countries X_i^k and imports from the set M_i^k observed from the trade-flow data. When there are positive exports from i to j the generalized gravity equation becomes

$$X_{ij}^k = \frac{E_j^k Y_i^k}{Y^k}\left(\frac{t_{ij}^k}{P_j^k \Pi_i^k}\right)^{1-\sigma_k} \qquad (29)$$

$$(\Pi_i^k)^{1-\sigma_k} = \sum_{j \in X_i^k}\left(\frac{t_{ij}^k}{P_j^k}\right)^{1-\sigma_k}\frac{E_j^k}{Y^k} \qquad (30)$$

$$(P_j^k)^{1-\sigma_k} = \sum_{i \in M_j^k}\left(\frac{t_{ij}^k}{\Pi_i^k}\right)^{1-\sigma_k}\frac{Y_i^k}{Y^k}. \qquad (31)$$

This setup allows for the case where firms in a country only sell to the domestic market (import-competing firms), sell to a limited set of foreign markets, or sell to all foreign markets. On the importing side it allows for the case where a country only purchases domestic varieties, where it purchases a limited set of domestic and foreign varieties, and where it purchases all varieties from all countries. It can therefore address several of the criticisms of gravity models discussed above. Note that (29)–(31) give rise to zero trade flows independently of fixed trade costs if either E_j^k or Y_i^k is equal to zero. These are unrestricted in conditional general equilibrium.

[63] They use their model to compute welfare gains from trade liberalization.

We can either estimate the structural form or estimate the log-transformation with OLS after replacing the multilateral resistance indices with importer and exporter dummies. In the latter case the estimates for proportional trade barriers remain the same as before when only positive trade flows are included in the regression. However, estimates based on aggregate gravity equations overstate proportional trade barriers since they aggregate zero and positive trade flows. We expect fixed costs to be a fruitful area for further research.[64]

4. *Evidence from Prices*

Prices provide another indirect source of information about the magnitude of trade costs. Two separate literatures shed light on trade barriers using price data, a trade literature and a macro literature. The trade literature has focused on comparing import or "world" prices to domestic wholesale prices. The aim is usually to estimate NTBs. The macro literature compares retail prices of similar goods across countries. The macro literature is focused on issues such as the speed of convergence of prices across countries and the relationship between exchange rates and prices. We will only discuss what can be learnt about the magnitude of trade costs from this literature. Broader surveys can be found elsewhere.[65]

4.1 *The Trade Literature*

It is useful to first introduce some notation. Consider two countries, i and j, with i a major exporter of a particular good and j an importer. The wholesale price of the good in country j is p_j. The c.i.f. import price at the port of entry is p_j^m, while the f.o.b. export price is p_i^x. The literature aims to extract information about policy barriers by comparing either the "world" price p_i^x or port of entry import price p_j^m to the domestic wholesale price p_j.

There are both conceptual and data problems associated with this method. The main conceptual problem for our purposes is that the price comparison captures only a limited component of the overall trade barrier t_{ij} between two countries. It also does not accurately capture NTBs, for which the method is designed. Data problems take the form of measurement problems and limited data coverage.

First consider the conceptual problems. In most reasonable models of economic activity, a large component of trade costs, often the most important component, is borne by the exporter and then shifted onto the importer. The c.i.f. import price includes not only the standard transport, insurance, and freight costs, but also the myriad of other costs borne by the exporter in order to bring the good to the foreign market. The price ratio p_j/p_j^m therefore does not capture this portion of full trade cost. It only captures the trade costs directly borne by the importing country, those associated with policy barriers in the form of tariffs and NTBs, as well as more informal trade costs borne by the importing country, such as information, regulatory, and contract costs.

Deardorff and Stern (1998), in their survey of the literature on the measurement of NTBs, explain why the price ratio p_j/p_j^m does not necessarily capture all trade costs associated with quotas and other NTBs. If for example the exporting firm has market power it may be able to extract the quota revenue for itself by raising the import price.[66] Another possibility is that a quota licence is

[64] Bergstrand (1985, 1989, 1990) and Baier and Bergstrand (2001) model the cost of distributing, marketing and tailoring a product to an export market as a CES transformation function. Total output of a particular good is a CES function of the quantities of that good sold to various markets. It is not exactly clear though what the microfoundations are for this transformation function. It leads to additional price indices in the gravity equation.

[65] See, for example, Kenneth Froot and Kenneth Rogoff (1995) and Pinelopi Koujianou Goldberg and Michael Knetter (1997).

[66] Michael Knetter (1997) provides evidence that German exporters charge substantially higher prices to Japan due to a variety of NTBs in Japan.

allocated directly to the final user of the imported good. In that case no price comparison will be able to capture it. Finally, voluntary export restraints (VERs) allocate the quota to exporters who presumptively build the value of their quota licenses into the export price. The Multi-Fibre Arrangement is the most prominent example.

The data problems besetting price comparisons are twofold: limitations of coverage and imperfections in the data which do exist. As to limitations of coverage, perhaps the main data problem, survey data on price comparisons, is mostly limited to agricultural products and, to a lesser extent, textiles and clothing. Even within these categories data are only available for certain countries and years. As to imperfections of the data that do exist, they can be categorized in three types. First, the wholesale price p_j is not a port of entry price and therefore already contains some local distribution costs. Second, the price p_j is usually for a domestic substitute of the import good or an index of imported and domestic goods. The analyst making price comparisons must confront the issue of comparability of the goods. Even a physically homogeneous good (such as Number 2 Red wheat) has variations with respect to terms of delivery. The price comparison method is most convincing where markets are thick and well organized. Third, there are timing problems, which are particularly an issue when the import or world price is denominated in another currency and a correct exchange rate needs to be used for price comparison.

The lack of availability of survey data on prices has led some researchers to compute unit values (e.g. Sazanami et al. 1995). This is only possible for categories of goods that are sufficiently broadly defined that domestic production exists and for which sensible quantity units exist. The domestic price p_j is then computed by dividing the value of domestic production by the quantity of output, and similarly for imports. The approximations resulting from this technique tend to be very crude and yield very different results than survey data when both are available. The comparison is often over very dissimilar goods.

Deardorff and Stern (1998) and Laird and Yeats (1990) both review studies that have estimated trade barriers based on price comparisons. The overall conclusion one reaches from the evidence is that trade barriers are very large in the agricultural sector. Table 9 reproduces from Deardorff and Stern a few price comparison measures in agriculture in the United States for 1991 and 1993. Agricultural policy commonly starts with a price support fixed by the government, and uses quotas or variable levies (in the case of the European Union) to avoid supporting farmers in the rest of the world. Sugar is the most notorious example of large distortions. In 1991 the sugar support price was equivalent to a 125-percent tariff, while in some years the tariff equivalent has exceeded 300 percent. Protection for U.S. dairy products is also very high. Canada also protects its agricultural sector. Deardorff and Stern report 1992 tariff equivalents of 165 percent in dairy products, 28 percent in chicken, and 29 percent in turkey. Other countries have even more extreme agricultural distortions—Japan is notorious for a domestic rice price over ten times the world price, with similar differentials for sugar in some years.

4.2 *The Macro Literature*

The purchasing power parity literature compares retail prices of individual goods or baskets of goods across countries. The main weakness of the literature to date is the absence of a theoretical foundation necessary to link evidence on relative prices across countries to trade barriers. We suggest various directions that trade theory can be employed to extract more information from price data.

Papers that attempt to draw a link from relative prices to trade barriers commonly refer to an arbitrage equation of the type

TABLE 9
Price-Gap Measures by Sector, 1991 and 1993 (Percent)

Sector	1991	1993				
Price-gap measures Agricultural Sector[a]	Quota Tar. Eq.	Quota Tar. Eq.	S Tariff	TW Tariff	Quota + TW Tar.	%US Imports
Sugar	124.8	93.7	1.0	0.2	93.9	0.1296
Butter	26.9	20.8	8.8	8.7	29.5	0.0008
Cheese	35.4	37.4	10.6	11.2	48.6	0.0889
Dry/condensed milk prod.	60.3	60.3	13.2	13.2	73.5	0.0033
Cream	60.3	60.3	0.0	0.0	60.3	0.0008
Meat	6.5	5.0	1.3	4.0	9.0	0.4296
Cotton	-	27.0	2.2	0.1	27.1	0.0001
Motor Vehicles[b]	-	0.4	2.2	1.7	2.1	9.5673
Maritime trans. (Jones Act)[c]	133.0	89.1				

Notes:
a. The price comparisons for the agricultural products were as follows. Sugar calculated as the difference between the U.S. price and the world price, inclusive of transport costs and import duties, expressed as a percentage of the world price; data from USDA, Sugar and Sweetener: Situation and Outlook Yearbook. Dairy products—based on domestic and world price data collected by the USDA for whole milk powder, butter and cheese; for dry/condensed milk products and cream, the price gap for whole milk powder was used as a proxy. Meat — based on the "market price support" portion of the producer subsidy equivalent (PSE) calculated by the OECD, comparing Sioux Falls (U.S) cutter prices with New Zealand milk cow prices, and domestic and world prices of Orleans/Texas "B" index cotton, including adjustments for transportation and marketing costs.
b. Motor vehicles based on an estimated 1.5 percent additional increase above the industrywide U.S. price increase needed in Japanese autos to equate supply and demand in the presence of the auto import restraint; weighted by the percent of Japanese auto imports to total whole imports.
c. Maritime transport calculated as the output-weighted average difference between the U.S. and world price for shipping "wet" and "dry" cargo, the weighted differences are between the U.S. price for shipping Alaskan North Slope crude petroleum to the U.S. west coast and to the U.S. gulf coast and the average world price for comparable tanker shipments transported equal distances; a separate estimate from the literature was used for the tariff equivalent for dry cargo. Additional estimates of the price gap attributed to the Jones Act can be found in White (1988) and Francois et al. (1996, p.186).
Sources: USITC (1993, 1995). Quota tariff equivalents and the notes for price comparisons above are reproduced from Deardorff and Stern, (1998), Table 3.6 (Source USITC 1993, 1995). Tariff averages, trade-weighted tariff averages and US product import percentages (except maritime transport) are calculated using data from the UNCTAD TRAINS database (Haveman repackaging). SIC to HS concordances from the US Census Bureau are used.

$$\frac{1}{t_{ij}} \le \frac{p_i}{p_j} \le t_{ij} \tag{32}$$

with t_{ij} the cost of arbitraging a price differential between i and j by a wholesaler. The relative price is assumed to move freely between the arbitrage points, also referred to as Heckscher's (1916) commodity points. Equation (32) is unfortunately of limited relevance in understanding the link between price differentials and trade costs in most markets.

Limitations on arbitrage pose a significant limitation to using (32). In many cases the arbitrage costs of wholesalers are prohibitive. Producers often obtain exclusive national marketing licences, which precludes arbitrage by wholesalers. Moreover, as pointed out by Obstfeld and Rogoff

(2000), even small firms that do not have exclusive distribution rights can price discriminate by dealing with a small number of wholesalers with whom they have developed long-term relationships. Other factors, such as warranties or small differences in products due to regulatory constraints, also contribute to limit the ability of wholesalers to arbitrage price differences. Pinelopi Koujianou Goldberg and Frank Verboven (2001) provide explicit evidence for the European car market showing that arbitrage activity by resellers is very limited even though price differences are large.

General equilibrium forces limiting price variation pose an even greater limitation to using (32). Even though (32) holds in most models, the relative price cannot freely fluctuate in the range given by Heckscher's (1916) commodity points. Below we will show that arbitrage alone leads to a much tighter condition. Assuming a specific trade model can tie down the relationship between relative prices and trade costs exactly. We will illustrate this point and examine the link between relative prices and trade costs below in the context of a version of the Ricardian model of Eaton and Kortum (2002).

While we will focus this survey on evidence from price levels at a point in time, some studies have employed evidence on changes in relative price over time to extract information about trade costs motivated by (32). We will first briefly review the time series evidence. The remainder of this section discusses what can be learned from price levels at a point in time.

4.2.1 *Time Series Evidence*

A well-known paper by Charles Engel and John Rogers (1996) computes the standard deviation of relative prices between Canadian and U.S. cities for fourteen consumption categories (such as alcoholic beverages and men's and boy's apparel). They show that the standard deviation of relative prices depends positively on distance and is much higher when two cities are separated by a border than when they are located in the same country. They argue that this may be a result of trade barriers, as arbitrage equation (32) suggests that larger relative price differences are possible if trade barriers are larger. But Engel and Rogers (2001) provide evidence that exchange-rate volatility may be the main explanation for the border effect. In that study they use evidence from cities in eleven European countries, which allows them to compute the border effect after controlling for bilateral exchange-rate volatility. The coefficient on the border dummy drops from 2.85 to 0.21.

Maurice Obstfeld and Alan Taylor (1997) is another example of a study that used price index information. Their starting point is again (32). They estimate a threshold autoregression (TAR) model[67], whereby the log-price differential follows a random walk inside a band $[-c,c]$. It can also exit the band and will then converge back to the band at a rate to be estimated. Their estimates for c tend to be rather small, on average about 0.08 for the United States relative to other countries. Obstfeld and Taylor interpret them as estimates of trade barriers based on arbitrage equation (32).

Since theory tells us that relative prices are not free to fluctuate in the range suggested by (32), it is not a correct starting point for either time series evidence or evidence about price levels at a point in time. There are many factors that contribute to time series variation in relative prices, such as changes in marginal costs, trade costs, taxes, markups, and exchange rates. It is hard to see how information can be extracted about the *level* of trade costs from evidence on *changes* in relative prices, especially without the guidance of theory.

4.2.2 *Extracting Information from Price Levels*

Recent survey evidence of price levels for individual goods in cities around the world

[67] See also Alan Taylor (2001) and references therein.

has led to a small but promising literature aimed at extracting information about trade costs. Since the approach that most authors have taken is rather atheoretical, we will first discuss some theoretical background to interpret the findings from this literature.

Some Theoretical Background. The price paid by the final user of a good generally contains four components: (i) the marginal cost of production, (ii) trade costs, (iii) various monopolistic markups over cost in the chain from producer to final user, and (iv) subsidies and taxes. In order to shed light on the relationship between trade costs and price differentials we will first abstract from the last two price components. The key force is a general equilibrium or multi-market version of arbitrage which constrains the behavior of relative prices.

If country i buys a good from country m the price in i will be $p_i=c_m t_{mi}$, where c_m is the cost of production in m and t_{mi} is one plus the tax equivalent of trade costs on shipments from m to i. Country i will source from the producer m for which $c_m t_{mi}$ is the lowest. Arbitrage is done here not by consumers or wholesalers, but by producers. If the price in market i is above $c_m t_{mi}$, it is profitable for the producer in country m to undercut the existing price in country i.

Without imposing any specific model structure, we can already say something about the price in location i relative to the price in location j. Specifically, if it is optimal for country i to source from country z_i and country j from z_j, then it must be the case that the price in i is no larger than if it had sourced from z_j and the price in j is no larger than if it had sourced from z_i. These arbitrage constraints lead to the following inequalities

$$\frac{t_{iz_i}}{t_{jz_i}} \leq \frac{p_i}{p_j} \leq \frac{t_{iz_j}}{t_{jz_j}}. \tag{33}$$

The arbitrage equation (33) is generally much tighter than the equation (32) commonly referred to in the literature. As an example, consider the case where i and j purchase the good from the same producer. If the producer is located in country m, the relative price is equal to

$$\frac{p_i}{p_j} = \frac{t_{im}}{t_{jm}}. \tag{34}$$

In this case the relative price is completely tied down by trade barriers. Analogous points were made in Deardorff (1979) in the context of forward and spot currency trade with transactions costs.

It also follows from (34) that trade costs do not necessarily lead to price differentials. If both i and j face the same trade barrier with m, their relative price is equal to one. On the other hand, in the specific case where m is one of the two countries, the relative price captures exactly what we intend to measure. If $m=j$ the relative price is t_{ij}/t_{jj}, the trade cost between i and j relative to the trade cost within country j. A natural strategy would be to identify the source country for each product. The price in country i relative to the source country is informative about international relative to local trade barriers. Unfortunately survey data often do not tell us which country produced the good. In some cases the price is not even for a specific good, but an index of similar goods. We will return to these issues below.

More can be learned about the relationship between relative price differentials and trade costs by adopting a specific trade model with a specific economic geography. This is useful for gaining perspective on the atheoretical literature which looks at the geographic dispersion of prices for evidence of trade costs. We simulate a model to generate distributions of prices and then relate them to the trade cost parameters we impose.

For the trade model we will consider a variation of Eaton and Kortum (2002). There are N countries that each have the same size labor force. There are G goods, with an elasticity of substitution σ between them. Productivity z for each good in each country is drawn from the Fréchet distribution with cumulative distribution function

$e^{-z^{-\theta}}$. The variance of z is inversely related to the parameter θ.

The countries are evenly spaced on a circle. The average trade cost for any pair of countries is proportional to their distance on the circle. Trade costs vary across both goods and location pairs. This extends Eaton and Kortum, who assume that trade costs are identical across goods. The average trade cost of good g (across location pairs) is $2t_{av}(g-1)/(G-1)$. Average trade costs therefore vary across goods from 0 to $2t_{av}$. The average trade cost (tax equivalent) across all goods and location pairs is t_{av}.

Each country buys from the cheapest producer, so that the price of good g in country i is

$$p_i^g = min(c_j^g t_{ji}^g,..,c_N^g t_{Ni}^g) \qquad (35)$$

where c_j^g is the production cost of good g in country j and t_{ji}^g is one plus the tax equivalent of trade costs on shipments of good g from j to i. When there is a continuum of goods, total labor demand will be the same across countries and wage rates will be equal. The level of this common wage rate is irrelevant for the price dispersion measures. We approximate the continuum model by simulating the model for a large number of goods, with G=3000.

In simulations of the model, we compute five different price dispersion measures that have been reported in the literature. The measures can be location-specific, location pair-specific or good-specific. The five measures are:

1. location-specific: for given location i the average of $|p_i^g - p_j^g|/(0.5p_i^g + 0.5p_j^g)$ across both goods and locations j;
2. good-specific: average of $|p_i^g - p^g|/p^g$ across locations i, with p^g the average price of good g across locations;
3. location-specific: expenditure weighted average of $(p_i^g - p_{low}^g)/p_{low}^g$ across goods, where p_{low}^g is the lowest price across countries of good g;
4. location pair-specific: the standard deviation of $ln(p_i^g) - ln(p_j^g)$ across goods for a given location pair (i,j);
5. standard deviation across both goods and countries of $ln(p_i^g) - \sum_j ln(p_j^g)/N$.

Particularly the third and the fourth measures have been computed with the aim of measuring trade barriers. Bradford and Lawrence (2003) and Mario Crucini, Chris Telmer, and Marios Zachariadis (2000) compute variations of measure 3 and interpret it respectively as a measure of "fragmentation" and "cost of arbitrage." This measure may both overstate and understate the trade-weighted average trade barrier. If the cheapest country is m, the price that i would pay to import from m is the price in m times the trade cost t_{mi}^g. If the trade cost is high, it may be optimal to import from another country with whom i has a lower trade barrier. In that case the measure overstates actual trade costs. It may also be optimal to purchase from a domestic producer in i. In that case the model tells us that the price difference with the cheapest country m is equal to the difference in production cost between i and m. The latter is smaller than the trade barrier between i and m (otherwise i would import from m) and may even be zero (if i is itself the cheapest country).

David Parsley and Shang-Jin Wei (2002) compute measure 4. They justify it based on the arbitrage equation (32). Assuming the same trade barrier for all goods, and that the relative prices between i and j are evenly dispersed in this no-arbitrage zone, there is indeed a direct relationship between the trade barrier between i and j and the standard deviation of these relative prices across goods. But arbitrage equation (32) is not a good starting point for relating price differences to trade barriers. Two extreme examples make the point. First assume that trade costs are so large that there is no trade. In that case price differences are entirely driven by differences in local production costs, which in general bear no relationship to trade costs. As a second example assume that i and j buy all goods from country m, with whom they have the same trade barrier for each good. The price dispersion measure

TABLE 10
TRADE COSTS AND PRICE DISPERSION

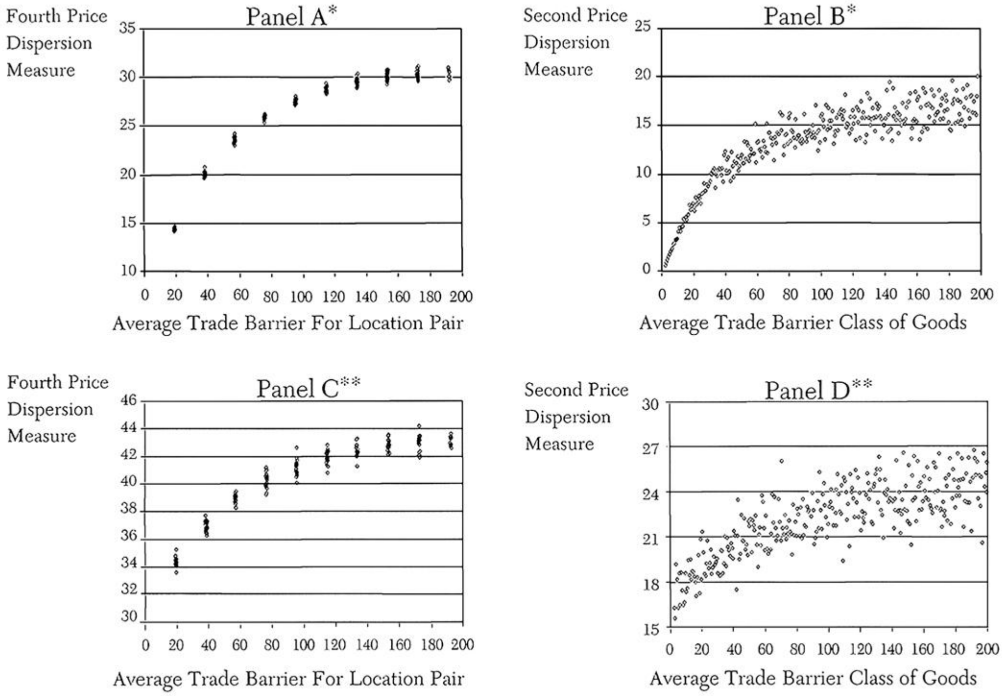
	parameters					Price Dispersion Measure					Trade Cost	
	N	σ	θ	trade cost average	vary/ same	1	2	3	4	5	trade-weighted average	ideal index
						Trade Cost Only Source of Price Dispersion						
(i)	10	4	5	100	vary	20	13	26	27	18	31	52
(ii)	10	4	5	200	vary	23	15	30	31	21	38	77
(iii)	10	1	5	100	vary	20	13	42	27	18	40	62
(iv)	10	1	5	200	vary	23	15	53	31	21	50	92
(v)	10	4	5	100	same	22	15	40	30	20	100	100
(vi)	10	4	3	100	vary	26	17	32	34	23	43	57
(vii)	10	4	8	100	vary	15	10	21	19	13	22	43
(viii)	20	4	5	100	vary	19	13	29	26	18	28	50
(ix)	20	4	5	200	vary	23	15	37	29	21	36	73
						With Other Sources of Price Dispersion						
(x)	20	4	5	0	vary	25	17	37	31	22	0	0
(xi)	20	4	5	100	vary	32	22	45	40	28	28	50
(xii)	20	4	5	200	vary	34	23	47	43	30	36	73
						Data						
(xiii)						32	21	40	44	24		

Notes: This table reports the average of five price dispersion measures described in the text. All numbers are in percentages. The model moments are based on the trade model described in the text, which is a variation of the Eaton and Kortum (2002) model. Results are shown for a variety of parameter assumptions in the model. N refers to the number of countries. σ is the elasticity of substitution between goods. θ is the parameter from the Fréchet distribution that is inversely related to productivity differences across countries. "vary/same" refers to whether the trade barriers t_{ij}^{g} vary across goods and countries as described in the text or are the same across al goods and countries. Trade weighted average trade costs and the ideal index of trade costs are shown as well. The latter is a common trade barrier across countries and goods that leads to the same overall expected trade level as implied by the trade costs assumed in the model. The last row reports data moments that have been computed by different authors. The samples are reported in Table 11.

will then be zero, even though trade costs may be very large.

Table 10 provides some results of model simulations for various model parameters. Only average price dispersion measures are reported (e.g. across location pairs or goods). Apart from the assumed average trade cost, the table also reports the trade-weighted average and the ideal index. The latter is the common international trade barrier (across goods and countries) that leads to the same aggregate international trade as implied by the assumed trade barriers. All reported numbers are in percentages. The trade-weighted average trade cost is much lower than the arithmetic average since international shipments tend to be limited to countries with whom the producer has a low trade barrier. Clusters of countries of varying sizes are formed, with one member of the cluster being the single source of production for all members of the cluster. The ideal index always lies in between the arithmetic average and trade-weighted average barrier.

Table 10 shows that even if trade barriers are the only source of price dispersion, it is hard to conclude much about the magnitude of trade barriers from the average price dispersion measures. Rows (i)-(ix) show results for nine different parameterizations. All price dispersion measures are far below the ideal



* Trade Costs Only Source of Price Dispersion

** Including Other Sources of Price Dispersion

Figure 2. Price Dispersion across Location Pairs and Goods

index. There is also no apparent relationship with the trade-weighted average index. If we lower the elasticity of substitution σ from four to one, the trade-weighted average barrier rises but the price dispersion measures (other than the third) remain unchanged. If we remove the variation of trade barriers across locations and goods, the trade-weighted average barrier rises from 31 percent to 100 percent but the price dispersion measures change very little. When we raise the number of countries from ten to twenty, the trade-weighted average barrier drops, while the third price dispersion measure rises. All price dispersion measures drop relative to the trade-weighted average barrier when the average trade cost is increased or θ is lowered.

Even if we knew the model and the parameters of the model other than trade costs, it would still be impossible to conclude much about the magnitude of trade costs, whether the arithmetic or trade-weighted average or the ideal index. The reason is that we do not know the distribution of trade costs across goods and location pairs.

While average price dispersion measures are not very informative about trade costs, the variation of price dispersion across location pairs and goods is. This is illustrated in panels A and B of figure 2 based on representative simulations of the model. Panel A shows the fourth price dispersion measure for each location pair as a function of the average trade cost for that location pair. Price dispersion is clearly higher for location pairs that have higher average trade costs. Panel B shows a similar result for goods based on the second price dispersion measure. Average price dispersion for a particular good (across location pairs) depends on the particular productivity draws for that good from the Fréchet distribution and therefore has some randomness to it. Each point in the panel represents the average for ten goods.

It is clear that price dispersion tends to rise for goods with higher trade barriers.

These findings can be exploited empirically. For example, one can specify a trade cost function like (11), relating trade cost across location pairs to distance and other observable characteristics that are associated with trade barriers. One can similarly add good-specific characteristics to the trade cost function. The parameters of the trade cost function can be estimated by using the variation of price dispersion across location pairs and goods. We will return to this in the discussion of empirical evidence below.

So far we have abstracted from the two other components of prices, associated with taxes and markups. We have also abstracted from differences in domestic trade costs across countries. Introducing domestic trade costs and taxes is relatively easy. The equilibrium prices in the model above then need to be multiplied by the ad-valorem tax equivalent of domestic trade costs and taxes. This affects the price dispersion measures to the extent that taxes and local trade costs vary across countries.

Variable markups also affect price differences. Markups depend on factors such as the price elasticity of demand and the market share of an oligopolist. In practice one of the most important factors affecting markups is nominal exchange rate volatility combined with nominal price rigidities in the buyer's currency. To the extent that exporters set prices in the buyer's currency, the relative price for the same good across two countries will fluctuate one-for-one with the exchange rate during the time that prices remain set. The profit margin of the exporter will fluctuate accordingly.[68]

[68] Essentially the same outcome occurs when the exporter sets the price in its own currency, but domestic distributors in the importing country (importers, wholesalers, retailers) absorb the exchange rate fluctuations in their profit margins by setting the price fixed in the local currency. See Philippe Bacchetta and van Wincoop (2003) for a model along this line, which shows that such price setting behavior may be optimal for both exporters and importers.

To illustrate the impact of these additional sources of price dispersion, we multiply the equilibrium prices by t_i^g for good g in location i. This captures local taxes and distribution costs. One can make an argument that it also captures ex-post markup variation due to exchange rate volatility and price rigidities, although introducing such features would require a substantial modification of the model. As an illustration we will assume that $ln(t_i^g)$ has a standard deviation of 23 percent across goods and countries, including country and good-specific components with standard deviations of both 5 percent.

Rows (x)–(xii) of table 10 show the average price dispersion moments which result from introducing variation in $ln(t_i^g)$. It further illustrates that average price dispersion measures tell us very little about trade barriers. For example, doubling the average trade barrier from 100 percent to 200 percent has remarkably little effect on the price dispersion measures. It is again the case that much more can be learned from the variation in price dispersion across location pairs and goods. Panels C and D of figure 2 show that across both location pairs and goods there remains a strong positive relationship between price dispersion and trade costs, even though it is somewhat clouded by the additional sources of price dispersion.

Evidence from Survey Data. Various authors have computed measures of price dispersion, using survey data for disaggregated goods from the OECD, Eurostat, and the Economist Intelligence Unit. Table 11 lists the papers that have computed the five price dispersion measures discussed above, as well as the data samples. Average price dispersion measures are listed in the last row of both tables 10 and 11. While some papers consider a much broader set of countries, for comparability we only report results for industrialized countries (mostly European countries).

Table 10 shows that the average price dispersion numbers in the data are not too far from those in the model when other sources

TABLE 11
DATA SOURCES FOR PRICE DISPERSION MEASURES

	Measure 1	Measure 2	Measure 3	Measure 4	Measure 5
author	Bradford & Lawrence (2003)	Crucini et.al (2001)	Bradford & Lawrence (2003)	Parsley & Wei (2002)	Crucini & Shintani (2002)
data source	OECD survey	Eurostat	OECD survey	Economist Intelligence Unit	Economist Intelligence Unit
countries	9 OECD	13 EC	9 OECD	11 EC	15 EC
goods	120	3545	120	95	270 (traded)
period	1990, 1999	1990	1990s	1990s	1990, 2000
average price dispersion	32	21	40	44	24

Notes: the Table reports the papers and associated data sources that have computed the five price-dispersion measures mentioned in the text. "Traded" refers to traded goods as an unknown subset of the total number of goods listed.

of price dispersion are included. But since this can be said for an average trade barrier in the model of 100 percent as well as for an average trade barrier of 200 percent, we cannot expect to learn much about trade barriers from this.

We saw that more can be learned by exploiting the variation of price dispersion across location pairs and goods. Parsley and Wei (2001, 2002) and Crucini et al. (2000) have related location pair-specific price dispersion measures to observables associated with trade barriers, while Crucini et al. (2000, 2001) have related good-specific price dispersion measures to various characteristics of goods. These studies confirm that variation of price dispersion across location pairs and goods is related to variation in trade barriers across locations and goods. This is consistent with the results from the model we discussed. The main weakness of the literature so far is that the empirical work is not based on any particular trade model.

Parsley and Wei (2002) regress the location pair-specific fourth price dispersion measure, the standard deviation of $ln(p_i^g)-ln(p_j^g)$ across goods, on various variables related to trade frictions. These include distance, common language, exchange rate volatility, membership of a currency union and average tariff rate. Using data for fourteen U.S. cities and cities in 69 other countries, they find that such regressions have a high adjusted R^2 in the range of 0.7 to 0.8.

Parsley and Wei are particularly interested in an estimate of the reduction in the tariff equivalent of trade barriers associated with currency unions. They find that the price dispersion measure is reduced by 3.2 percent for non-EMU country pairs with the same currency and by 4.3 percent for EMU countries, relative to countries that do not have a common currency. They transform these into tax equivalents by using that a 1-percent tariff reduction reduces the price dispersion measure by 0.86 percent. The reduction in the tax equivalent of trade barriers is then 3.7 percent for non-EMU countries in a currency union and 5 percent for EMU countries. These are much lower estimates of international trade barriers than those based on trade volume data in the gravity literature. The important difference though is that there is a theoretical foundation for estimates based on gravity equations. Theory does not predict a simple linear relationship between price dispersion and trade barriers.

Parsley and Wei (2001) apply the same method to relative prices between Japanese and U.S. cities for 27 goods. They find that price dispersion is positively related to distance and the presence of a border between two cities. They also find that the importance of the border has declined during their 1976–97 sample. Nominal exchange rate volatility and shipping costs help account for the border effect, but do not explain the decline in the border effect over time.

Crucini et al. (2000) run a regression of a different price dispersion measure on distance and distance squared. Their price dispersion measure for location pair (i,j) is $\sum_{g=1}^{G}(1/G)|p_i^g - p_j^g|/p^g$. This is similar to the first price dispersion measure discussed above. Applying it to data for thirteen European countries from the mid-1980s, they find distance to be highly significant, with an R^2 of 0.53.

Crucini et al. (2001) consider the relationship between price dispersion and goods characteristics. They do not particularly focus on trade costs, but one of their results is closely related to trade costs. They find a negative relationship between good-specific price dispersion (measure 2 above) and the level of trade for each good. The latter is defined as aggregate trade among the countries in the sample divided by aggregate output. Since more trade is associated with lower trade barriers, this result is consistent with a positive relationship between price dispersion and trade barriers.

While the results from this literature do not yet reveal much about the magnitude of actual trade barriers, they suggest that exploiting the variation of price dispersion across goods and location pairs is a natural direction to go to learn about the size of trade barriers. This needs to be done in the context of a trade model that incorporates all major sources of price differentials: international trade costs, domestic trade costs, taxes and markups.

Which theoretical approach to adopt depends on the nature of the survey data. There are three different types of data. The first category gives us information about prices of a particular brand (particular make and model) in different locations and also tells us the location of the producer of that brand. An example of this is data on automobile prices for particular makes and models. The second type of data also gives detailed price information for a particular brand in various locations, but does not provide sufficient information that allows for identification of the producer of the brand. An example of this is the Eurostat survey data. The third type of price data is not for particular brands, but for an average of representative brands in a sector. An example of this is the OECD survey data.[69]

The first type of data is the most informative about trade costs, but detailed information of this type is quite rare. In the absence of local trade costs, markups and taxes, the trade barrier is revealed by the price relative to that in the producer's country. But of course local trade costs, markups and taxes are not zero and some modeling is still required to extract information about trade barriers. A nice illustration is Goldberg and Verboven (2001), who develop a model for the European car market. They only consider one type of trade barrier, quotas for Japanese cars, but their approach could be used more broadly to estimate trade barriers. The price paid by consumers is equal to marginal cost times a wholesale price markup, times an exogenous dealer markup, times the gross value added tax rate, times trade costs associated with quotas. The marginal cost is defined to include the cost of local distribution and

69 Bradford and Lawrence (2003) provide a description of the aggregation procedures adopted by the OECD.

marketing in the destination market, which depends on the wage rate in the destination market. The markup charged by producers is endogenous and depends on both the supply and demand side of the model. Estimation of the parameters of the model, together with data on taxes and exogenous dealer markups, then yields an estimate of trade costs associated with quota.

Next consider the second type of data, where we do not know the location of the producer. One can of course make an informed guess about who the producer is. Otherwise we need a general equilibrium model to tell us which country is the likely producer of a product. As an illustration, consider again the Eaton and Kortum (1992) model. Production costs for each country are not known exactly, only the distribution from which these costs are drawn. For each good in country i we know the probabilities that it is sourced from various countries. The distribution from which the set of relative prices is drawn is therefore known and can be used to estimate trade barriers after assuming a specific trade cost function. For realism the Eaton and Kortum model would need to be extended to allow for local distribution costs, taxes, and markups.

Finally consider the case where the price data are not for individual brands, but for an average of prices of various brands of a category of consumption. This may actually make it easier to extract information about trade costs. In this case we can straightforwardly apply the gravity model without having to make any specific assumptions about the production structure. Simply equating aggregate demand and aggregate supply, assuming a constant elasticity demand structure, yields (6) and (7):

$$\left(\Pi_i^k\right)^{1-\sigma_k} = \sum_j \left(\frac{t_{ij}^k}{P_j^k}\right)^{1-\sigma_k} \frac{E_j^k}{Y^k}$$

$$\left(P_j^k\right)^{1-\sigma_k} = \sum_i \left(\frac{t_{ij}^k}{\Pi_i^k}\right)^{1-\sigma_k} \frac{Y_i^k}{Y^k}.$$

Theory implies that the equilibrium price indices P_i^k for consumption category (or sector) k are an implicit function of aggregate demand and supply in every country and all bilateral trade barriers. After adopting a specific trade cost function, the equilibrium price indices will depend on observables and parameters of the trade cost function that need to be estimated. If we interpret the price data for consumption categories as the P_i^k, plus measurement error, the trade cost parameters can be estimated with nonlinear least squares.[70]

In applying this approach some other realistic features need to be added. First, one needs to allow for tax differences across countries and goods. Second, local trade costs t_{ii} are probably better modeled as a function of both local wages and internal distance. Direct information on local distribution costs may be used as well. Third, as discussed in section 3, introducing fixed costs makes it possible to capture the fact that brands from only a limited number of suppliers are purchased by a country. Using information about the set of suppliers to each country will then make it possible to solve for the price indices P_i^k from (30) and (31).

5. *Conclusion*

Trade costs are large when broadly defined to include all costs involved in getting a good from producer to final user. Both international trade costs and local distribution costs are very large and together dominate the marginal cost of production. Trade costs also vary widely across countries. On average, developing countries have significantly larger trade costs, by a factor of two or more in some important categories. Trade costs also vary widely across product lines, by factors of as much as ten or more.

[70] It is possible that the average price data are not representative of the various brands consumed and therefore not a good proxy for P_i^k. An alternative approach is to use price data for individual brands from the second type of data and aggregate those up to the price indices .

The patterns of variation make some economic sense, but we think more sense can still be extracted.

Better measurement of trade costs is highly desirable. The quality of the existing measures is low and can be improved. Direct measures of policy barriers are scandalously difficult to find and to use, considering the importance of trade policy in overall international policy making and to potential welfare-improving changes. Transport-cost data could relatively easily be improved greatly. Estimates based on the structural gravity models can also be improved. Extensions of existing gravity models, better treatment of aggregation and endogeneity problems, and better estimates of substitution elasticities are all likely to improve our understanding of trade costs.

Our focus is on measuring the costs, only glancingly with their explanation. Ultimately, the profession should proceed to explanation. There is undoubtedly a rich relationship between domestic and international trade costs, market structure, and political economy. Some trade costs provide benefits, and it is likely that the pursuit of benefits partly explains the costs.

References

Alesina, Alberto; Robert J. Barro and Silvana Tenreyro. 2002. "Optimum Currency Unions," NBER Macroeconomics Annual 2002.

Anderson, James E. 1979. "A Theoretical Foundation for the Gravity Equation," *Amer. Econ. Rev.* 69:1. pp. 106–16.

———. 1994. "The Theory of Protection," in *Surveys in International Trade.* David Greenaway and L. Alan Winters, eds. Oxford: Blackwell, pp. 107–38.

———. 1998. "Effective Protection Redux," *J. Int. Econ.* 44:1, pp. 21–44.

———. 2000. "Why Do Nations Trade (So Little)?" *Pacific Econ. Rev.* 5:2, pp. 115–34.

Anderson, James E. and Douglas Marcouiller. 2002. "Insecurity and the Pattern of Trade: An Empirical Investigation," *Rev. Econ. Statist.* 84:2, pp. 345–52.

Anderson, James E. and J. Peter Neary. 1994. "The Trade Restrictiveness of the Multi-Fibre Arrangement," *World Bank Econ. Rev.* 8:2, pp. 171–89.

———. 1996. "A New Approach to Evaluating Trade Policy," *Rev. Econ. Stud.* 63:1, pp. 107–25.

———. 2003. "The Mercantilist Index of Trade Policy," *Int. Econ. Rev.* 44:2, pp. 627–49.

Anderson, James E. and Eric van Wincoop. 2002. "Borders, Trade and Welfare," in *Brookings Trade Forum 2001*. Susan Collins and Dani Rodrik, eds. Washington: Brookings Institution, pp. 207–44.

———. 2003. "Gravity with Gravitas: A Solution to the Border Puzzle," *Amer. Econ. Rev.* 93:1, pp. 170–92.

Anderson, Michael A. and Stephen L.S. Smith. 1999a. "Canadian Provinces in World Trade: Engagement and Detachment," *Can. J. Econ.* 32:1. pp. 23–37.

———. 1999b. "Do National Borders Really Matter? Canada-US Regional Trade Reconsidered," *Rev. Int. Econ.* 7:2. pp. 219–27.

Bacchetta, Philippe and Eric van Wincoop. 2003. "Why Do Consumer Prices React Less than Import Prices to Exchange Rates?" *J. Europ. Econ. Assoc.* forthcoming.

Baier, Scott L. and Jeffrey H. Bergstrand. 2001. "The Growth of World Trade: Tariffs, Transport Costs and Income Similarity," *J. Int. Econ.* 53:1, pp. 1–27.

———. 2003. "Endogenous Free Trade Agreements and the Gravity Equation," Notre Dame U.

Balistreri, Edward J. and Russell H. Hillberry. 2002. "Trade Frictions and Welfare in the Gravity Model: How Much of the Iceberg Melts?" work. pap., US Int. Trade Commission.

Barro, Robert J. and Silvana Tenreyro. 2002. "Economic Effects of Currency Unions," work. pap. Harvard U.

Bergstrand, Jeffrey H. 1985. "The Gravity Equation in International Trade: Some Microeconomic Foundations and Empirical Evidence," *Rev. Econ. Statist.* 67:3. pp. 474–81.

———. 1989. "The Generalized Gravity Equation, Monopolistic Competition, and the Factor-Proportions Theory in International Trade," *Rev. Econ. Statist.* 71:1. pp. 143–53.

———. 1990. "The Heckscher-Ohlin-Samuelson Model, the Linder Hypothesis and the Determinants of Bilateral Intra-Industry Trade," *Econ. J.* 100:403, pp. 1216–29.

Bernard, Andrew B. and J. Bradford Jensen. 2003. "Why Some Firms Export," *Rev. Econ. Statist.* forthcoming.

Bernard, Andrew B. and Joachim Wagner. 2001. "Export Entry and Exit by German Firms," *Weltwirtschaftliches Archiv* 137:2. pp. 105–23.

Bradford, Scott C. and Robert Z. Lawrence. 2003. "Paying the Price: The Cost of Fragmented International Markets," Inst. Int. Econ.:Washington DC, forthcoming.

Broda, Christian and John Romalis. 2003. "Identifying the Relationship between Trade and Exchange Rate Volatility," work. pap., U. Chicago GSB.

Brown W. Mark. 2003. "Overcoming Distance, Overcoming Borders: Comparing North American Regional Trade," research pap. 8, Micro-Economic Analysis Division, Statistics Canada.

Brown, W. Mark and William P. Anderson. 2002. "Spatial Markets and the Potential for Economic Integration between Canadian and US Regions," *Pap. Reg. Sci.* 81:1, pp. 99–120.

Brun, Jean-Francois; Celine Carrere, Patrick Guillaumont and Jaime de Melo. 2002. "Has Distance Died?" CEPR discus. pap. 3500.

Burstein, Ariel; Joao C. Neves and Sergio Rebelo. 2003. "Distribution Costs and Real Exchange Rate Dynamics During Exchange-Rate-Based Stabilizations," *J. Monet. Econ.* 50:6, pp. 1189–214.

Chen, Natalie A. 2002. "Intra-national versus International Trade in the European Union: Why Do National Borders Matter?" CEPR Discus. Pap. 3402.

Coe, David T.; Arvind Subramanian and Natalia Tamirisa. 2002. "The Missing Globalization Puzzle," work pap. IMF.

Coulombe, Serge. 2002. "Border Effects and North American Economic Integration," work. pap. U. Ottawa.

Crucini, Mario J. and Mototsugu Shintani. 2002. "Persistence in Law-of-One Price Deviations: Evidence from Micro-Data," work. pap. 02-W22, Vanderbilt U.

Crucini, Mario J.; Chris I. Telmer and Marios Zachariadis. 2000. "Dispersion in Real Exchange Rates," work pap. Vanderbilt U.

———. 2001. "Understanding European Real Exchange Rates," work pap. Vanderbilt U.

Davis, Donald. 1998. "The Home Market Effect, Trade and Industrial Structure," *Amer. Econ. Rev.* 88:5, pp. 1264–76.

Deardorff, Alan V. 1979. "One-Way Arbitrage and Its Implications for the Foreign Exchange Markets," *J. Polit. Econ.* 87:2, pp. 351–64.

———. 1998. "Determinants of Bilateral Trade: Does Gravity Work in a Neoclassical World?" in *The Regionalization of the World Economy*. J.A. Frankel, ed. Chicago: U. Chicago Press, pp. 7–32.

Deardorff, Alan V. and Robert M. Stern. 1998. *Measurement of Nontariff Barriers*. Ann Arbor: U. Michigan Press.

Eaton, Jonathan and Samual Kortum. 2002. "Technology, Geography and Trade," *Econometrica* 70:5, pp. 1741–79.

Engel, Charles. 2002. "Comment on Anderson and van Wincoop" in *Brookings Trade Forum 2001*. Susan Collins and Dani Rodrik, eds. Washington: Brookings Inst.

Engel, Charles and John H. Rogers 1996. "How Wide Is the Border?" *Amer. Econ. Rev.* 86:5, pp. 1112–25.

———. 2000. "Relative Price Volatility: What Role Does the Border Play?" in *Intranational Macroeconomics*. Gregory Hess and Eric van Wincoop, eds. Cambridge: Cambridge U. Press, pp. 92–111.

.———. 2001. "Deviations from Purchasing Power Parity: Causes and Welfare Costs," *J. Int. Econ.* 55:1, pp. 29–57.

Estevadeordal, Antoni; Brian Frantz and Alan M. Taylor. 2003. "The Rise and Fall of World Trade, 1870–1939," *Quart. J. Econ.* 68:2, pp. 359–408.

Evans, Carolyn L. 2001. "The Costs of Outsourcing," work. pap. 2001b. Board Gov. Fed. Reserve System.

———. 2003a. "The Economic Significance of National Border Effects," *Amer. Econ. Rev.* 93:4. pp. 1291–312.

———. 2003b. "Home Bias in Trade: Location or Foreign-ness?" work. pap. 2001a, Board Gov. Fed. Reserve System.

———. 2003c. "Border Effects and the Availability of Domestic Products Abroad," work. pap. Board Gov. Fed. Reserve System.

Evenett, Simon J. and Anthony Venables. 2002. " Export Growth in Developing Countries: Market Entry and Bilateral Trade Flows," work pap., London School Econ.

Feenstra, Robert C. 1994. "New Product Varieties and the Measurement of International Prices," *Amer. Econ. Rev.* 84:1. pp. 157–77.

———. 1998. "Integration of Trade and Disintegration of Production in the Global Economy," *J. Econ. Perspect.* 12:4, pp. 31–50.

———. 2002. "The Gravity Equation in International Economics: Theory and Evidence," *Scot. J. Polit. Econ.* 49:5, pp. 491–506.

———. 2003. *Advanced International Trade: Theory and Evidence*, Princeton: Princeton U. Press.

Frankel, Jeffrey, Ernesto Stein and Shang-jin Wei. 1998. "Continental Trading Blocs: Are They Natural or Supernatural?" in *The Regionalization of the World Economy*. J.A. Frankel, ed. Chicago: U. Chicago Press, pp. 91–113.

Frankel, Jeffrey A. and Shang-Jin Wei. 1997. "ASEAN in a Regional Perspective," in *Macroeconomic Issues Facing ASEAN Countries*. J. Hicklin, D. Robinson and A. Singh, eds. Washington, DC: IMF.

Froot, Kenneth and Kenneth Rogoff. 1995. "Perspectives on PPP and Long-Run Real Exchange Rates," in *Handbook of International Economics*, vol. 3. Gene Grossman and Kenneth Rogoff, eds. Amsterdam: Elsevier, pp. 1645–729.

Ghosh, Atish and Holger C. Wolf. 1994. "Pricing in International Markets: Lessons from the Economist," NBER work. pap. 4806.

Glick, Reuven and Andrew K. Rose. 2002. "Does a Currency Union Affect Trade? Time Series Evidence," *Europ. Econ. Rev.* 46:6, pp. 1125–51.

Goldberg, Pinelopi Koujianou and Michael M. Knetter. 1997. "Goods Prices and Exchange Rates: What Have We Learned?" *J. Econ. Lit.* 35:3, pp. 1243–72.

Goldberg, Pinelopi Koujianou and Frank Verboven. 2001. "The Evolution of Price Dispersion in European Car Markets," *Rev. Econ. Stud.* 68:4, pp. 811–48.

Gould, David M. 1994. "Immigrant Links to the Home Country: Empirical Implications for US Bilateral Trade Flows," *Rev. Econ. Statist.* 76:2, pp.302–16.

Grossman, Gene. 1998. "Comment on Deardorff," in *The Regionalization of the World Economy*. J.A. Frankel, ed. Chicago: U. Chicago Press, pp. 33–57.

Harrigan, James. 1993. "OECD Imports and Trade Barriers in 1983," *J. Int. Econ.* 34:1-2, pp. 91–111.

———. 2002. "Specialization and the Volume of Trade: Do the Data Obey the Laws?" in *Handbook of International Trade*. Kwan Choi and James Harrigan, eds. London: Basil Blackwell.

Haveman, Jon and David Hummels. 2004. "Alternative Hypotheses and the Volume of Trade: Evidence on the Extent of Specialization," *Can. J. Econ.* forthcoming.

Head, Keith and Thierry Mayer. 2000. "Non-Europe: The Magnitude and Causes of Market Fragmentation in the EU," *Weltwirtschaftliches Archiv* 136:2, pp. 284–314.

———. 2001. "Illusory Border Effects: Distance Mismeasurement Inflates Estimates of Home Bias in Trade," work. pap. U. British Columbia.

Head, Keith and John Ries. 1998. "Immigration and Trade Creation: Evidence from Canada," *Can. J. Econ.* 31:1, pp. 47–62.

———. 2001. "Increasing Returns vs National Product Differentiation as an Explanation for the Pattern of US-Canada Trade," *Amer. Econ. Rev.* 91:4. pp. 858–76.

Heckscher, E.F. 1916. "Växelkursens Grundval vid Pappersmyntfot," *Ekonomisk Tidskrift* 18:Oct. pp. 309–12.

Helliwell, John F. 1996. "Do National Boundaries Matter for Quebec's Trade?" *Can. J. Econ.* 29:3, pp. 507–22.

———. 1997. "National Borders, Trade and Migration," *Pacific Econ. Rev.* 2:3. pp. 165–85.

———. 1998. *How Much Do National Borders Matter?* Washington DC: Brookings Inst.

———. 1999. "Perspectives on North American Free Trade: Effects of the Canada-US Free Trade Agreement on Interprovincial Trade," Industry Canada Research Publications Program, pap.5.

———. 2003. "Border Effects: Assessing Their Implications for Canadian Policy in a North American Context," work pap., U. British Columbia.

Helliwell, John F. and Frank C. Lee. 1999. "Effects of the Canada-United States Free Trade Agreement on Interprovincial Trade," Industry Canada Research Publication Program, pap. 5.

Helliwell, John F. and John McCallum. 1995. "National Borders Still Matter for Trade," *Policy Options* 16, pp. 44–48.

Helliwell, John F. and Genevieve Verdier. 2001. "Measuring Internal Trade Distances: A New Method Applied to Estimate Provincial Border Effects in Canada," *Can. J. Econ.* 34:4. pp. 1024–41.

Hillberry, Russell. 1998. "Regional Trade and the Medicine Line: The National Border Effect in US Commodity Flow Data," *J. Borderlands Stud.* 8:2, pp. 1–17.

———. 2002. "Aggregation Bias, Compositional Change and the Border Effect," *Can. J. Econ.* 35:3, pp. 517–30.

Hillberry, Russell and David Hummels. 2002a. "Intra-National Home Bias: Some Explanations," NBER work. pap. 9022, *Rev. Econ. Statist.* forthcoming.

———. 2002b. "Explaining Home Bias in Consumption: The Role of Intermediate Input Trade," work. pap. 9020, NBER Cambridge, MA.

Hufbauer, Gary and Kimberly Ann Elliott. 1994. *the Cost of Protection in the United States.* Washington: Institute Int. Econ.

Hummels, David. 1999. "Have International Transportation Costs Declined?" work pap., Purdue U.

———. 2001a. "Toward a Geography of Trade Costs," work pap., Purdue U.

———. 2001b. "Time as a Trade Barrier," work pap., Purdue U.

Hummels, David; Jun Ishii and Kei-Mu Yi. 2001. "The Nature and Growth of Vertical Specialization in World Trade," *J. Int. Econ.* 54:1, pp. 75–96.

Kelejian, H.H. and Ingmar R. Prucha. 1999. "A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model," *Int. Econ. Rev.* 40:2, pp. 509–33.

Kim, Namdoo. 1996. *Measuring the Costs of Visible Protection in Korea.* Washington: Institute Int. Econ.

Klenow, Peter J. and Andres Rodriguez-Clare. 1997. "Quantifying Variety Gains from Trade Liberalization," work. pap. Stanford U.

Knetter, Michael. 1997."Why Are Retail Prices in Japan So High? Evidence from German Export Prices," *Int. J. Ind. Organ.* Aug., pp. 549–72.

Lai, Huiwen and Daniel Trefler. 2002. "The Gains from Trade with Monopolistic Competition: Specification, Estimation and Mis-specification," NBER work. pap. 9169.

Laird, Sam and Alexander Yeats. 1990. *Quantitative Methods for Trade-Barrier Analysis.* NY: NYU. Press.

Lee, Jong-Wha and Phillip Swagel. 1997. "Trade Barriers and Trade Flows Across Countries and Industries," *Rev. Econ. Statist.* 19:3, pp. 372–82.

Limao, Nuno and Anthony J. Venables. 2001. "Infrastructure, Geographical Disadvantage, Transport Costs and Trade," *World Bank Econ. Rev.* 15:3, pp. 451–79.

Lopez-Cordova, Ernesto J. and Christopher M. Meissner. 2003. "Exchange Rate Regimes and International Trade: Evidence from the Classical Gold Standard Era," *Amer. Econ. Rev.* 93:1. pp. 344–53.

McCallum, John. 1995. "National Borders Matter: Canada-US Regional Trade Patterns," *Amer. Econ. Rev.* 85:3, pp. 615–23.

Melitz, Jacques. 2003. "Language and Foreign Trade," work pap., U. Strathclyde.

Messerlin, Patrick. 2001. *Measuring the Cost of Protection in Europe.* Washington DC: Institute Int. Econ.

Millimet, Daniel L. and Thomas Osang. 2001. "Do State Borders Matter for US Intranational Trade? The Role of History and Internal Migration," work pap., Southern Methodist U.

Minondo, Asier. 2002. "Borders and Trade in Europe: A Fresh Analysis on the Extent of the Home-Bias Effect," work pap., U. Deusto-ESTE, San Sebastian.

Nitsch, Volker. 2000. "National Borders and International Trade: Evidence from the European Union," *Can. J. Econ.* 33:4. pp. 1091–105.

Obstfeld, Maurice and Kenneth Rogoff. 2000. "The Six Major Puzzles in International Macroeconomics. Is There a Common Cause?" in *NBER Macroeconomics Annual 2000*. B.S. Bernanke and K. Rogoff, eds. Cambridge, MA: MIT Press, pp. 339–90.

Obstfeld, Maurice and Alan M. Taylor. 1997. "Nonlinear Aspects of Goods-Market Arbitrage and Adjustment: Heckscher's Commodity Points

Revisited," *J. Japan. Int. Econ.* 11:Dec. pp. 441–79.

O'Connel, Paul G.J. and Shang-Jin Wei. 2002. "The Bigger They Are, the Harder They Fall: Retail Price Differences across US Cities," *J. Int. Econ.* 56:1, pp. 21–53.

O'Rourke, Kevin H. and Jeffrey G. Williamson. 1999. *Globalization and History: The Evolution of a Nineteenth Century Atlantic Economy.* Cambridge: MIT Press.

Parsley, David C. and Shang-Jin Wei. 2001. "Explaining the Border Effect: The Role of Exchange Rate Variability, Shipping Costs and Geography," *J. Int. Econ.* 55:1, pp. 87–105.

———. 2002. "Currency Arrangements and Goods Market Integration: A Price Based Approach," work pap. Vanderbilt U.

Portes, Richard and Helene Rey. 2002. "The Determinants of Cross-Border Equity Transaction Flows," work pap., London Bus. School.

Rauch, James E. and Vitor Trindade. 2002. "Ethnic Chinese Networks in International Trade," *Rev. Econ. Statist.* 84:1, pp. 116–30.

Ritschl, Albrecht and Nikolaus Wolf. 2003. "Endogeneity of Currency Areas and Trade Blocs: Evidence from the Inter-War Period," work pap., Humboldt U.

Roberts, Mark and James Tybout. 1997. "The Decision to Export in Colombia: An Empirical Model of Entry with Sunk Costs," *Amer. Econ. Rev.* 87:4, pp. 545–64.

Rose, Andrew K. 2000. "One Money, One Market: Estimating the Effect of Common Currencies on Trade," *Econ. Pol.* 30, pp. 7–45.

———. 2004. "The Effect of Common Currencies on International Trade: A Meta-Analysis," in *Monetary Unions and Hard Pegs: Effects on Trade, Financial Development and Stability.* Volbert Alexander, Jacques Melitz and George M. von Furstenberg, eds. Oxford U. Press.

Rose, Andrew K. and Eric van Wincoop. 2001. "National Money as a Barrier to Trade: The Real Case for Currency Union," *Amer. Econ. Rev. Pap. Proceed.* 91:2. pp. 386–90.

Sazanami, Yoko, Shujiro Urate and Hiroki Kawai. 1995. *Measuring the Costs of Pretection in Japan.* Washington: Institute Int. Econ.

Staiger, Robert W. and Frank A. Wolak. 1994. "Measuring Industry Specific Protection: Antidumping in the United States," *Brookings Pap. Econ. Act.: Microecon.* pp. 51–118.

Taylor, Alan M. 2001. "Potential Pitfalls for the Purchasing-Power-Parity Puzzle? Sampling and Specification Biases in Mean-Reversion Tests of the Law of One Price," *Econometrica* 69:2, pp. 473–98.

Tchamourliyski, Yuriy. 2002. "Distance and Bilateral Trade: The Role of Non-Homothetic Preferences," work pap. Boston College.

Trefler, Daniel. 1993) "Trade Liberalization and the Theory of Endogenous Protection: An Econometric Study of US Import Policy," *J. Polit. Econ.* 101:1, pp. 138–60.

Wei, Shang-Jin. 1996. "Intra-national versus International Trade: How Stubborn Are Nations in Global Integration?" NBER work. pap. 5531, Cambridge, MA.

Wolf, Holger. 2000a. "(Why) Do Borders Matter For Trade?" in *Intranational Macroeconomics*. G.D. Hess and E. van Wincoop, eds. Cambridge: Cambridge U. Press, pp. 112–28.

———. 2000b. "Intranational Home Bias in Trade," *Rev. Econ. Statist.* 82:4. pp. 555–63.

Yi, Kei-Mu. 2003a. "Can Vertical Specialization Explain the Growth of World Trade?" *J. Polit. Econ.* 111:1. pp. 52–102.

———. 2003b. "A Simple Explanation for the Border Effect," work pap., Fed. Reserve Bank NY.

Zhang Shuguang; Zhang Yansheng and Wan Zhongxin. 1998. *Measuring the Costs of Protection in China.* Washington DC: Institute Int. Econ.