

Trade Costs ¹

James Anderson
Boston College
NBER

Eric van Wincoop
University of Virginia
NBER

September 10, 2003

¹In preparation for the Journal of Economic Literature. Preliminary draft; please do not quote without permission.

Abstract

This paper surveys trade costs — what we know, and what we don't know but may usefully attempt to find out. Partial and incomplete data on direct measures of costs go together with inference on implicit costs from the pattern of trade across countries. Representative margins for full trade costs in rich countries exceed 170% based on our pushing the data very hard. Poor countries face even higher trade costs. There is a lot of variation across countries and across goods within countries, much of which makes economic sense. Theory looms large in our survey, providing interpretation and perspective on the one hand and suggesting improvements for the future on the other hand. Some new results are presented to apply and interpret gravity theory properly and to handle aggregation appropriately.

I Introduction

“... *the report of my death was an exaggeration.*” Mark Twain, 1897

The death of distance is exaggerated. Trade costs are large, even in the absence of formal barriers to trade 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 representative tariff equivalent estimate discussed below is a 170% total trade barrier from foreign producer to final user in the domestic country. Trade barriers therefore dominate production costs, even though the latter have been the focus of most trade theory. (2) Obstfeld and Rogoff (2000) make a case that all the major puzzles of international macroeconomics hang on trade costs. (3) Economic geography is founded on trade costs and the details matter. 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 (Davis, 1998). (4) Reducing current trade costs has large welfare implications. For example, Anderson and van Wincoop (2002) argue that policy related costs are often worth more than 10% of national income. (5) Trade costs are richly linked to economic policy — some important trade costs are direct policy instruments (e.g., tariffs, the tariff equivalents of quotas and trade barriers associated with the exchange rate system), while others are greatly influenced by policy (transport infrastructure investment, law enforcement). (6) The cross-commodity structure of policy barriers is important — e.g., see the large literature on the theory of trade policy surveyed in Anderson (1994). This makes the inevitable aggregation of barriers a crucial issue, a topic

we develop below.

Broadly Defined

Trade costs are broadly defined to include all costs incurred in getting a good to a final user other than the production cost of the good itself. Among others this includes transportation costs (both freight costs and time costs), policy barriers (tariffs and non-tariff barriers), information costs, contract enforcement costs, costs associated with the use of different currencies, legal and regulatory costs, and local distribution costs (wholesale and retail). We do not cover the structural determinants of these trade costs except in passing. Our focus is on the prior step of measuring the costs. Ultimately, with a firmer understanding of the size and pattern of the costs, the profession can and should proceed to the explanation of the costs. There is undoubtedly a rich structure of endogeneity between various types of domestic and international trade costs, market structure and political economic structure. Some trade costs provide benefits, and it is likely that the pursuit of benefits partly explains the costs.

Three Sources

We report on trade costs from three broad sources. Direct measures of trade costs are obviously best and are discussed in section II. In an ideal world, only direct measures would be needed, complete and accurate. In the actual world, direct measures of trade costs are remarkably sparse and inaccurate. We review the best available direct evidence and make a plea for better information collection and especially for its ready provision to researchers. Two types of indirect measures complement the direct measures: inference from quantities (trade volumes), discussed in section III, and inference from prices, discussed in section IV. In both cases, economic structure is needed to infer trade costs from the data, so the quality of the inferred trade cost is dependent on the quality of the underlying economic reasoning as well as on the quality of the data.

Theory Looms Large

Theory looms large in our survey, despite its focus on the very applied topic of trade costs. 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 theoretical developments in trade theory have led to a better understanding of the link between trade barriers and trade flows. This has begun to bridge the gap between practice and theory in the inference of trade costs. We linger on the bridge in the belief that a reader who pauses with us will produce better work in the future.

The gravity model is the main work horse linking trade barriers to trade flows. Assumptions needed to derive gravity equations are not well understood, with gravity often taken to be rather atheoretic or justified only under highly restrictive assumptions. We will show that only minimal structure needs to be imposed to derive a gravity equation from theory. Gravity equations represent a conditional general equilibrium. Under a key assumption, which we will refer to as trade separability, the allocation of trade across countries is separable from the allocation of production and spending within countries. Standard trade theory is mostly concerned with the latter, the allocation of demand and supply within countries. Gravity equations develop a link between the trade allocation and trade barriers 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. This makes gravity a powerful tool for inferring information about trade barriers.

Our treatment features extensive consideration of the appropriate aggregation of trade costs. 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. The literature has paid relatively little attention to such aggregation issues. In order to stimulate more work in this direction we show how theory can be used to replace common a-theoretic 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

Trade costs are large. As an illustration consider the example of Mattel's Barbie

doll, discussed in Feenstra (1998). The production costs for the doll are \$1, while it sells for about \$10 in the United States. The main difference is in the form of transportation, marketing, wholesaling and retailing. In this example the tariff equivalent of trade barriers is 900%.

A very rough estimate of the tariff equivalent of trade costs for industrialized countries is 170%. This includes all markups over the production costs. This number breaks down as follows: 21% transportation costs, 44% border related trade barriers and 55% retail and wholesale margins ($2.7=1.21*1.44*1.55$). The first and the last category can be directly measured. The 21% transport cost includes both freight costs and a 9% tariff equivalent of the time value of goods in transit. Both are based on estimates for U.S. data. Representative retail and wholesale margins are at least 40% for both rich and poor countries. The U.S. margin is equal to 68% while more compact developed countries have lower margins (Germany has 53%, Netherlands 50%). As an illustration we have used a 55% margin, which is close to the average for industrialized countries.

Estimates from gravity equations suggest that national borders pose a barrier in the range of 25-50%. The 44% number used above is based on a mixture of direct evidence and evidence from the gravity literature. Direct evidence is available for tariff barriers and to a lesser extent for non-tariff barriers. Tariff barriers are now low in most countries, on average (trade-weighted or arithmetic) less than 5% for rich countries, and with a few exceptions are on average between 10% and 20% for developing countries. In contrast, non-tariff policy barriers are high and increasing, especially in the form of antidumping and its effects. Our overall estimate of policy barriers for industrialized countries is about 8%. Less precisely measured border costs appear on average to dwarf the effect of tariff and non-tariff policy barriers. An extremely rough breakdown of the 44% number reported above is as follows: a 8% policy barrier, a 7% language barrier, a 14% currency barrier (from the use of different currencies), a 6% information cost barrier, and a 3% security barrier. Although the breakdown is suggestive, the overall range of border barriers of 25-50% is probably more meaningful. The pure international component of trade barriers, including transport costs and border barriers but not local distribution margins, is estimated to be in the range of 40-80% for industrialized countries.

Trade costs are also highly variable, across both goods and countries. The example of Mattel's Barbie doll shows that some trade barriers are much larger than

the average 170% barrier. Others are much lower. The variability of trade costs makes statements about the average height of trade costs quite misleading unless accompanied by study of variability and some consideration of appropriate aggregation. The pattern of variability has economically intriguing meaning. In some cases it makes straight economic sense. Trade barriers in developing countries are higher than those for industrialized countries reported above. 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, acting differentially across countries. A more intriguing pattern that is not fully understood is that sectoral trade barriers appear to vary inversely with elasticities of demand. Policy barriers, particularly NTB's, also vary significantly across goods. Non-tariff barriers are highly concentrated. Coverage ratios in many sectors are close to zero, but for textiles and apparel the U.S. coverage ratio is 71% and tariff equivalents are in the range of 5% to 33%.

II 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 policy barriers first. Next, we review evidence on transport costs. Finally, we review evidence on wholesale and retail distribution costs, which provide evidence on domestic trade costs.

An important theme is the many limitations and difficulties faced in obtaining accurate measures of trade costs. Substantially better data on trade costs are quite feasible to collect and would yield a high payoff in understanding international and interregional economics and in improved policy making advice. Particularly egregious is the paucity of good data on policy barriers. 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. These may be reflected in price comparisons covered below and by inference from quantity flows covered in the next section.

The available evidence shows that trade costs are high on average but highly variable across products and countries. Tariffs are low for many developed coun-

tries, but nontariff policy barriers are pervasive in certain sectors and usually have high tariff equivalents. Transport costs are high, considerably higher than tariffs in low-tariff countries, and highly variable across goods and trading partners. Wholesale and retail margins are much larger than transport or tariff cost margins (sometimes by an order of magnitude) and are highly variable across sectors and countries.

II.A Policy Barriers

II.A.1 Measurement Problems and Limitations

Economists new to the analysis of international trade are always shocked at the poor quality of direct measures of the policy barriers to trade. Economic theory has at least since the mercantilists pointed to the importance of international trade and trade policy, so it is natural to assume that trade policy is well documented. Indeed, the grossly incomplete and inaccurate information available on policy barriers is a scandal and a puzzle.¹ Trade negotiations preoccupy substantial bureaucracies in each country and multilaterally, trade policy is a key concern of the World Bank and other development institutions, and revenue from control of trade is important for poor countries. Yet the seemingly simple question ‘how high are policy barriers to trade?’ 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 which are useful only in combination with other missing or fragmentary data, and aggregation bias. Each of the difficulties is amplified in the discussion below.

The most complete data available on policy barriers to trade is based on the United Nations Conference on Trade and Development’s Trade Analysis & Information System, TRAINS. It contains information on Trade Control Measures (tariff, para-tariff and non-tariff measures) at tariff line level for a maximum of 137 countries beginning in the late 1980’s. TRAINS reports all data on bilateral tariffs, non-tariff barriers and bilateral trade flows at the six-digit level of the Harmonized

¹ Political economy does not provide a convincing explanation for the scarcity of data. Recipients of rent from protection have an interest in obscurity, but reduction of barriers benefits other politically active groups. Why isn’t the collection of better information at low cost an equilibrium?

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 6-digit aggregate. It is in some cases possible to drill down to the national tariff line level.

The main problem with TRAINS is the substantial incompleteness. Table 1 gives a sense of this. It reports for each year from 1989 to 2000 the fraction of countries that report some lines (though possibly a very limited number) for tariffs, NTB’s and trade flows. Of 121 reporting countries in 1999, 43% report tariffs, 30% report NTB’s, 55% report trade flows and 17.4% 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% have tariff and NTB information recorded in 1999, with considerable variability in coverage across the years. The current data is available for purchase from UNCTAD, but another problem is that it customarily comes with front end software which defeats use of the data for normal social science purposes.² Another key limitation is that 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 misleading.

Another useful data source is the World Trade Organization’s trade barrier databases. The Consolidated Tariff Schedule database lists the Most Favored Nation (MFN) bound tariffs and the Initial Negotiating Rights 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 Integrated Data Base contains information on the applied rates at the national tariff line level. Neither database is public. The WTO additionally periodically reports on individual member country trade policy with a published Trade Policy Review which uses applied tariffs and other useful information. Another periodic data source for policy barriers to trade is various studies published by the Institute for International Economics which take up the considerable labor

²UNCTAD sells the data each year to commercial customers who use it to provide current information on trade costs to potential traders. The front end is 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.

of evaluating a single country's trade policy.³

Making use of the available data, 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.⁴ The World Bank has very recently put together the most comprehensive system, a huge improvement over anything available before, and in principle it is committed to making access public.⁵ Its World Integrated Trade System (WITS) software is coupled to TRAINS and to the WTO Integrated Data Base and Consolidated Tariff Schedules along with the UN Statistical Division's COMTRADE trade flow data. Essentially 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. WITS has some other data handling and modeling functions as well. The World Bank Trade and Production database is produced and published using WITS with the databases internally available at the Bank. It is published on the Bank website and presumably will be regularly updated. The needs of many users will be satisfied with this data. The Trade and Production database contains trade, production and tariff data for 67 developing and developed countries at the industry level over the period 1976-1999. 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 3 digit level (28 industries) for 67 countries and at the 4 digit level (81 industries) for 24 of these countries. The tariff data is drawn from TRAINS with supplemental information from the WTO Integrated Trade Database and Trade Policy Reviews. Importantly, the limitations of the missing specific tariff information are substantially addressed.

³See Hufbauer and Elliott (1994) on the US, Messerlin (2001) on the EU, Sazanami et al. (1995) on Japan, Kim (1996) on South Korea, and Zhang et al. (1998) on China.

⁴For 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 14 years times 137 countries.

⁵See <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 will need to pay fees to UNCTAD for use of COMTRADE and TRAINS. Email queries are not answered, in this reporter's experience, without using friends at the Bank as intermediaries.

The source of trade data in TRAINS and in the World Bank data is the United Nations Statistics Department's Comtrade database. The World Bank database is the best one stop source, but it unfortunately does not provide NTB data. For this purpose a user must use TRAINS or more specialized databases directed at particular NTB's.

Data limitations can make even the available trade barrier information difficult to interpret. Many tariffs are in the form of specific taxes (on quantity), and must be converted to ad valorem equivalents. This requires price information which must be obtained and matched up to the tariff line items. Imperfections of classification or other information introduce potential measurement error in the ad valorem equivalent. TRAINS reports the percentage of underlying lines which have specific tariffs in order to provide the user with some information about how widespread the measurement problem may be. 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.

Nontariff barriers are much more problematic than tariff barriers, both in basic information and in conceptual issues. Deardorff and Stern (1998) have an excellent and extensive treatment of the measurement of nontariff barriers. See also Laird and Yeats (1990), who in addition review many more efforts to quantify nontariff barriers. The TRAINS database records the presence or absence of a nontariff barrier (NTB) on each 6 digit line. Many differing types of nontariff barriers are recorded in TRAINS (a total of 18 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 dataset.⁶ Haveman follows what has become a customary grouping of NTB's 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 which are covered by NTB's. No information about

⁶See the Ultimate Trade Barrier Catalog at <http://www.eiit.org/Protection>.

NTB restrictiveness is provided in TRAINS; unsurprisingly, since measuring the restrictiveness of each type of nontariff barrier requires a well specified economic model. Nontariff barrier information in TRAINS is particularly prone to incomplete 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.⁷

In a limited number of cases for important nontariff barriers, individual analysts have developed direct measures of the restrictiveness of nontariff barriers. The Multi-Fibre Arrangement controlling trade in textiles and apparel is the most important quota system in the world. Exporting countries set quotas negotiated with importing countries, and in some cases the export countries permit markets in the transfer of licenses. The largest export market is in Hong Kong, where newspaper reports on license prices have been used to create data on the restrictiveness of the quota system. Less reliable license price data is available for some other markets. In combination with tariffs charged by importing countries (which serve to retain some of the quota rent which would otherwise go to the exporter), license prices can be used to construct tariff equivalents.

Indirect methods of measuring the restrictiveness of NTB's 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 IV. The restrictiveness of nontariff barriers can also be inferred from trade quantities in the context of a well specified model of trade flows. Evidence about nontariff barriers from trade flows is discussed in Section III. Deardorff and Stern (1998) and Laird and Yeats (1990) provide other detailed discussion of inference about the restrictiveness of NTB's.

Finally, aggregation and the associated bias are very important problems in the analysis of trade barriers. Tariffs and NTB's are extremely high dimensional and

⁷National 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.

exhibit large variation across product 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 which 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 beyond the 6 digit HS level 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 trade weighted percentage of component sectors with nontariff barriers.

How should aggregation be done? Anderson and Neary (1996, 2003) propose theoretically ideal trade policy aggregators based on the idea of a uniform tariff equivalent of differentiated tariffs and NTB's. Theoretically consistent aggregation depends on the purpose of the analysis, so the analyst must specify tariff equivalence with reference to an objective which makes sense for the task at hand. Anderson and Neary develop and apply indices for the small country case which are equivalent in terms of welfare and in terms of distorted aggregate trade volume,⁸ and show that atheoretic aggregation can significantly bias the measurement of trade restrictiveness. 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.

II.A.2 Evidence on Policy Barriers

Policy induced trade costs are interesting in themselves for such purposes as trade negotiations, while they also provide a benchmark against which to set the magni-

⁸See also Anderson (1998) for equivalence in terms of sector-specific factor income and its relationship to the commonly reported effective rate of protection.

tude of the non-policy trade costs reported in the other parts of this survey. This survey concentrates on current trade costs, in this section those due to tariff and nontariff barriers.⁹

Tariffs

First we present recent tariff data taken from TRAINS for 50 countries for 1999. The relatively small number of countries available reflects reporting difficulties typical of TRAINS; some earlier years contain data for more countries. Trade weighted and arithmetic average tariffs are reported in Table 2. These are the commonly used atheoretic indices of the thousands of individual tariff lines in the underlying data.¹⁰ The tariff data confirm that tariffs are low among most developed countries (under 5%), while developing countries continue to have higher tariff barriers (mostly over 10%).

The variation of tariffs across goods is quite large in all countries; typically many are small and a few are quite large.¹¹ Intuition based on the first course in economics implies that variation of tariffs adds to the welfare cost of protection. The reasoning is that the marginal deadweight loss of a tariff is proportional to the size of the tariff, hence the cumulated loss (the dead weight loss triangle) varies with the square of the tariff. The details of a full welfare analysis are complex and qualify this insight but there is something right about it. In light of the intuitive idea about the cost of variation of tariffs, it is fairly common to compute coefficients of variation (the ratio of the standard deviation to the mean) of tariffs, either arithmetic or trade-weighted. Anderson and Neary (2003) show that the

⁹See the work of Williamson and co-authors (e.g., Williamson and O'Rourke, 1999) for historical evidence.

¹⁰Anderson 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% for the trade weighted average and 12% for the uniform volume equivalent, while the US numbers in 1990 are 4% and 4.8% 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.

¹¹Political economy elements explain a good bit of the variation. See for example Rodrik (1995).

welfare cost is increasing in an appropriately weighted coefficient of variation with intuitive weights. They report coefficients of variation of tariffs for 25 countries around the year 1990, ranging from 0.14 to 1.67, many being clustered around 1.

Bilateral variation of tariffs can be rather large. Bilateral variation of tariffs is based on preferential trade at the most basic level (on each individual tariff line, a free trade agreement partner faces a zero tariff while all others face the MFN tariff), but aggregation over goods induces further bilateral variation due to the differing composition of trade across partners. Harrigan (1993) developed data on bilateral production-weighted average tariffs in 28 product categories for OECD countries for 1983, reporting the averages over commodities in his paper and making the full dataset available. For Canada, the reported range of bilateral averages runs from 1.2% to 3.2% and for Japan from 2.3% to 4.5%. The U.S. has more modest differences, from 1.6% to 2.3%. Lai and Trefler (2002) are notable for compiling 3-digit bilateral tariffs for 14 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.

Nontariff Barriers

Next we turn to nontariff barriers. Table 3 gives data on the prevalence of nontariff barriers for 34 countries in 1999 based on data from TRAINS and Haveman's extensive work rendering it more usable. The smaller number of countries available than for tariffs reflects previously noted reporting difficulties with TRAINS. We report arithmetic and trade weighted NTB coverage ratios; the percentage of tariff lines subject to NTB's. 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% while the trade-weighted NTB coverage is 5.5%. This reflects the fact that NTB's tend to fall on important traded goods such as textiles and apparel.

The NTB coverage ratio is calculated in two different ways, narrow (basically price and quantity control measures and quality control measures) and broad (the narrow classification plus threat measures related to antidumping) . Under the narrow definition, NTB's cover less than 10% of trade for the rich countries and modest amounts of trade in most except for Argentina and Brazil. The broader definition includes operating under the threat of an ongoing antidumping action.

Threats are significant, since Staiger and Wolak (1994) have shown that the threat of antidumping impedes trade considerably (on average for the U.S. they find a 17% reduction in trade due to an ongoing investigation). Table 3 implies that antidumping is quite common. For example, the U.S. has 1.5% of tariff lines subject to NTB without antidumping but 27.2% subject to NTB when antidumping is included.

The use of NTB's is concentrated in certain sectors in most economies. To give a sense of this, Table 4 reports sectoral NTB coverage ratios for the U.S., E.U., Japan and Canada for 1999. NTB's are widely used by developed countries in food products (for example, the trade-weighted NTB coverage of Agriculture, Forestry and Fishery Products is 74% for the U.S. and 24% for the E.U.), textiles/apparel (71% for the U.S. and 42% for the E.U.), wood and wood products (39% for the U.S. and 26% for the E.U.) 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.

The differential and discriminatory nature of NTB's suggests that tariff equivalents will be highly differentiated. Where markets subject to quota 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 which are found on most quota-constrained products provide a useful measure of the restrictiveness of quotas.

The limited information directly available on the restrictiveness of quotas shows that some barriers are quite high. Deardorff and Stern (1998) survey most of what limited data is available on quota license prices. Particularly interesting is the Multi-Fibre Arrangement, the most important quota system in the world. Table 5 (based on Table 3.6 of Deardorff and Stern, 1998) gives estimates of tariff equivalents based on annual averages of weekly Hong Kong license prices (which are themselves averages of transactions within the week) for textiles and apparel subject to quota between controlled exporters and the U.S. 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 tariff equivalents using Hong Kong export prices for the underlying textile and apparel items, trade weighted across suppliers.¹² To the license prices are added the U.S. tariffs on the corresponding items to form the full tariff equivalent (which is of course split between importer and exporter). The table shows fairly high tariff equivalents, especially in the largest trade categories (23% for products of Broadwoven fabric mills, 33% for Apparel made from purchased material). There is also high variability of license prices and tariff equivalents across commodities (from 5% to 23% for textiles, from 5% to 33% for apparel). Earlier years reveal higher tariff equivalents. Anderson and Neary (1994) report trade weighted average tariff equivalents across the MFA commodity groups for a set of exporters to the U.S. in the mid-1980's. The Hong Kong average exceeded 19% in each year and ranged to over 30% in some years, while tariff equivalents (very likely biased upward) for other countries were much larger, some over 100%. The evidence demonstrates fairly convincingly that there is (i) substantial restrictiveness of MFA quotas and (ii) very large differentials in quota premia across commodity lines and across exporters. In comparison to tariffs, NTB's are concentrated in a smaller number of sectors and in those sectors they are much more restrictive.

Price comparison measures confirm this picture of the high and highly concentrated nature of NTB's 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 IV.

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

¹²Because 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.

protection rate in industrial goods was 7.7%.

II.B Transport Costs

Transport costs include direct and indirect elements. Direct elements include freight charges and insurance on shipments which is customarily added into the freight charge data. Indirect costs include those incurred by the transport user which vary with the shipment's characteristics, such as 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 with the aid of an explicit economic model.

II.B.1 Measurement Problems and Limitations

There are three main sources of data for transport costs. Most directly, industry or shipping firm information can be used. For example, Limao and Venables (2001) obtain quotes from shipping firms for a standard container shipped from Baltimore to various destinations. 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 unit values in transport costs. National customs data in some cases allow quite detailed unit values to be constructed. For example, the U.S. Census provides data on U.S. imports at the 10 digit Harmonized System level by exporter country, mode of transport and entry port, valued at f.o.b. and c.i.f. bases. Dividing the former into the latter yields a unit value estimate of bilateral transport cost. Hummels (2001a) makes use of this source for the U.S. and several other countries. The most widely available (many countries and years are covered) but least satisfactory unit value data 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.). The IMF draws its data from the UN's Comtrade database, supplemented in some cases with national data sources, and reports it in the *Direction of Trade Statistics* and the *International Financial Statistics*. Hummels (2001a) points out that a high proportion of observations are

imputed; this and the compositional shifts which occur over time in unit value data based on aggregate trade flows lead him to conclude that “quality problems should disqualify these data from use as a measure of transportation costs in even semi-careful studies.” Nevertheless, because of their availability and the difficulty of obtaining better estimates for a wide range of countries and years, even recent careful empirical work such as Harrigan (1993) and Baier and Bergstrand (2001) uses the IMF data.

II.B.2 Evidence on Transport Costs

Hummels (2001a,b;1999) provides the most detailed estimates on direct costs. Hummels (2001a) shows the wide dispersion in freight rates over commodities and across countries in 1994 based on use of national customs data. The all-commodities trade weighted average ranges from 3.8% for the U.S. to 13.3% for Paraguay. The all-commodities arithmetic average ranges from 7.3% for Uruguay to 17.5% for Brazil. The U.S. average is 10.7%. Across commodities for the U.S. the range of trade weighted averages is from less than 1% (for transport equipment) to 27% for crude fertilizer. The arithmetic averages range from 5.7% for machinery and transport equipment to 15.7% for mineral fuels.

Hummels (1999) considers variation over time. The overall trade-weighted average transport cost for the U.S. declined over the last 30 years, from 6% to 4%. Composition problems are acute in assessing average transport costs over time because world trade in high-value-to-weight manufactures has grown much faster than trade in low-value-to-weight primary products, as Hummels notes. He 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 40 years.

Notice that alongside tariffs and NTB’s, 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. Thus policy protection tends to complement natural protection, amplifying the variability of total trade costs.

Limao and Venables (2001) emphasize the role that infrastructure plays in al-

tering costs at each end of trade. 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%. The median landlocked country has transport costs which are 55% higher than the median coastal economy. (Limao and Venables also report similar results using the c.i.f./f.o.b. ratios of the IMF.) The infrastructure variables, not surprisingly, also have much explanatory power in predicting trade volume. It is hard to avoid the conclusion that understanding trade costs and their role in determining international trade volumes must incorporate the internal geography of countries and the associated interior trade costs. This is particularly true for large developing economies but is likely to be so for spatially large developed economies. The simplicity of treating countries as dimensionless points is very appealing but no longer tenable in the context of trade costs. Since some high level of geographic aggregation is inevitable, future work should consider appropriate aggregation.

As for indirect costs, Hummels (2001b) provides estimates of the time cost of shipments. His method allows imputation of a willingness-to-pay for saved time. He finds that for manufactured goods each day in travel is on average worth 0.8 percent of the value of the good, equivalent to a 16% ad valorem tariff for the average length ocean shipment. Modal choice is the key feature of his work, with shippers switching from ocean to air when the full (shipping plus time) cost of ocean exceeds that for air.¹³ 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 must also have an effect on the time costs of trade, likely a very considerable one.

In 1998, half of U.S. shipments was done by air and half by boat. This ignores trucking and rail modes, which are important to trade with Canada and Mexico, the two largest trade partners of the U.S.. Assigning 1 day to shipment by air

¹³Linking port of entry for US 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.

anywhere in the world, as Hummels does, and using the 20-day average for ocean shipping, leads to an average 9% tariff equivalent of time costs. Hummels argues that faster transport (shifting from shipping to air, and faster ships) has reduced the tariff equivalent of time costs for the U.S. from 32% to 9% over the period 1950-1998.¹⁴

II.C Wholesale and Retail Margins

Wholesale and retail margins add large location-specific components to trade costs in the narrow sense which we have associated with the movement of goods over distance and across jurisdictional borders. In contrast to international trade costs, these local trade costs are unavoidable. They apply to the purchase of both foreign and domestic goods as all goods need to go through the local distribution system before reaching the final user. Local trade costs therefore do not affect trade, as we will show more formally in section III. They do however have a big impact on prices, as we discuss in section IV.

A basic source of data on wholesale and retail margins is national input-output accounts. Burstein et al. (2001) construct domestic margins for tradable consumption goods (which correspond most closely to the goods for which trade costs are also relevant), reporting a weighted average of 41.9% for the U.S. in 1992 based on individual input-output commodity margins ranging from 0 to 64.2%. These numbers are as a fraction of retail prices. They also show that for the U.S. the input-output margins 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. The latter source yields combined wholesale and retail margins on all goods equal to 46.2%. A similar calculation from Argentina's 1993 Census of Wholesale and Retail Trade yields a combined margin of 61%. This a plausible finding since developing countries have inferior infrastructure, which is

¹⁴The calculation is based on observing that US imports, excluding Canada and Mexico, had 0% air shipment in 1950 and 50% air shipment in 1998. The average ocean shipment time was halved from 40 days to 20 days over the same time. The net effect is a saving of 29.5 days, equal to 40 days for 1950 minus 10.5 days for 1998. The latter is equal to .5 times 20 days for ocean shipping plus .5 times 1 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% to 9% ad valorem.

likely to affect wholesale margins in particular.

Bradford and Lawrence (2003) use the same input-output method to measure distribution margins for the U.S. and eight other countries for which the data is available (Australia, Canada, Japan; Belgium, Germany, Italy, Netherlands, UK). Table 6 reports the ratio of consumer to producer prices for selected tradable household consumption goods and an arithmetic average for 125 goods.¹⁵ The averages range over countries from 42% in Belgium to 70% in Japan. The U.S. margin is 68%. This is higher than reported margins for the U.S. by Burstein et al. (2001), but they divide by the retail price rather than the producer price. We will report all trade barriers in this survey in terms of ad valorem tariff equivalents, which implies division by producer prices as in Table 6. The range of margins is much larger across goods than across countries, for example running from 14% on Electronic Equipment to 216% on Ladies Clothing in the U.S..

Some of these margins include profits rather than value added. Only the value added component should be counted as part of local trade costs. Although no information is available on this breakdown, one would expect that the bulk of the reported margins are value added because of standard competition arguments.

II.D An Impressionistic Total

We now piece together a simple average trade cost from the mass of previously presented data, each datum being itself a highly qualified average of underlying elements. For the representative developed country we begin with Messerlin's estimate for the E.U. of a 7.7% arithmetically weighted average of tariffs and tariff equivalents of various NTB's. NTB's are more important in the U.S. but its tariff average is lower. We take the E.U. number as representative for developed countries; no correspondingly ambitious figure exists for the U.S.. With this we combine Hummels' average 10.7% arithmetic average transport cost for the U.S. (Hummels does not report any E.U. data). Add Hummels' estimate of the average value of time of 9%. The total cost for trade policy and transport costs is 30% ($1.3=1.077*1.107*1.09$), used in the introduction to create an impression. We

¹⁵The reported numbers include local wholesale and retail margins, plus local transportation costs, plus taxes levied on producers. The latter two categories are only as small component of the total though.

concede that this is pushing the available data very hard, but we think the picture is not likely to be grossly in error. Trade costs are large even for rich countries.

III Inference of Trade Costs from Trade Flows

Trade costs can be inferred from trade flows in the context of an economic model which explains trade in terms of observable exogenous variables and the unobservable trade cost, itself a function of a set of observable variables. The effect of the trade cost proxy on trade flows having been estimated, the analyst can calculate a tariff (or other trade cost) equivalent using an elasticity of import demand when one is available, preferably estimated in the study.

The main device that has been used to link trade barriers to trade volume is the gravity model. A subset of the gravity literature uses it to discriminate among theories of the determinants of trade, aside from trade costs. It is not very well suited for this purpose, since many trade models will lead to gravity (Deardorff, 1998). We develop the theory of the gravity model below rather extensively, revealing some new and useful general equilibrium properties which clarify procedures for good empirical work, reporting results and doing sensible comparative statics. Even more significantly, our development may help link gravity to the classic concern of trade economists with the determinants of trade based on the equilibrium allocation of production and expenditure within nations.

The literature has used a variety of ad hoc trade cost functions to relate the unobservable cost to observable variables, depending on the purposes of the analysis. This is quite useful since many trade costs cannot be measured directly — some examples are trade costs associated with different languages, different cultures and customs, different institutions, regulations and product standards, asymmetric information, taste bias towards domestic goods, and contract enforcement problems. The plausible hypothesis is that cost falls with common language and customs, better information, better enforcement and so forth. Policy experiments are feasible with this approach, such as computing the impact of customs unions or monetary unions on trade volume. A small subset of the literature computes the tariff equivalent of trade costs associated with the various frictions. But since the trade cost function is a reduced form, it is limited in some of its natural uses. For example, although one can obtain estimates of trade barriers associated with national

borders, it is much harder to pin down exactly what factors generate these border related barriers.

Further economic theory is needed to identify the underlying structure of trade costs. Many implementations of the ad hoc trade cost function impose restrictions which seem quite implausible, as we argue below. In contrast, one could in principle compute all bilateral trade barriers for a group of countries or regions from all the bilateral trade data. This would have the disadvantage of attributing to trade costs all possible deviations of trade flows from the prediction of the frictionless model. Statistically speaking there is an unboundedly large confidence interval about the point estimates because all degrees of freedom are used up.

III.A Traditional Gravity

Most commonly estimated gravity equations take the form

$$x_{ij} = \alpha_1 y_i + \alpha_2 y_j + \sum_{m=1}^M \beta_m \ln(z_{ij}^m) + \epsilon_{ij} \quad (1)$$

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 ($m = 1, \dots, M$) is a set of observables to which bilateral trade barriers are related. A particularly 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.¹⁶ 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 1 if the two regions are located in the same country and equal to zero otherwise. He found that trade between provinces is more than 20 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. Until recently it was common to simply refer to some standard gravity theory references, such as Anderson (1979) and Deardorff (1998), to justify estimation of this gravity equation.

¹⁶See Anderson and Smith (1999a,b), Chen (2002), Evans (2000, 2001a,2003), Helliwell (1996,1997,1998), Helliwell and McCallum (1995), Helliwell and Verdier (2000), Hillberry (1998, 2002), Nitsh (2000), Wei (1996), and Wolf (2000a,b).

III.B Theory-Based Gravity

As emphasized by Anderson and van Wincoop (2003), equation (1) does not have an interpretation that is grounded in theory. Gravity equations can be derived from a variety of different theories.¹⁷ None of them lead to the traditional gravity specification (1).

We embed our development of gravity in a much more general treatment of trade theory than is common in the gravity theory literature in order to clear up some confusions in the literature and point the way to better work in the future. Enormous economy of thought is achieved by the methodological strategy of analyzing the allocation of trade across countries separately from the allocation of production and consumption within countries. To be precise, 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. We will refer to *trade separability* as the property of models where the allocation of production and consumption within a country, the set $\{Y_i^k, E_i^k\}$ for country i , is separable from the allocation of trade across countries. The trade allocation is the allocation of expenditure within a product class k across goods produced by different countries.

Trade separability obtains under the assumption of separable preferences and technology. By this we mean that each product class of final and intermediate goods has a distinct natural aggregator of varieties 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.

In the class of trade separable models one can derive gravity equations without having to make any assumptions about what specific model accounts for the observed output structure $\{Y_i^k\}$ and expenditure allocations $\{E_i^k\}$. 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. We can focus on a *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 takes place within this conditional

¹⁷For recent surveys of gravity theory, see Harrigan (2002), and Feenstra (2002,2003).

general equilibrium. If one wishes to conduct comparative statics, such as the impact of a change in trade barriers on trade flows, it is of course necessary to consider the full general equilibrium as a change in trade barriers will generally affect the allocations $\{Y_i^k, E_i^k\}$ by changing the market clearing prices of the conditional equilibrium sub-model. Additional information about the structure of allocation within countries will be necessary. If the goal is only to estimate trade barriers, we can limit our attention to the conditional general equilibrium.

Two additional restrictions to the class of trade separable structures yield gravity. The added restrictions are: (ii) the aggregator of varieties is identical across countries and CES, and (iii) trade costs are proportional to the quantity of trade. The CES form imposes homothetic preferences and the homogeneity equivalent for intermediate input demand. These key assumptions allow simplification of the system of demand equations and market clearing equations. We discuss extensions below which relax restrictions (ii) and (iii) in various useful ways. Our discussion 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..".

We now show how gravity equations can be derived by imposing goods market equilibrium for all varieties conditional on the allocation $\{Y_i^k, E_i^k\}$. 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)¹⁸

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

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:

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

Assumption (iii) 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

¹⁸With (more plausibly) differing weights for country of origin, the prices p_{ij}^k are interpreted as effective prices which incorporate the differing weights. See Anderson and van Wincoop (2003) for details.

country k and t_{ij}^k is the markup associated with trade costs. The latter is one plus the tariff equivalent of trade barriers.

Imposing the market clearing conditions

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

for all i and k yields gravity. To be precise, one can 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 and the entire set $\{Y_i^k, E_i^k\}$. Trade flows therefore also depend on trade barriers and the set $\{Y_i^k, E_i^k\}$. Since the $\{Y_i^k, E_i^k\}$ and X_{ij}^k are observable, one can use these conditional general equilibrium equations to draw inferences about trade barriers.

While we have derived the gravity equations under very minimal assumptions, it is often believed that much more structure is behind these equations. The misunderstanding may not be surprising as papers where the gravity equation is derived have often made many other specific assumptions that are not critical to the gravity equation itself. The first paper to formally derive a gravity equation from a general equilibrium model with trade costs is Anderson (1979). Anderson (1979) makes the Armington assumption that every country produces a particular variety. Bergstrand (1985) makes the same assumption. In the gravity equations above it may very well be though that $Y_i^k = 0$, so that some countries do not produce any variety of class k . Bergstrand (1989,1990) and Baier and Bergstrand (2001) derive gravity equations in models with monopolistic competition. This assumption again plays no role. Anderson and van Wincoop (2003) assume an endowment economy, which also plays no role for the gravity equation. Such assumptions do matter for

the results from comparative statics, such as the impact of national borders on trade in Anderson and van Wincoop (2003).

Theoretical gravity equations in Anderson (1979), Bergstrand (1985,1989,1990) and Baier and Bergstrand (2001) look far more complicated than (5). Those gravity equations contain a large number of prices and price indices. A contribution of Anderson and van Wincoop (2003) is the realization that all these prices can be summarized by just two price indices, one for the importer and one for the exporter, and that in the conditional general equilibrium these price indices can themselves be solved as a function of trade barriers and total supply and demand in each location.

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

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

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

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

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 ($E_i = Y_i$).

Anderson and van Wincoop (2003) refer to Π_i and P_j as “multilateral resistance indices”. They summarize the average trade resistance between a country and its trading partners in an ideal aggregation sense which we develop below. The main insight from the theory is that bilateral trade depends on *relative* trade barriers. Bilateral trade between i and j depends on the bilateral barrier t_{ij} divided by the product of their multilateral resistance indices for imports of j and for exports of i . Under the restriction of symmetric trade costs ($t_{ij} = t_{ji}$), the system simplifies further since $\Pi_i = P_i$.

A couple of comments are in order about homogeneous goods trade. When we let the elasticity of substitution σ_k in (5) go to infinity, trade converges to that in a homogeneous goods model. However, no information about trade bar-

riers 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, exports is equal to $E_2^k - Y_2^k$. Gravity equation (5) 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 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 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. Trade flows in the conditional general equilibrium are therefore not very sensitive to trade barriers and we cannot learn much about their size.¹⁹

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 Eaton and 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

¹⁹In this argument we implicitly assume that domestic trade barriers are normalized at zero, so $t_{ii} = 1$. More generally, trade flows depend on relative trade barriers and one can normalize by dividing all trade barriers by the lowest one.

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.²⁰

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 $1/\beta$.

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).

III.C The Trade Cost Function

Empirical applications of this theory assume that the bilateral trade barrier is a function of observables z_{ij}^m . The most common functional form is a loglinear one:

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

²⁰Eaton and Kortum only derive a gravity specification for X_{ij}/X_{ii} .

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 tariff 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. A clear limitation of gravity theory is that it has nothing to say about the trade costs function (11). This has led to often arbitrary assumptions regarding functional form, 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 $t_{ij} = d_{ij}^\rho$. Grossman (1998) argues that a perhaps 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. 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 for example the Grossman specification is correct with $\rho = 0.3$, 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. It is therefore possible to obtain highly misleading results for trade barrier estimates when the wrong functional form is adopted.

One way out of this is the approach adopted by Eaton and Kortum (2002). They assume that there are different trade barriers for 6 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. It can for example reasonably approximate both of the specifications mentioned above.

Another issue is that the most common specification (11) is multiplicative in the various cost factors. Hummels (2001a) argues that an additive specification may be more sensible and estimates the gravity equation for both trade cost specifications. 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 variables to include in the trade costs specification. There is again a need here for grounding empirics in a more theoretical approach. 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 for the trade cost specification, it is often useful to think through carefully the relationship between trade barriers and observed variables. As an example, consider language. It is common in gravity specifications to include a language variable that is one if two countries speak the same language and zero otherwise. 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% 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? Better economic geography will produce clearer results.

The trade cost function used in gravity models need *not* include domestic margins, even though these differ widely across countries. This rather surprising and important (for the validity of empirical work which has previously omitted it) proposition follows from basic gravity theory. The intuition is that domestic margins are like a lump sum tax — they do not alter relative prices. If we multiply all trade barriers t_{ij} by destination margins m_j , it is easily verified from (8)-(10) that the P_j are multiplied by m_j , the Π_i are unchanged, and therefore trade flows are also unchanged.²¹

The invariance of trade patterns to domestic retail margins which apply to all goods has another important practical implication. We can only identify relative

²¹While changes in the margins have no effect on trade in the conditional general equilibrium, they do affect trade in the full general equilibrium as they generally change production and consumption within countries.

costs with the gravity model. The most plausible way to interpret a system of trade costs $\{t_{ij}\}$ is to pick some region i and normalize $t_{ii} = 1$ (if $t_{ij} = d_{ij}^p$, the procedure normalizes distances so $d_{ii} = 1$). Essentially this procedure treats the trade cost of i with itself as a pure retail margin applied to all goods, and deflates all other trade costs by that margin. The best choice for a numeraire region is one that is quite compact and in other ways likely to have little internal trade cost, so that this procedure essentially sweeps out pure domestic margins.

A final concern about the trade cost specification is what type of regularity conditions to impose. Such conditions are often imposed implicitly. When considered explicitly they are often implausibly strong. 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 NTB's 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 trade cost variables z_{ij}^m such as common language or adjacency dummies.

NTB's cause particularly difficult problems for the inference of trade costs. The effect on trade barriers of NTB's in a country i will generally vary across trading partners j , goods k and time t . With a perfectly fitting trade model one could in principle identify the variation of the impact of NTB's on trade barriers over all dimensions (i, j, k, t) . Since this places enormous reliance on the accuracy of the economic model, some analysts (often implicitly) impose equality on the effect of NTB's over one or more dimensions (i, j, k, t) to increase the degrees of freedom. For example, Harrigan (1993) 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. NTB's have the same trade-reducing effect for all goods k that it imports from the rest of the world.²²

²²A further regularity condition is also implicitly imposed by both authors. They use NTB coverage ratios for each good as an explanatory variable, so they are also implicitly assuming that all changes in this ratio are equally important across the dimension on which equal trade flow effects of NTB's is being assumed.

We have considerable sympathy for attempts to discover useful descriptions of trade costs through the use of simple forms of (11). Our purpose in criticizing the ad hoc functional form and the regularity assumptions is to stimulate those who follow to learn more about trade costs by thinking through the relationship of the proxy variables to the unmeasured trade cost while retaining operationality.

III.D Estimation of Trade Barriers

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

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

where $x_{ij} = \ln(X_{ij})$, $y_i = \ln(Y_i)$, and $\lambda_m = (1 - \sigma)\gamma_m$.

The theoretical gravity equation can be estimated in three different ways. Anderson and van Wincoop (2003) estimate the structural equation with non-linear least squares after solving for the multilateral resistance indices as a function of the observables z_{ij}^m and the parameters λ_m . Another approach, which also gives an unbiased estimate of the parameters λ_m , is to replace the multilateral resistance indices and production variables, $y_i - (1 - \sigma)\ln(P_i)$, with region-specific dummies. This approach is adopted by Anderson and van Wincoop (2003), Eaton and Kortum (2002), Minondo (2002), Rose and van Wincoop (2002) and Hummels (2001a). Head and Mayer (2001) and Head and 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

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

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 Bergstrand (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 non-tradables 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 NTB's affect trade but not prices. 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 IV.

The estimate of the tariff equivalent of trade barriers between i and j associated with variable m is then

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

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 tariff equivalent of estimated trade barriers is approximately 9 times larger when using $\sigma = 2$ instead of $\sigma = 10$.

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 a bit tricky though since different countries or regions have different barriers for internal trade. The results will also depend a lot on how one measures barriers within a region or country. For example, the appropriate average distance within a country or region is not easy to measure. This is essentially an aggregation problem since a country or region is itself an aggregate. Head and Mayer (2001), Helliwell and Verdier (2001) and Hillberry and Hummels (2002a) show that the proper measurement of internal distance can make a big difference for the results.

III.E Estimation Bias with Traditional Gravity Equations

The most commonly estimated gravity equation (1) differs from the theoretically consistent counterpart (12) in that the multilateral resistance terms are missing

from the former. In the absence of the multilateral resistance indices 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 tariff equivalent of trade barriers associated with this variable. That is indeed how the results from estimating (1) are commonly interpreted. This interpretation is generally incorrect as 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 U.S. 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)

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

If we ignore the multilateral resistance terms, 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 U.S. (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

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

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. This point was emphasized by Anderson and van Wincoop (2003). The large multilateral resistance of provinces implies that relative trade barriers for within-country trade are very low, leading to large within-country trade. In contrast, the lower multilateral resistance for the states implies that trade within

the U.S. is not as high. Brown (2003), Brown and Anderson (2002) and Anderson and van Wincoop (2003) all find that (size-adjusted) inter-provincial trade is much higher than inter-state trade. Anderson and van Wincoop (2003) estimate that border barriers raise trade between provinces by a factor 6, while they raise trade between states by only 25%. Consistent with that, the results from that paper imply that ignoring multilateral resistance terms leads to an estimate of the tariff equivalent of border barriers of 36% and 4.7% when respectively only intra-provincial and intra-state data are used for within-country trade (assuming $\sigma = 10$).

The example above emphasizes size. But estimation bias holds more generally. As an additional illustration, consider the role of distance. When $t_{ij} = d_{ij}^{\rho} b^{\delta_{ij}}$, the gravity equation is the same as (15), 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 an old style gravity equation that omits multilateral resistance terms. 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. They can be different 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. Coulombe (2002) emphasizes the role of these topological issues related to the special structure of the regions.²³

²³While 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 is 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.

III.F The elasticity of substitution

Since estimates of trade costs from trade flows are very sensitive to assumptions about the elasticity of substitution σ , a look at the evidence is worthwhile. Although many papers have estimated this elasticity, only a few have done so in the context of equations that are theoretically well grounded, while also allowing for positive trade costs.

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

Hummels (2001a) assumes

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

where tar_{ij} is the tariff rate and f_{ij} is a freight factor, equal to one plus the tariff equivalent of freight costs. Hummels also considers an additive form of this equation, which is arguably more realistic. With the multiplicative trade cost function, the gravity equation becomes (dropping the constant term)

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

The elasticity of substitution can now be estimated through the coefficient on the log of directly observed trade costs. Hummels estimates this regression for 1992 data of imports of six countries from a large number of other countries. The gravity equation is estimated at the sectoral level, with multilateral resistance terms replaced by country dummies for each sector. The estimated elasticity rises from 4.79 at the one-digit SITC level to 8.26 at the 4-digit SITC level.

Head and Ries (2001) adopt a similar method. They consider two countries, the U.S. and Canada, and assume that the only trade barrier is a border-related barrier, which varies across industries. Estimating of the gravity equation gives an estimate of $(\sigma - 1) \ln(b_{it})$, with b_{it} equal to one plus the tariff equivalent of the

border barrier in industry i at time t .²⁴ They then decompose the border barrier into tariff and non-tariff components, $b_{it} = (1 + tar_{it})(1 + NTB_{it})$. It is assumed that $\ln(1 + NTB_{it}) = K_t + \epsilon_{it}$, 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 non-tariff 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. This is based on 3-digit industry data from 1990 to 1995.

Baier and Bergstrand (2001) estimate a theoretical gravity equation where tariffs and transport costs are the only trade barriers. Both are observed, which allows for estimation of σ . They use aggregate trade data for OECD countries and focus on changes in trade flows from the period 1958-1960 to the period 1986-1988. Their point estimate 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 $\ln f_{ij}$ and $\ln(1 + tar_{ij})$ in the gravity regressions. His restricted regressions for 28 sectors report most elasticity point estimates in the range from 5 to 10, with 4 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 8-digit SITC. The estimated elasticities range from 3 for typewriters to 8.4 for TV receivers.

²⁴In this setup the estimate is obtained as a simple analytical function of the shares that both countries spend on their own goods.

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 50 manufactured products in 19 countries, they approximate $\ln(P_j) - \ln(P_i)$ as the unweighted 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 5 to 10.

III.G The Size of International Trade Barriers

In this section we report some results for international trade barriers relative to domestic trade barriers. Recall that it is impossible for gravity to deliver absolute measures of trade barriers, so all measures are relative to some benchmark. We first present summary measures of all international trade barriers and then discuss results which decompose border barriers into several likely sources.

III.G.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 tariff 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 5, 8 and 10. 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 relative to domestic trade barriers. The results are clearly 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% for $\sigma = 10$ to 97% 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 (2001) imply an average U.S.-Canada trade barrier of 47% based on average results from 1990 to 1995. Anderson and van Wincoop (2003) do not report this barrier, but their estimated trade cost parameters imply a 46% U.S.-Canada trade barrier for 1993, virtually the same as Head and Ries (2001) report. 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 U.S. and Canada.²⁵ Eaton and Kortum (2002) report results for 19 OECD countries for 1990. For countries that are 750-1500 miles apart the trade barrier is 58-78%. The lower 58% number, which applies to countries that share a border and language, is not much larger than the estimates of the U.S.-Canada barrier. We may summarize that international trade barriers are in the range of 40-80% for a representative elasticity estimate.

Balistreri and Hillberry (2002) have argued that trade barrier estimates from these types of studies are implausibly large. They compute some numbers based on parameter estimates in Anderson and van Wincoop (2003). Table 1 of their paper reports trade barriers for Alberta and Alabama. All trade barriers are reported relative to the barrier of trade within Maryland, which has the lowest trade barrier. As an illustration of their findings, they report a trade barrier between Alabama and Quebec of 1684% for $\sigma = 3$ and 322% for $\sigma = 5$. These numbers are much larger than the 46% U.S.-Canada trade barrier reported above. There are several reasons for this discrepancy. First, elasticities of 3 and 5 are low relative to estimates in the literature reported above. If we set $\sigma = 8$, the Alabama-Quebec barrier drops to 133% relative to the Maryland barrier. Second, Balistreri

²⁵The latter includes trade within states and provinces, trade between states and trade between provinces.

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 margins reported in section II. Even if one wishes to compare to trade barriers within a state, it would be better to compare to the average trade barrier within states since the choice of the lowest trade barrier state puts a lot of weight on the rather arbitrary internal distance measure.²⁶ We find that the trade weighted average barrier for trade within the United States is 45% relative to trade within Maryland. The trade barrier for Alabama-Quebec trade relative to within-U.S. trade is 61% ($1.61=2.33/1.45$). The latter is still a bit above the 46% 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 $\sigma = 8$, the estimated U.S.-Canada border barrier from Anderson and van Wincoop (2003) is 26%. With a total barrier of 46%, it implies a distance related barrier of 16% ($1.16=1.46/1.26$). This is not much different than direct estimates of trade port costs reported in section II. Including both freight costs and the time cost of transportation, the total transport cost estimate for the U.S. is 21% ($1.21=1.107*1.09$, with 10.7% freight cost and 9% time cost).

Although based on traditional gravity equations, estimates of border barriers in Wei (1996) and Evans (2000) are not too far off, respectively 14-38% and 30%. Both apply to OECD countries. The results from Eaton and Kortum (2002) are a bit higher, 39-55%. Eaton and Kortum do not actually report barriers associated with national borders. The 39-55% barrier applies to barriers of countries 0-375 miles apart relative to domestic barriers. Since even for those countries the average distance is larger than within countries, it overstates a bit the barrier associated

²⁶Anderson and van Wincoop (2003) use the Wei (1996) measure of one fourth the distance of a region's capital from the nearest capital of another region. For specific regions this can of course be very far off. While Anderson and van Wincoop show that their comparative statics results are not sensitive to internal distance assumptions, it does make a big difference for the Balistreri and Hillberry exercise. For $\sigma = 8$, the average trade barrier within states is 21% larger than within Maryland.

with borders. But our own calculations based on estimates from Anderson and van Wincoop (2003) indicate that this overstatement is only about 6%. We can summarize that for a representative elasticity estimate, barriers associated with national borders are in the range of 25-50%.²⁷

III.G.2 Decomposition of Border Barriers

It is of great interest to better understand what factors drive these border barriers. We saw in section II that policy barriers, in the form of tariffs and NTB's, amount to no more than about 8% for OECD countries.²⁸ The evidence based on trade volumes suggest that there are important additional barriers associated with national borders. We will now discuss estimates of five different types of trade frictions that can be attributed to 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 trade barrier associated with speaking different languages of about 7% when $\sigma = 8$.

Currency Barriers

The use of different currencies may pose a barrier to trade as well. At a minimum there are costs involved in exchanging currencies and hedging currency risks. 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 barrier associated with using different currencies is 14% 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

²⁷There 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 Millimet and Osang (2001).

²⁸Recall that Messerlin (2001) reports a 7.7% arithmetic average protection rate in industrial goods for the European Union.

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 19th and early 20th century, Estevadeordal et al. (2003) and Lopez-Cordova and Meissner (2003) document a big impact on trade of belonging to the same commodity regime, such as the gold standard. Rose (2003) considers the evidence from 19 studies on the effect of currency unions on trade and concludes that the combined evidence from all studies suggest a doubling of trade if countries belong to a currency union.

One problem with this evidence is that it remains unclear exactly why a currency union raises trade levels so much. There exists substantial consensus that the impact of exchange rate volatility on trade is very small at best, with even the sign uncertain. If monetary unions do not raise trade by removing bilateral exchange rate volatility, it remains unclear why the estimated effect is so large.

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. The natural approach is to use instrumental variables. The most convincing work along this line is by Alesina et al. (2002) and Barro and Tenreyro (2002). Using a probit analysis they first compute the probability that a country adopts each of four potential anchor currencies. This probability depends on relatively 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. For any two 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.

Another approach to deal with endogeneity is to consider 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. Glick and 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. On the

other hand, Ritschl and 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 its members.

Information Barriers

Border costs associated with information barriers are potentially important as well. Evidence along this line has been documented by several authors. One example is Portes and Rey (2002). They first estimate a traditional gravity equation (1), regressing bilateral trade on GDP's, per capita GDP's and distance. After that 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 are highly significant and have a positive sign. Moreover, the coefficient on bilateral distance is reduced from -0.55 to -0.23.

Rauch and Trindade (1999) have conducted the most careful work so far. They argue that information barriers to trade may be reduced when two countries both have substantial Chinese networks. They estimate a traditional gravity equation for both 'reference price' goods and differentiated goods which do not have reference prices.²⁹ The Chinese network variable is the product of the population percentages of Chinese ethnicity in two countries. 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% Chinese population) to be

²⁹For the reference price goods they also distinguish between goods traded on organized exchanges and those not so traded.

worth as much as a 47% increase in trade.³⁰ Earlier work done on the effect of Chinese immigrant links on the bilateral trade of the U.S. 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 8, these Chinese networks save an information cost worth 6%. This may be both a lower bound and an upper bound to the tariff 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 (2000) also suggest that the 6% information barrier is an overstatement. She estimates a traditional gravity equation for OECD bilateral trade flow data for 12 industries, with trade dependent on GDP's, 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

³⁰Table 9 of Rauch and Trindade (1999) 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% and those with Chinese population shares less than 1%. For the smaller share group, the volume effects are modest — 6.2% 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% rise in trade for differentiated goods, a 128% rise in trade in reference priced goods, and a 89% 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: $47\% = 100 * (2.78 / 1.89)$.

will probably be illuminating.

Contracting Costs and Insecurity

Contracting costs are another source of border barriers. There are costs both in writing contracts and enforcing them or self-insuring the costs of default on unenforced contracts.³¹ Evans (2001b) provides evidence that internal contracting costs within a firm are much lower than external contracting costs. Specifically, the tariff equivalent of the trading costs of a foreign affiliate of a U.S. multinational with unaffiliated U.S. firms is on average 37% higher (for a demand elasticity of 5) 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 (1999) also 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 a 89% 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 very high elasticities of substitution, the tariff equivalent is likely to be small (3% when $\sigma = 20$).

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.³² 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

³¹See Anderson (2000) for a discussion of the literature on this topic.

³²They divide imports of j from i by US imports from i , which cancels exporter price index. The importer relative price index is approximated with a Törnqvist index.

their maximum likelihood point estimate of the elasticity of substitution equal to 8, the implied tariff equivalent of relatively low quality institutions is 16%. 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 raises the U.S. security level to that of Singapore, the highest in the data. This yields a bit more than a 20% increase in trade, which with an elasticity of 8 implies a tariff equivalent of almost 3%.³³

Non-Tariff Policy Barriers

Finally, Harrigan (1993), Head and Mayer (2000) and Chen (2002) consider the role of non-tariff 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 non-tariff barriers. Harrigan (1993) estimates the traditional gravity equation (1) for 1983 bilateral trade data for 28 industries among OECD countries and also finds that non-tariff 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 NTB's suggested by political economy. Treffer (1993) shows that the effect of NTB's 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 NTB's with instruments commonly used in the political economy literature.³⁴ In contrast, in a related approach which uses a set of both rich and poor countries, Lee and Swagel (1997) jointly estimate an equation relating sectoral imports to trade barriers and an equation relating sectoral NTB's to various driving factors from political economy. Once appropriately controlling for industry and country dummies in the trade regression, they find no evidence that NTB's affect trade

³³The US security composite score (0.651) is raised to the maximum score in the data, for Singapore (1.241). The trade created by this change, using the coefficient 0.285 from their regression with the log composite security index as a regressor, is around 20%

³⁴Treffer uses a cross section of US sectors for his study, implicitly assuming a common trade cost effect of NTB's as well as a common elasticity of import demand. These are not very plausible restrictions, and are clearly rejected by Harrigan's results. Nevertheless, Treffer's results very strongly indicate that accounting for endogeneity of NTB's will make a big difference in gravity models incorporating NTB's.

flows. The difference between their results and Treffer's may partly reflect differences across countries which are not controlled for, and partly the richer set of political economy variables which Treffer is able to deploy for the U.S. alone.

There is extensive evidence that free trade agreements and customs unions increase trade and therefore reduce trade barriers (e.g. Frankel et al. (1998) and Frankel (1997)), but it is less clear what elements of these trade agreements play a role (tariffs, NTB's, or regulatory issues). As noted above, all gravity model analyses of NTB's and regional trade agreements impose very strong and presumptively implausible regularity restrictions on the effect of non-tariff barriers and customs union membership upon trade volume. Moreover, most of this evidence is cross-section evidence, which raises causality issues. Time series evidence suggests that free trade agreements are less important. Although Helliwell (1998,1999) 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% policy related barrier (based on direct evidence from tariffs and NTB's), a 7% language barrier, 14% currency barrier, a 6% information cost barrier, and a 3% security barrier, overall border barriers are 44%. This falls within the estimated 25-50% range reported for OECD countries in Table 7. It should be clear though from the discussion above that this breakdown is extremely rough at best.

III.H Aggregation Issues

The gravity equations (5)-(7) form the starting point for our discussion of aggregation issues. We repeat them here for convenience:

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

$$\begin{aligned}
(\Pi_i^k)^{1-\sigma_k} &= \sum_j \left(\frac{t_{ij}^k}{P_j^k} \right)^{1-\sigma_k} \frac{E_j^k}{Y^k} \\
(P_j^k)^{1-\sigma_k} &= \sum_i \left(\frac{t_{ij}^k}{\Pi_i^k} \right)^{1-\sigma_k} \frac{Y_i^k}{Y^k}.
\end{aligned}$$

After a log-transformation, and substituting the trade cost function, the parameters can be consistently estimated with OLS by replacing $\ln(E_j^k) - (1 - \sigma_k)\ln(P_j^k)$ with an importer dummy and $\ln(Y_i^k) - (1 - \sigma_k)\ln(\Pi_i^k)$ with an exporter dummy. This is the approach taken in Hummels (2001a).

There are two different types of aggregation issues. The first is that in practice gravity equations are estimated from aggregate data. Even for disaggregated data no estimation is ever feasible at the level of detail of reality, which involves precisely defined goods at particular points in the geographic space. This raises the question of aggregation bias involved in estimating trade costs from aggregate data when trade costs vary substantially at the disaggregate level. Second, estimated trade costs at a disaggregate level overwhelm the comprehension of the analyst. Trade costs vary both across goods and geographically between and within regions. One would like to have an ideal summary index. We first discuss ideal aggregation and then aggregation bias resulting from estimating aggregate gravity equations.

III.H.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.

As an illustration, Table 8 reports the multilateral resistance indices for U.S. states and Canadian provinces based on Anderson and van Wincoop (2003), assuming $\sigma = 8$. In that model $P_i = \Pi_i$, so the import and export trade cost measures are the same. In computing these multilateral resistance indices, we adopt the approach of Balistreri and Hillberry (2002) of 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 being close to a greater share of the output produced.

Aggregating Across Trading Partners

It is desirable to have an ideal summary measure for all trade barriers t_{ij} where $i \neq j$. While trade barriers vary across countries, one would like to have one summary measure for international trade barriers. A natural way to do that is to replace all t_{ij} where $i \neq j$ with one single international trade barrier that leaves aggregate international trade unchanged. In order to illustrate this, consider the following 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. These simplifying assumptions have the advantage that $P_i = \Pi_i$ and that these multilateral resistance indices are equal across all countries. There are no trade barriers within countries. 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)

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

Since $b^{1-\sigma}$ is convex in b , it follows that b is less than the average trade barrier.

Some more algebra, which we will omit here, leads to the conclusion 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

One can also aggregate across goods. In order to illustrate that, 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

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

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 $\sum_{k=1}^N \theta_k b_k$ 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:

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

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 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} .³⁵

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 $\text{corr}(b, \sigma) < 0$. Similar conclusions can be drawn for language and adjacency as factors driving trade costs in Hummels (2001a). Evans (2000) provides similar evidence for border related trade barriers. Using 1990 OECD country data for 12 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.³⁶

Why b and σ should be negatively correlated across sectors is an interesting empirical puzzle. It suggests an element of monopoly pricing in the trade services industry. Monopoly markups vary inversely to the elasticity of demand, as the data show. Since international trade services are associated with monopoly power (Cargill dominates U.S. agricultural trade, international shipping is notoriously collusive, international air travel is still heavily regulated and so forth), it perhaps should not be surprising to find evidence of monopoly pricing in inferred trade costs.

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 industries is twice as large as an output weighted average.

³⁵Simulation 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.

³⁶Chen (2002) finds that for European trade there is no relationship between b^σ and the Rauch (1999) measure of product differentiation, again pointing to a negative relationship between b and σ across industries.

The example shows that there are many 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 II, 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.³⁷

III.H.2 Estimation Bias and Aggregation

Consider estimating an aggregate gravity equation by replacing multilateral resistance indices with country dummies. Although in the examples above international trade barriers vary across countries or goods, estimation based on aggregate data pretends that there is only one border barrier b and one elasticity σ . The aggregate gravity equation can be written as

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

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

$$\ln(\hat{b}^{1-\sigma}) = \frac{1}{N^2 - N} \sum_{i \neq j} x_{ij} - \frac{1}{N} \sum_{i=1}^N x_{ii} = \frac{\sum_{i=1}^{0.5N-0.5} \ln(b_i^{1-\sigma})}{0.5N - 0.5} \quad (24)$$

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

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

³⁷One can also compare the ideal index to the trade weighted average index. In contrast to the example above for aggregation across trading partners, in the current example for aggregation over goods the ideal index can be either larger or smaller than a trade weighted average barrier.

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

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

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 (2002). 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).³⁸

³⁸If 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 still be shown to lead to the standard aggregate gravity equation (8).

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 .³⁹ 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 it has 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 as opposed to the theoretically derived gravity model, 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 (2002) regress the product of sectoral demand

³⁹If 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.

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 Yi (2003a,2003b) are also relevant in this context and are closely related to Hillberry and Hummels (2002). These papers are based on evidence of an important role of vertical specialization in international trade, as documented in Hummels et al. (2001). Vertical specialization is defined as the use of imported intermediate goods in the production of goods that are exported. Yi (2003a) develops a model 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 produce a non-traded final consumption good. These stages can take place in different locations, leading to “back and forth” trade.

Yi argues that such a model does a much better job in accounting for the growth of world trade over the past four decades than models with only one stage of production. Yi (2003a) attributes this to two factors. First, a reduction in trade barriers can lead to vertical specialization in that the first two stages will be produced in different locations based on comparative advantage. Second, a reduction in trade barriers has a magnified impact on import demand by the final goods sector. Yi’s reasoning on magnification is that if all stages are produced in different locations, the value added from stage one faces international trade barriers twice: first from its shipment to stage two producers in another location, and second from the shipment of stage two producers to final goods producers. Using a similar model, Yi (2003b) argues that this can lead to lower estimates of trade barriers associated with international borders since trade flows become more sensitive to trade barriers.

We think this reasoning is incorrect. Assume that stage 1 is outsourced, produced in another country than the stage 2 production. If a final goods producer sources a stage 2 intermediate good from a foreign supplier, the stage one component will indeed have crossed borders twice. But if the final goods producer sources a stage 2 intermediate good from domestic producers, the stage 1 component will still have crossed the border one time. The stage 1 component therefore will have

crossed the border just one more time when stage 2 intermediate goods are sourced abroad. There is no magnification effect.

The model by Eaton and Kortum (2002) confirms that multiple stages of production do not lead to a magnification effect. 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. The same bundle of intermediate goods is used for the production of each good. All intermediate goods, including the one produced at home, have components that have effectively crossed borders infinitely many times. When a consumer buys a foreign good, its components have crossed the border just one more time than for a domestic good. The Eaton and Kortum model leads to a standard gravity equation as shown above, confirming that there is no magnification effect.

In contrast, Yi (2003a,b) is correct that trade barriers have a bigger impact on trade flows due to their effect on the production location. The argument is similar to that in Hillberry and Hummels (2002). With positive trade barriers production tends to be located close to demand. If stage 2 production of a particular good is located in country 1, demand for stage 1 intermediates for that good is high in country 1. Production of stage 1 is therefore likely to be located in country 1 as well if international trade barriers are sufficiently high. The fact that there are multiple stages of production is not critical here. If for example all stage 2 goods are produced with the same index of stage 1 goods, demand for stage 1 goods is the same everywhere and there is no reason for stage 1 and 2 firms to locate near each other.⁴⁰ Such a symmetric demand structure is a key element of the Eaton and Kortum model.

These models all suggest that the production location response to trade barriers can lead to an upward bias of trade costs estimates based on aggregate data. The magnitude of this bias remains unknown though and will be an important area

⁴⁰If we define stage 1 and stage 2 in the Yi model as two different sectors, trade in each of these sectors does obey the sectoral gravity equation (5) if the technology parameter in stage 1 is drawn from a Fréchet distribution, as in Eaton and Kortum (2002), while the second stage has the same productivity parameters across countries. If the technology parameter in stage 2 production also varies across countries, and is drawn from a Fréchet distribution, trade in the stage 2 sector does not follow (5) for the technical reason that the product of two Fréchet distributions is not itself a Fréchet distribution.

for future research. One obvious recommendation is to disaggregate. Indeed it is important to note that 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.

III.I Criticisms of the Gravity Approach

Various criticisms have been directed against using gravity equations such as (8) and (5) as a tool for analyzing trade volumes and trade costs. Here is a sampling:

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”-Coe et al.(2002)).
2. There are no import-competing sectors or non-tradables sectors that only supply to the domestic market (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 (Haveman and Hummels (2001)).
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. 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. One needs to take into account though that McCallum estimated a traditional gravity equation without theoretical foundations. The distance elasticity is much lower (-0.79) 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 tariff 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:

$$\alpha(1 - \sigma)\partial \ln(\tau_{ij})/\partial \ln(d_{ij}) \tag{26}$$

One would expect the transport costs to be at most proportional to distance, so that the elasticity $\ln(\tau_{ij})/\partial \ln(d_{ij})$ is less than 1. Hummels (2001a) estimates it to be about 0.3.⁴¹ As discussed in section I, the tariff equivalent for transport costs for the U.S. are on average about 11%. 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 tariff equivalent of U.S. transport costs are on average 21%. 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].

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. As discussed above, Portes and Rey (2002) find that the distance elasticity drops 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 has the advantage that it can deal with zero trade

⁴¹Grossman (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} .

observations. Although it is not clear why, this significantly reduces the absolute value of the distance elasticity, from -1.08 to -0.35 for 2000 trade data. A possible explanation is that their finding is consistent with a model in which 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.

Coe et al. also find that the distance elasticity has fallen over time, from -0.51 in 1980 to -0.35 in 2000. A large recent literature surveyed in Coe et al. (2002) has found that the distance elasticity has not declined or even risen over time, which has been considered puzzling since transport costs have declined over time. Coe et al. (2002) refer to it as the missing globalization puzzle. It is not intuitively clear why estimation in levels would resolve this puzzle. An alternative resolution of the non-declining distance elasticity puzzle is presented by Brun et al. (2002). By including the cost of fuel in the trade cost function, they succeed in reversing the implication of rising distance elasticity, at least for the set of bilateral trades between rich countries. 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 second criticism, raised by Engel (2002), is that there are no import-competing sectors or non-tradables sectors that do not export. Engel makes this comment in the context of the model by 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 non-tradables. Assume that as a result of a rise in border barriers there is a shift in resources out of tradables into non-tradables. The non-tradables are only sold locally within a region and not between regions. As a result a given 10% drop in exports of provinces will lead to a smaller increase in trade with other provinces than in the absence of non-tradables. 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. Non-tradables do not affect the gravity equation for tradables and trade barrier estimates are therefore unaffected.⁴²

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 non-tradables and the estimated border barrier does not.

In contrast, full general equilibrium comparative statics is affected by the presence of non-tradables. In the presence of non-tradables a rise in border barriers will shift resources to the non-tradables 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% drop in international trade, the increase in inter-provincial trade will be lower.

In short, the gravity equation and estimates of trade barriers are unaffected by non-tradables but comparative statics is. 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 same will be the case under a modification of the CES preference structure discussed below.

⁴²Under trade separability, non-tradables do not affect the marginal utility (or marginal productivity) of different types of tradable goods within a sector.

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. In the “new open economy macro” literature it is often assumed that the elasticity of substitution between domestic brands is higher than between domestic and foreign brands. Evidence quoted in defense of this is very weak though. But there is another good reason to consider the case of a higher elasticity of substitution among domestic goods. 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.

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 U.S. and Canada trade between them rose rapidly while trade within Canada (inter-provincial trade) did not fall much. This stands in sharp contrast to the finding by Anderson and van Wincoop (2003) that border barriers have increased inter-provincial trade much more (factor 6) than they reduced state-province trade (about 44%).⁴³

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.⁴⁴ This does not affect the estimation of the trade cost parameters, which only depends on the gravity structure.

One possible explanation for the Helliwell puzzle is a non-tradables sector. As discussed above, a border barrier will then lead to a larger reduction in international trade and a smaller reduction in state-province trade. Unfortunately, this explanation is not very plausible. It would only work if trade barriers lead to a significant shift out of tradables into non-tradables. Because of its much larger size, the share of the tradables sector would not be much affected in the U.S. and

⁴³Possibly related to the Helliwell evidence is the point raised by Brown (2003) that industries where border barriers have disappeared when comparing data on inter-state and state-province trade tend to still have much more inter-provincial trade than inter-state trade.

⁴⁴Anderson and van Wincoop (2003) adopt a simple endowment economy.

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 (1958-1960 and 1986-1988) 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. In order to effectively exploit the time series properties of the data, it is best to estimate gravity equations in their structural form. Otherwise one needs to include separate country dummies for each year, so that estimates only use cross-sectional aspects of the data. In accounting for changes in world trade, Estevadeordal *et al.* ignore the changes in the multilateral resistance variables.

The fifth criticism, launched by Haveman and Hummels (2001), is that gravity models imply that all countries purchase goods from all suppliers. Using 1990 bilateral trade data among 173 countries for 4-digit SITC categories, Haveman and Hummels show that in 58% of cases importers buy from fewer than 10% 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 one supplier (the cheapest one), even though many suppliers may exist. It can therefore account for zero trade flows at a disaggregated level. The Eaton and Kortum model cannot account for zero trade flows at an aggregate level, 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% higher in Canada than in the United States (assuming $\sigma = 5$). 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 prices fluctuate 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% 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.

III.J Extensions of the Gravity Approach

We show that simple extensions can accomplish much greater flexibility of the model while retaining the essential simplicity of the gravity model (5)-(7). Recall that we made three assumptions to derive the gravity equation (5): (i) trade separability, (ii) the aggregator of varieties is identical across countries and CES, and (iii) trade costs are proportional to the quantity of trade. Below we only discuss a variety of extensions of (ii) and (iii). The most ambitious relaxation of assumptions takes us out of the world of trade separable structures entirely. If the separability property is not satisfied, two-stage budgeting is impossible. It is worthwhile to at least consider how estimates of trade barriers depend on this. As

far as we know no work has yet been done in this direction.

Extensions of the CES Structure

Tchamourliysky (2002) considers non-homothetic CES preferences. He modifies a standard CES consumption index by adding a constant γ_{ij} to consumption of goods from country i by country j consumers. He then derives a modified gravity equation and estimates γ_{ij} to be negative. This makes trade flows more sensitive to trade barriers. Incorrectly assuming CES preferences would make consumption less 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).

Another direction in which the CES preference structure can be generalized is to allow for zero consumption of some goods to be optimal if trade barriers are sufficiently high. Any change in the preference structure that keeps the marginal utility from consuming goods finite when consumption is zero accomplishes this. To our knowledge, this extension has not yet been implemented. An advantage of this generalization is that it can explain the observed zero trade flows, the concern of Haveman and Hummels (2001). It can also explain that certain firms do not export at all.

The CES preference structure can also be generalized by assuming that certain goods within a sector are better substitutes than other goods using nested CES structures. For example, in the automobile industry, compact cars are better substitutes for each other than compact cars are with vans. One direction in which this assumption can be generalized is to allow domestic goods to be closer substitutes for each other than domestic goods are with foreign goods. This would particularly address the criticism listed above by Engel (2002). It also corresponds to allowing for lower productivity differences within than between countries in the context of the Eaton and Kortum (2002) model.

Consider for example data on trade flows for U.S. states and Canadian provinces.

Assume a one-sector world. Let σ_H between 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:

$$X_{ij} = \frac{Y_i Y_j}{Y_w} \frac{t_{ij}^{1-\sigma_H}}{P_{j,c}^{\sigma_F - \sigma_H} P_j^{1-\sigma_F} \Pi_i^{1-\sigma_H}} \quad (27)$$

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 . All these indices can be solved as explicit functions of trade barriers. 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 for example easy 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.

Differences in Preferences and Technology

Another key restriction on preferences and technology in the gravity model is that they are the same for all agents. This is clearly a strong assumption and can be generalized 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. It is well known though that differences in preferences are empirically indistinguishable from trade costs. Consider the following utility function for country j consumers:

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

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} , so that trade barriers and differences in preferences are empirically indistinguishable. Under sufficiently strong restrictions on taste differences, it is possible to distinguish trade costs from taste differences.⁴⁵ Another

⁴⁵Bergstrand (1985) derives a gravity equation under a different type of heterogeneity in pref-

promising route to distinguishing the two is to exploit time series variation. Trade costs plausibly move over time while tastes presumptively are stable.

Evans (2003) 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 (2003) 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 is critical and not the nationality of the firm that sells the goods.

Preferences can also be different across individual consumers in that they prefer different “ideal varieties”, as in Lancaster (1979) preferences. Helpman and Krugman (1985) claim that in most essentials, including trade volume and presumably inference about trade costs, the ideal varieties approach offers little difference from the Dixit-Stiglitz love of varieties approach. The latter being easier to work with, it is preferable. This claim has been the foundation for the vast theoretical and empirical literature which has ensued, including many of the gravity model studies cited above. If the claim is valid, the distribution of consumers within countries being different across countries in the ideal varieties approach would be presumably act like different ‘tastes’ in the love of variety approach. But at the very least, the ‘taste difference’ would be related to distributional parameters and so have empirical implications. We speculate that the ideal varieties approach may be worth investigating in future work on home bias.

Demand can also be different when it comes from firms buying intermediate goods. We assume that two stage budgeting is possible, so that the production function of a particular firm can be written as a function of separate indices of intermediates for each sector. We will ignore aggregation problems by assuming that the intermediate products in each sector are at the most disaggregated level of goods produced by individual firms. Two types of heterogeneity in production

erences. 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.

functions are important. First, different firms naturally have a different aggregate demand for a particular sector. This can lead to differences in E_j^k in (5). As discussed above, variation in E_j^k across countries in a way that is correlated with trade barriers can give rise to aggregation bias when estimating an aggregate gravity equation. But it is irrelevant when estimating the disaggregated gravity equation (5). The second type of heterogeneity is that for a specific industry the index of intermediates in the production function differs across firms. For example, assume that each country has only one firm producing a particular type of sector k goods. For that sector one can interpret (28) as an index of intermediates in the production functions of firms in country j . For a given production structure the same comment can 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 firms tend to give more weight to particular types of sector k 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.

Fixed Costs of Trade

We can usefully relax the standard trade costs assumption that trade costs are proportional to the quantity 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 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 (2003) attempts to recompute the tariff 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% 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 (2003) 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 (2003), Klenow and Rodriguez (1997) develop a model in which fixed costs are borne by importers.⁴⁶ In that setup it is possible that firms export only 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 (1997) find that a drop in tariffs from 48.5% to 22.1% over this period lead to an increase in the number of imported varieties of 30% for consumer goods and 20% for intermediate goods.

It is not hard to derive a simple generalization of (5)-(7) that 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 . We know these sets from the trade flow data. When there are positive exports from i to j the generalized gravity equation becomes

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

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

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

This is a very general setup. It 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

⁴⁶They use their model to compute welfare gains from trade liberalization.

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.

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.⁴⁷

IV 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 to estimate NTB’s, although the evidence generally applies to a much broader range of trade barriers. The macro literature compares domestic retail prices across countries of similar goods instead of comparing import to domestic prices. This literature is based on examining deviations from purchasing power parity as opposed to measuring trade barriers, so it is focused on issues such as the speed of convergence of prices across countries or 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.⁴⁸

⁴⁷Bergstrand (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 goods 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.

⁴⁸See for example Froot and Rogoff (1995) and Goldberg and Knetter (1997).

IV.A 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 NTB’s, 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 NTB’s, 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 NTB’s, discuss evidence based on this type of price comparison. They explain in great detail why the price ratio p_j/p_j^m does not necessarily capture all trade costs associated with quotas and other NTB’s. If for example the exporting firm has market power (its good is differentiated) it may be able to extract the quota revenue for itself by raising the import price. Knetter (1997) argues that this is the case for exports to Japan. He finds that German exporters charge substantially higher prices when exporting to Japan than to the United States, United Kingdom and Canada, and argues that this is the result of a variety of NTB’s in Japan. Another possibility is that a quota licence is 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 (VER's) 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 markups. 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. But even so, atheoretic averaging of transactions prices is inescapable (using either trade weights or arithmetic weights) but will generally be biased. 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. 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 a large number of studies that have estimated trade barriers based on price comparisons. Data and conceptual problems aside, 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 agri-

culture 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 E.U.) to avoid supporting farmers in rest of the world. The data is convincing as a demonstration that price supports and the associated NTB's create quite large and varied distortions in agricultural trade. Sugar is the most notorious example, where the support price far exceeds the world price in any year. In 1991 the sugar support price was equivalent to a 125% tariff, while in some years the tariff equivalent has exceeded 300%. 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% in dairy products, 28% in chicken and 29% in turkey. Other countries have even more extreme agricultural distortions — Japan is notorious for a domestic rice price over 10 times the world price, with similar differentials for sugar in some years.

Sazanami et al. (1995) use the unit value approach to compute the tariff equivalent of trade barriers for Japan in 1989. Although there are serious measurement problems with this approach, the advantage is that coverage is broad, including not only agriculture and textiles, but the entire manufacturing sector. They find average trade barriers of 59.5% for metal products, 128.3% for chemical products and 140.2% for machinery.

IV.B 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

$$\frac{1}{t_{ij}} \leq \frac{p_i}{p_j} \leq t_{ij} \quad (32)$$

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. Goldberg and 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.

IV.B.1 Time Series Evidence

A well-known paper by Engel and Rogers (1996) computes the standard deviation of relative prices between Canadian and U.S. cities for 14 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 11 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.

Obstfeld and Taylor (1997) is another example of a study that has used price index information. Their starting point is again (32). They estimate a threshold autoregression (TAR) model⁴⁹, 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 U.S. 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 production 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.

IV.B.2 Extracting Information from Price Levels

Some detailed recent survey evidence of price levels for individual goods in cities around the world has led to a promising, but still small and very recent, literature aimed at extracting information about trade costs. Since the approach that most authors have taken is rather a-theoretical, 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 cost of production, (ii) trade costs, (iii) various markups over cost in the

⁴⁹See also Taylor (2000) and references therein.

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 tariff 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 by imposing such arbitrage by producers. 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

$$\frac{t_{iz_i}}{t_{jz_i}} \leq \frac{p_i}{p_j} \leq \frac{t_{iz_j}}{t_{jz_j}} \quad (33)$$

Under the reasonable assumption that the trade cost for shipments from i to j is no larger than the trade cost for shipment from i to m and then from m to j , it follows that $t_{ij} < t_{im} t_{mj}$. Equation (32) then follows directly from (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

$$\frac{p_i}{p_j} = \frac{t_{im}}{t_{jm}} \quad (34)$$

In this case the relative price is completely tied down by trade barriers.

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. We are not aware of any papers that have attempted to measure trade barriers this way. A problem that arises is that 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, with each having 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 a Fréchet distribution with cumulative distribution function $e^{-z^{-\theta}}$. The variance of z is inversely related to the parameter θ .

As for economic geography, 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 (tariff equivalent) across all goods and location pairs is t_{av} .

General equilibrium arbitrage implies that each country buys from the cheapest producer, so that the price of good g in country i is

$$p_i^g = \min(c_1^g t_{1i}^g, \dots, c_N^g t_{Ni}^g) \tag{35}$$

where c_j^g is the production cost of good g in country j and t_{ji}^g is one plus the tariff 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 $G = 3000$. Further raising G makes no difference for the results reported below.

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 first measure is location-specific. For location i it is the average of $|p_i^g - p_j^g| / (0.5p_i^g + 0.5p_j^g)$ across both goods and locations j . The second measure is good-specific and is the average of $|p_i^g - p^g| / p^g$ across locations i , with p^g the average price of good g across locations. The third measure is location-specific and is the 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 . The fourth measure is location pair-specific. For location pair (i, j) it is the standard deviation of $\ln(p_i^g) - \ln(p_j^g)$ across goods. The final measure is the 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) compute the third price dispersion measure. They first subtract local distribution costs from the final goods price and call the result the producer price. They compare this to the “landed price,” which is the lowest producer price plus transport cost to that location. The percentage difference is interpreted as a measure of “fragmentation”.⁵⁰ 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 costs 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

⁵⁰Crucini et.al.(2000) compute the percentage saving when countries purchase a common basket of goods from the lowest price locations and refer to it as a comprehensive measure of the cost of arbitrage. This is a slight variation on the third measure with the denominator being p_i^g rather than p_i^{low} .

cheapest country).

Parsley and Wei (2002) compute measure 4. They justify it based on the arbitrage equation (32), which says that $\ln(p_i^g) - \ln(p_j^g)$ is bound between $-\ln(t_{ij})$ and $\ln(t_{ij})$. 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 this relative price across goods. But as discussed above, the 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 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 (multiplied by 100). The trade weighted average trade cost is much lower than the arithmetic average. The reason is that 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. Members of a cluster tend to be located next to each other on the circle. 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 9 different parameterizations. All price dispersion measures are far below the ideal index. There is also no apparent relationship with the trade weighted average index. A couple of examples illustrate this. If we lower the elasticity of substitution σ from

4 to 1, 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, holding the average barrier at 100%, the trade weighted average barrier rises from 31% to 100%. The price dispersion measures change very little though and all fall far below the trade weighted average. When we raise the number of countries from 10 to 20, 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 one representative simulation 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. There are multiple pairs of countries that have the same average bilateral trade barrier. They should also have the same price dispersion. Panel A shows some minor variation though as a result of the fact that the number of goods is less than infinity. The main point is that price dispersion is higher for location pairs that have a higher average trade costs. Panel B shows a similar result for goods. It is 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. Panel B shows results from a representative simulation. Each point in the panel represents the average for 10 goods. It is clear that price dispersion tends to rise for goods with higher average trade barriers (across all locations).

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 by assuming that there are no internal trade costs. Taxes simply add to the prices that have been computed from the model, without altering the equilibrium otherwise. Under a reasonable assumption the same is the case for internal trade costs. If we write $t_{ij} = \tilde{t}_{ij}t_{jj}$, where \tilde{t}_{ij} is the international trade cost and t_{jj} is the local trade cost in j , the equilibrium prices net of local trade costs are the same as before. We simply need to add local trade costs and taxes to this equilibrium price. 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. 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.⁵¹

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.⁵²

⁵¹See 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.

⁵²If the prices from the model are interpreted as ex-ante prices in the buyer's currency, ex-post prices can be compared by multiplying them by an exchange rate E_i relative to a numeraire currency, which then becomes a component of t_i^g .

As an illustration we will assume that $\ln(t_i^g)$ has a standard deviation of 23% across goods and countries, including country and good-specific components with standard deviations of both 5%.

Rows (x)-(xii) of Table 10 shows the average price dispersion moments which result from introducing variation in $\ln t_i^g$ for average trade barriers ranging from 0 to 200%. It further illustrates that average price dispersion measures tell us very little about trade barriers. For example, doubling the average trade barrier from 100% to 200% 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. Panel C of Figure 2 shows that there remains a strong positive relationship across country pairs between the fourth measure of price dispersion and their average trade barrier. Panel D shows that goods with high average trade barriers also continue to have higher price dispersion. Even though the additional sources of price dispersion cloud these relationships somewhat, they still come out strong.

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 Table 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 of price dispersion are included. But since this can be said for an average trade barrier in the model of 100% as well as for an average trade barrier of 200%, 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 countries. 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 14 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% for non-EMU country pairs with the same currency and by 4.3% for EMU countries, relative to countries that do not have a common currency. They transform these into tariff equivalents by using that a 1% tariff reduction reduces the price dispersion measure by 0.86%. The reduction in the tariff equivalent of trade barriers is then 3.7% for non-EMU countries in a currency union and 5% for EMU countries. If both cities are located in the United States the tariff equivalent of trade barriers is reduced by 12%. 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 1976-1997 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 13 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 use the second price dispersion measure, the average of $|p_i^g - p^g|/p^g$ across locations i . They do not particularly focus on trade costs, but one of their results is closely related to trade costs. They regress their good-specific measure of price dispersion on the trade level for each good and the share of non-traded inputs for each good. The trade level is defined as total trade among the 13 European countries in their sample, divided by total output for that industry across these countries. The coefficient on trade is negative and highly significant. 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, local trade or distribution costs, different 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.⁵³

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 relative to the producer 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

⁵³Bradford and Lawrence (2003) provide a description of the aggregation procedures adopted by the OECD.

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 the wholesale price times an exogenous dealer markup and the gross value added tax rate. In order to extract information about trade barriers it is necessary to model local trade costs in the destination market and the markup in the wholesale price. They assume an oligopoly with multiproduct firms. The optimal wholesale price charged by a firm is equal to marginal cost plus a markup plus a term associated with quotas. Marginal cost depends on the characteristics of the car, the wage rate in the producer country, the wage rate in the destination country (associated with local distribution and marketing), other factor prices and the total quantity of production. We believe though that for the purposes of extracting information about trade barriers it is sufficient to model only the local distribution costs in the destination market since the cost in the producer country drops out when comparing relative prices. The optimal markup also depends on the demand side of the model, where they assume a discrete choice model that leads to a logit specification. Aggregate demand for a particular car in a particular market depends on price, product characteristics, income and income distribution. They separately estimate the demand and supply side of the model.

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 may be. 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. Trade costs can also be assumed to depend on various characteristics of goods. The larger the number of goods, the more precise the estimates of trade barriers from this method. For realism the Eaton and Kortum model would need to be extended to allow for other sources of price dispersion, such as local distribution costs and taxes.

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):

$$\begin{aligned} (\Pi_i^k)^{1-\sigma_k} &= \sum_j \left(\frac{t_{ij}^k}{P_j^k} \right)^{1-\sigma_k} \frac{E_j^k}{Y^k} \\ (P_j^k)^{1-\sigma_k} &= \sum_i \left(\frac{t_{ij}^k}{\Pi_i^k} \right)^{1-\sigma_k} \frac{Y_i^k}{Y^k}. \end{aligned}$$

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 non-linear least squares.

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 P_i^k .

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 model as a function of both local wages and internal distance. Direct information on local distribution margins may be used as well. Third, as discussed in section IV, 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).

V Conclusion

Trade costs are large on average. Broadly defined, they dominate production costs everywhere. Excluding location-specific distribution costs halves their size but they still loom large on average. Trade costs 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 10 or more. 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 feasibly 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.

Inference about trade costs other than those associated with policy and with transport can also be improved greatly. The structural gravity model, particularly as expositied and extended in the survey, offers the potential for theoretically consistent and more precise estimates of key implicit trade costs. This is obviously so for inference about trade costs from trade flows but we argue that it is likely to be so for inference about trade costs from international price comparisons.

Aggregation of trade costs makes a big difference. We provide ideal alternatives to existing atheoretic procedures of aggregation based on trade weights, production weights or arithmetic weights. We provide a sample application of ideal average trade costs associated with distance and borders for states and provinces based on our study of U.S.-Canada trade. Future work should provide useful guides to the obstacles to trade posed by their own and their partners' resistance using ideal indices to aggregate across the many sources of trade costs, directly measured and implicit, for many regions and countries.

References

- Alesina, Alberto and 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," *American Economic Review*, 1979, 69(1), pp. 106-116.
- (1994), "The Theory of Protection", *Surveys in International Trade*, David Greenaway and L. Alan Winters, eds., Oxford: Blackwell Publishers, 107-138.
- (1998), "Effective Protection Redux", *Journal of International Economics*, 44, 21-44.
- (2000), "Why Do Nations Trade (So Little)?" *Pacific Economic Review* 5, 115-134, 2000.
- Anderson, James E. and Douglas Marcouiller (2002), "Insecurity and the Pattern of Trade: An Empirical Investigation," *Review of Economics and Statistics*, 84(2), 345-352.
- Anderson, James E. and J. Peter Neary, (1994), "The Trade Restrictiveness of the Multi-Fibre Arrangement", *World Bank Economic Review*, 8(2), 171-189.
- , (1996), "A New Approach to Evaluating Trade Policy", *Review of Economic Studies*, 63, 107-125.
- , (2003), "The Mercantilist Index of Trade Policy", *International Economic Review*, 44, 627-649.
- Anderson, James E. and Eric van Wincoop (2003), "Gravity with Gravitas: A Solution to the Border Puzzle," *American Economic Review*, March 2003, 93, 170-192.
- Anderson, James E. and Eric van Wincoop, (2002), "Borders, Trade and Welfare", *Brookings Trade Forum 2001*, Susan Collins and Dani Rodrik, eds., Washington: The Brookings Institution, 207-244.
- Anderson, Michael A. and Smith, Stephen L.S. (1999a), "Canadian Provinces in World Trade: Engagement and Detachment," *Canadian Journal of Economics* , 32(1), pp. 23-37.
- (1999b), "Do National Borders Really Matter? Canada-U.S. Regional Trade Reconsidered," *Review of International Economics*, 7(2), pp. 219-227.
- Bacchetta, Philippe and Eric van Wincoop. "Why do Consumer Prices React Less than Import Prices to Exchange Rates?," *Journal of the European Economic Association*, forthcoming, 2003.
- Baier, Scott L and Jeffrey H. Bergstrand (2001), "The Growth of World Trade: Tariffs,

- Transport Costs and Income Similarity,” *Journal of International Economics* 53, 2001, 1-27.
- Balistreri, Edward J. and Russell H. Hillberry (2002), “Trade Frictions and Welfare in the Gravity Model: How much of the Iceberg Melts?,” working paper, U.S. International Trade Commission, 2002.
- Barro, Robert J. and Silvana Tenreyro (2002), “Economic Effects of Currency Unions,” working paper, Harvard University, 2002.
- Bergstrand, Jeffrey H. (1985), “The Gravity Equation in International Trade: Some Microeconomic Foundations and Empirical Evidence,” *Review of Economics and Statistics*, 1985, 67(3), pp. 474-81.
- (1989), “The Generalized Gravity Equation, Monopolistic Competition, and the Factor-Proportions Theory in International Trade,” *Review of Economics and Statistics*, 1989, 71(1), pp. 143-53.
- (1990), “The Heckscher-Ohlin-Samuelson Model, the Linder Hypothesis and the Determinants of Bilateral Intra-Industry Trade,” *The Economic Journal*, 1990, 100, pp. 1216-29.
- Bernard, Andrew B. and J, Bradford Jensen (2003), “Why Some Firms Export,” *Review of Economics and Statistics*, 2003, forthcoming.
- Bernard, Andrew B. and Joachim Wagner (2001), “Export Entry and Exit by German Firms”, *Weltwirtschaftliches Archiv*, 2001, 137(2), pp. 105-123.
- Bradford, Scott C. and Robert Z. Lawrence (2003) “Paying the Price: The Cost of Fragmented International Markets,” Institute of International Economics: Washington DC, forthcoming.
- Brown W. Mark (2003), “Overcoming Distance, Overcoming Borders: Comparing North American Regional Trade,” Research Paper 8, 2003, 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 U.S. Regions,” *Papers in Regional Science*, 2002, 81, 99-120.
- Burstein, Ariel, Joao C. Neves and Sergio Rebelo. “Distribution Costs and Real Exchange Rate Dynamics During Exchange-Rate-Based Stabilizations,” *Journal of Monetary Economics*, forthcoming, 2003..
- Chen, Natalie A. (2002), “Intra-national Versus International Trade in the European Union: Why Do National Borders Matter?”, CEPR Discussion Paper 3402, 2002.

- Coe, David T. and Natalia T. Tamirisa (2002), "The Missing Globalization Puzzle," IMF Working Paper, August 2002.
- Coulombe, Serge (2002), "Border Effects and North American Economic Integration," working paper, University of Ottawa.
- Crucini, Mario J., Chris I. Telmer and Marios Zachariadis (2001) "Understanding European Real Exchange Rates," working paper, Vanderbilt University.
- Crucini, Mario J., Chris I. Telmer and Marios Zachariadis (2000) "Dispersion in Real Exchange Rates," working paper, Vanderbilt University.
- Crucini, Mario J. and Mototsugu Shintani (2002) "Persistence in Law-of-One Price Deviations: Evidence from Micro-Data," working paper No 02-W22, Vanderbilt University.
- Davis, Donald R. (1998), "The Home Market Effect, Trade and Industrial Structure", *American Economic Review*, 88, 1264-1276.
- Deardorff, Alan V (1998), "Determinants of Bilateral Trade: Does Gravity Work in a Neoclassical World?," in J.A. Frankel, ed., *The Regionalization of the World Economy*. Chicago: University of Chicago Press, 1998, pp. 7-32.
- Deardorff, Alan V. and Robert M. Stern, (1998), *Measurement of Nontariff Barriers*, Ann Arbor: University of Michigan Press.
- Eaton, Jonathan and Samuel Kortum (2002) "Technology, Geography and Trade," *Econometrica*, 2002, 70(5), 1741-1779.
- Grossman, Gene (1998), "Comment on Deardorff", in J.A. Frankel ed., *The Regionalization of the World Economy*. Chicago: University of Chicago Press, 1998, pp. 33-57.
- Engel, Charles (2002), "Comment on Anderson and van Wincoop" in *Brookings Trade Forum 2001*, Susan Collins and Dani Rodrik, eds., Washington: The Brookings Institution.
- Engel, Charles and John H. Rogers (2001). "Deviations from Purchasing Power Parity: Causes and Welfare Costs," *Journal of International Economics* , 55, 29-57.
- , (1996), "How Wide is the Border?", *American Economic Review*, 86:1112-1125.
- , (2000), "Relative Price Volatility: What Role Does the Border Play?," in Gregory Hess and Eric van Wincoop, eds., *Intranational Macroeconomics*, (Cambridge University Press, Cambridge), 92-111.
- Estevadeordal, Antoni, Brian Frantz and Alan M. Taylor (2003), "The Rise and Fall of

- World Trade, 1870-1939,” *Quarterly Journal of Economics* 68(2), 359-408, 2003.
- Evans, Carolyn L (2000). “The Economic Significance of National Border Effects,” Mimeo, Federal Reserve Bank of New York, 2000.
- (2001a). “Home Bias in Trade: Location or Foreign-ness?,” working paper, 2001a.
- (2001b). “The Costs of Outsourcing,” working paper, 2001b.
- , (2003). “Border Effects and the Availability of Domestic Products Abroad,” working paper, Board of Governors, 2003.
- Evenett, Simon J. and Anthony Venables, (2002), “Export Growth in Developing Countries: Market Entry and Bilateral Trade Flows,” working paper, London School of Economics.
- Feenstra, Robert C. (2003), *Advanced International Trade: Theory and Evidence*, Princeton University Press, 2003.
- , (2002), “The Gravity Equation in International Economics: Theory and Evidence,” *Scottish Journal of Political Economy*, 49(5), 2002, 491-506.
- (1994), “New Product Varieties and the Measurement of International Prices,” *American Economic Review*, 1994, 84(1), 157-177.
- (1998), “Integration of Trade and Disintegration of Production in the Global Economy”, *Journal of Economic Perspectives*, 12, 31-50.
- Frankel, Jeffrey A. and Shang-Jin Wei (1997), “ASEAN in a Regional Perspective,” in J. Hicklin, D. Robinson and A. Singh, eds., *Macroeconomic Issues Facing ASEAN Countries*. Washington, DC: International Monetary Fund, 1997.
- Frankel, Jeffrey, Ernesto Stein and Shang-jin Wei (1998), “Continental Trading Blocs: Are They Natural or Supernatural?” in J.A. Frankel ed., *The Regionalization of the World Economy*. Chicago: University of Chicago Press, pp. 91-113.
- Froot, Kenneth A. and Kenneth Rogoff (1995) “Perspectives on PPP and Long-Run Real Exchange Rates,” in Gene M. Grossman and Kenneth Rogoff, eds., *Handbook of International Economics*, volume 3, Elsevier: Amsterdam, 1645-1729.
- Ghosh, Atish and Holger C. Wolf (1994), “Pricing in International Markets: Lessons from the Economist,” NBER Working Paper 4806, July 1994.
- Glick, Reuven and Andrew K. Rose (2002), “Does a Currency Union Affect Trade? The Time Series Evidence”, *European Economic Review*, 46, 1125-1151.
- Goldberg, Pinelopi Koujianou and Michael M. Knetter (1997) “Goods Prices and Exchange Rates: What Have we Learned?,” *Journal of Economic Literature*, vol.

35, no. 3, 1243-1272.

- Goldberg, Pinelopi Koujianou and Frank Verboven (2001), "The Evolution of Price Dispersion in European Car Markets", *Review of Economic Studies*, 68, 811-48.
- Gould, David M. (1994), "Immigrant Links to the Home Country: Empirical Implications for U.S. Bilateral Trade Flows," *Review of Economics and Statistics* 76:2, May, pp.302-316.
- Harrigan, James (1993), "OECD Imports and Trade Barriers in 1983," *Journal of International Economics* 34, 1993, 91-111.
- Harrigan, James (2002) "Specialization and the Volume of Trade: Do the Data Obey the Laws?," in Kwan Choi and James Harrigan, eds. *The Handbook of International Trade*, London: Basil Blackwell, 2002.
- Haveman, Jon and Hummels, David (1999), "Alternative Hypotheses and the Volume of Trade: Evidence on the Extent of Specialization," Mimeo, Krannert School of Management, Purdue University, 1999.
- Head, Keith and Thierry Mayer (2000), "Non-Europe: The Magnitude and Causes of Market Fragmentation in the EU," *Weltwirtschaftliches Archiv* 136(2), 2000, 284-314.
- (2001) "Illusory Border Effects: Distance Mismeasurement Inflates Estimates of Home Bias in Trade," working paper, 2001.
- Head, Keith and John Ries (2001), "Increasing Returns versus National Product Differentiation as an Explanation for the Pattern of U.S.-Canada Trade," *American Economic Review*, 91(4), pp. 858-876.
- (1998), "Immigration and Trade Creation: Evidence from Canada", *Canadian Journal of Economics*, 31, 47-62.
- Heckscher, E.F. "Växelkursens Grundval vid Pappersmyntfot, *Ekonomisk Tidskrift*, October 1916, 18, 309-312.
- Helliwell, John F. (1996), "Do National Boundaries Matter for Quebec's Trade?", *Canadian Journal of Economics*, 29, pp. 507-22.
- (1997), "National Borders, Trade and Migration," *Pacific Economic Review*, 2(3), pp. 165-185.
- (1998), *How Much Do National Borders Matter?*. Washington, D.C.: The Brookings Institution.
- (1999), "Perspectives on North American Free Trade: Effects of the Canada-United States Free Trade Agreement on Interprovincial Trade," *Industry*

- Canada Research Publications Program, No. 5, April 1999.
- (2003), “Border Effects: Assessing Their Implications for Canadian Policy in a North American Context,” working paper, University of 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, Paper 5, April 1999.
- Helliwell, John F. and McCallum, John (1995), “National Borders Still Matter for Trade,” *Policy Options*, 16, pp. 44-48.
- Helliwell, John F. and Verdier, Genevieve (2001), “Measuring Internal Trade Distances: A New Method Applied to Estimate Provincial Border Effects in Canada,” *Canadian Journal of Economics*, 34(4), 1024-1041.
- Helpman, Elhanan and Paul Krugman (1985), *Market Structure and Foreign Trade*, Cambridge: MIT Press.
- Hillberry, Russell (1998), “Regional Trade and the Medicine Line: The National Border Effect in U.S. Commodity Flow Data,” *Journal of Borderlands Studies*, 8(2), pp. 1-17.
- Hillberry, Russell (2002), “Aggregation Bias, Compositional Change and the Border Effect”, *Canadian Journal of Economics*, 35, 3, 2002, 517-530.
- Hillberry, Russell and David Hummels (2002a), “Intra-National Home Bias: Some Explanations,” NBER Working Paper 9022, *Review of Economics and Statistics*, forthcoming.
- Hillberry, Russell and David Hummels (2002b), “Explaining Home Bias in Consumption: The Role of Intermediate Input Trade”, National Bureau of Economic Research (Cambridge, MA) Working Paper No. 9020.
- Hufbauer, Gary and Kimberly Ann Elliott (1994), *Measuring the Cost of Protection in the United States*, Washington: Institute for International Economics.
- Hummels, David (2001a), “Toward a Geography of Trade Costs”, working paper, Purdue University.
- (2001b), “Time as a Trade Barrier,” working paper, Purdue University.
- (1999), “Have International Transportation Costs Declined?”, working paper, Purdue University.
- Hummels, David, Jun Ishii and Kei-Mu Yi (2001), “The Nature and Growth of Vertical Specialization in World Trade,” *Journal of International Economics*, 54, 75-96.
- Kim, Namdoo (1996), *Measuring the Costs of Visible Protection in Korea*, Washington:

Institute for International Economics.

- Klenow, Peter J. and Andres Rodriguez-Clare (1997), "Quantifying Variety Gains from Trade Liberalization," working paper.
- Knetter, Michael. "Why Are Retail Prices in Japan So High?: Evidence from German Export Prices," *International Journal of Industrial Organization*, August 1997, 549-572.
- Lai, Huiwen and Daniel Treffer (2002), "The Gains from Trade with Monopolistic Competition: Specification, Estimation and Mis-specification", NBER Working Paper No. 9169.
- Laird, Sam and Alexander Yeats (1990), *Quantitative Methods for Trade-Barrier Analysis*, New York: New York University Press.
- Lancaster, Kelvin (1979), "Variety, Equity and Efficiency," New York: Columbia University Press.
- Lee, Jong-Wha and Phillip Swagel. "Trade Barriers and Trade Flows Across Countries and Industries," *The Review of Economics and Statistics*, 1995.
- Limao, Nuno and Anthony J. Venables (2001), "Infrastructure, Geographical Disadvantage, Transport Costs and Trade," *World Bank Economic Review*, 15, 451-79.
- Lopez-Cordova, Ernesto J. and Christopher M. Meissner (2003), "Exchange Rate Regimes and International Trade: Evidence from the Classical Gold Standard Era", *American Economic Review*, 93(1), 344-353.
- Melitz, Jacques (2003), "Language and Foreign Trade", working paper, University of Strathclyde.
- Messerlin, Patrick (2001), *Measuring the Cost of Protection in Europe*, Washington: Institute for International Economics.
- Millimet, Daniel L. and Thomas Osang (2001), "Do State Borders Matter for U.S. Intranational Trade? The Role of History and Internal Migration," working paper, Southern Methodist University.
- McCallum, John (1995), "National Borders Matter: Canada-U.S. Regional Trade Patterns," *American Economic Review*, 85(3), pp. 615-623.
- Minondo, Asier (2002), "Borders and Trade in Europe: A Fresh Analysis on the Extent of the Home-Bias Effect," working paper, Universidad de Deusto-ESTE, San Sebastian.
- Nitsch, Volker (2000), "National Borders and International Trade: Evidence from the European Union," *Canadian Journal of Economics* 33(4), 1091-1105.

- Obstfeld, Maurice and Kenneth Rogoff (2000), "The Six Major Puzzles in International Macroeconomics. Is There a Common Cause?", in B.S. Bernanke and K. Rogoff, eds., *NBER Macroeconomics Annual 2000*. Cambridge, MA: MIT Press, 2000, pp. 339-390.
- Obstfeld, Maurice and Alan M. Taylor. "Nonlinear Aspects of Goods-Market Arbitrage and Adjustment: Heckscher's Commodity Points Revisited," *Journal of the Japanese and International Economies*, December 1997, vol. 11, 441-479.
- O'Connell, Paul G.J. and Shang-Jin Wei (2002), "The Bigger They Are, the Harder They Fall: Retail Price Differences across U.S. cities," *Journal of International Economics* 56, 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," *Journal of International Economics*, 55(1), 87-105.
- (2002) "Currency Arrangements and Goods Market Integration: A Price Based Approach," working paper, Vanderbilt University.
- Portes, Richard and Helene Rey (2002), "The Determinants of Cross-Border Equity Transaction Flows," working paper, London Business School.
- Rauch, James E. Vitor Trindade (1999), "Ethnic Chinese Networks in International Trade", *Review of Economics and Statistics*, 84, 116-30.
- Rauch, James E. and Vitor Trindade (2002), "Information, International Substitutability, and Globalization," working paper, University of California, San Diego, *American Economic Review*, forthcoming.
- Ritschl, Albrecht and Nikolaus Wolf (2003). "Endogeneity of Currency Areas and Trade Blocs: Evidence from the Inter-War Period," working paper, Humboldt University.
- Roberts, Mark and James Tybout (1997), "The Decision to Export in Colombia: An Empirical Model of Entry with Sunk Costs," *American Economic Review*, 87, pp. 545-564.
- Rodrik, Dani (1995), "The Political Economy of Trade Policy," in G. Grossman and K. Rogoff (eds.), *Handbook of International Economics*, vol. 3, Amsterdam, North-Holland, 1995.
- Rose, Andrew K. (2003), "The Effect of Common Currencies on International Trade: A Meta-Analysis", forthcoming in Volbert Alexander, Jacques Melitz and George

- M. von Furstenberg, eds., *Monetary Unions and Hard Pegs: Effects on Trade, Financial Development and Stability*.
- Rose, Andrew K. and Eric van Wincoop (2001), "National Money as a Barrier to Trade: The Real Case for Currency Union," *American Economic Review Papers and Proceedings*, 2001, 91(2), pp. 386-390.
- Rose, Andrew K. (2000), "One Money, One Market: Estimating the Effect of Common Currencies on Trade", *Economic Policy*, 30, pp. 7-45.
- Sazanami, Yoko, Shujiro Urata and Hiroki Kawai (1995), *Measuring the Costs of Protection in Japan*, Washington: Institute for International Economics.
- Staiger, Robert W. and Frank A. Wolak (1994), "Measuring Industry Specific Protection: Antidumping in the United States", *Brookings Papers in Economic Activity: Microeconomics*, 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, 473-98.
- Tchamourliyski, Yuriy (2002), "Distance and Bilateral Trade: The Role of Non-Homothetic Preferences," working paper, Boston College.
- Trefler, Daniel (1993) "Trade Liberalization and the Theory of Endogenous Protection: An Econometric Study of U.S. Import Policy," *Journal of Political Economy* 101:1, 138-160.
- Wei, Shang-Jin (1996), "Intra-national versus Inter-national Trade: How Stubborn are Nations in Global Integration?," National Bureau of Economic Research (Cambridge, MA) Working Paper No. 5531.
- Wolf, Holger (2000a), "(Why) Do Borders Matter For Trade?," in G.D. Hess and E. van Wincoop, eds., *Intranational Macroeconomics*. Cambridge: Cambridge University Press, pp. 112-128.
- (2000b) "Intranational Home Bias in Trade," *Review of Economics and Statistics*, 82(4), 555-563.
- Yi, Kei-Mu (2003a) "Can Vertical Specialization Explain the Growth of World Trade?," *Journal of Political Economy*, 111(1), 52-102.
- (2003b). "A Simple Explanation for the Border Effect," working paper, Federal Reserve Bank of New York.
- Zhang Shuguang, Zhang Yansheng and Wan Zhongxin (1998), *Measuring the Costs of Protection in China*, Washington: Institute for International Economics.

Tables for JEL Trade Costs Survey

September 10, 2003

Table 1: Percentage of Countries with Data in TRAINS

Year	All Countries				OECD Countries		OECD Countries	
	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.

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.

Table 3: Non-Tariff Barriers - 1999

Country	NTB ratio (narrow)	TW NTB ratio (narrow)	ratioNTB 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 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

ISIC	Description	United States 1999				EU-12 1999				Japan 1996				Canada 1999			
		Narrow		Broad		Narrow		Broad		Narrow		Broad		Narrow		Broad	
		NTB Ratio		NTB Ratio		NTB Ratio		NTB Ratio		NTB Ratio		NTB Ratio		NTB Ratio		NTB Ratio	
		S	TW	S	TW	S	TW	S	TW	S	TW	S	TW	S	TW	S	TW
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.

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
Curatins 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 US import percentages are calculated using data from the UNCTAD TRAINS dataset. SIC to HS concordances from the US Census Bureau are used.

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 Netherlands, so an average of the four European countries' margins is used.

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 (1999) 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 ($\sigma = 5$) ($\sigma = 5$)	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 σ 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.

Table 8: Trade Cost Indices for US 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 USA	1.50

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

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

Sector	1991	1993				
	Quota Tar. Eq.	Quota Tar. Eq.	S Tariff	TW Tariff	Quota + TW Tar.	%US Imports
Price-gap measures Agricultural Sector ^a						
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.

Table 10: Trade Costs and Price Dispersion

	N	σ	parameters		Price Dispersion Measure					Trade Cost		
			θ	trade cost average	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 all 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.

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.

Figure 1 Elasticity of Substitution versus Distance Elasticity of Trade Costs

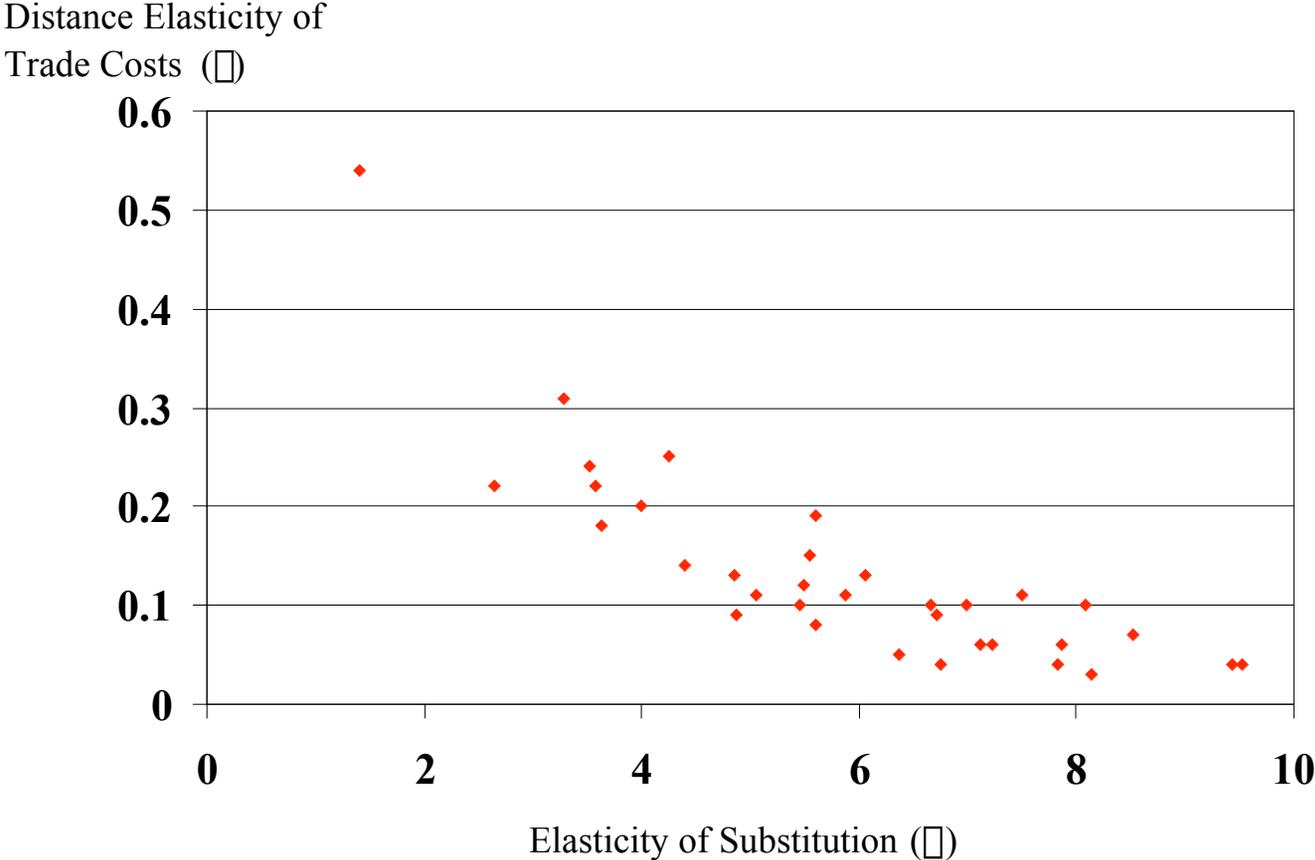
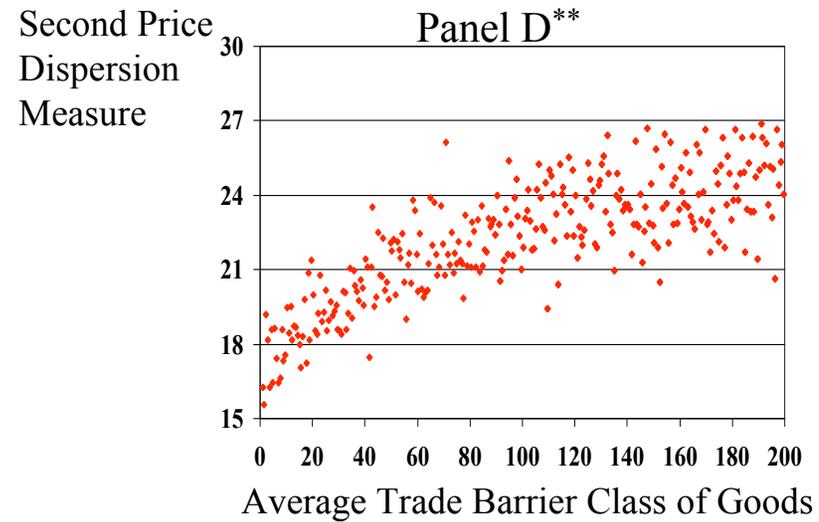
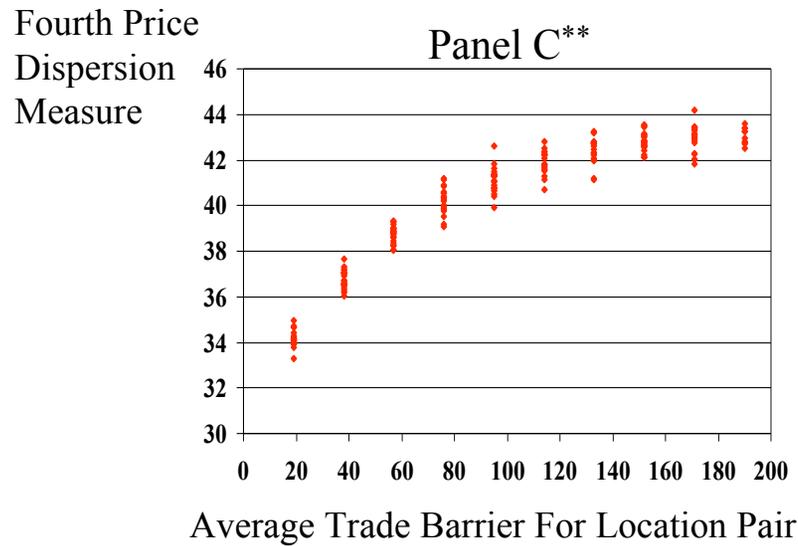
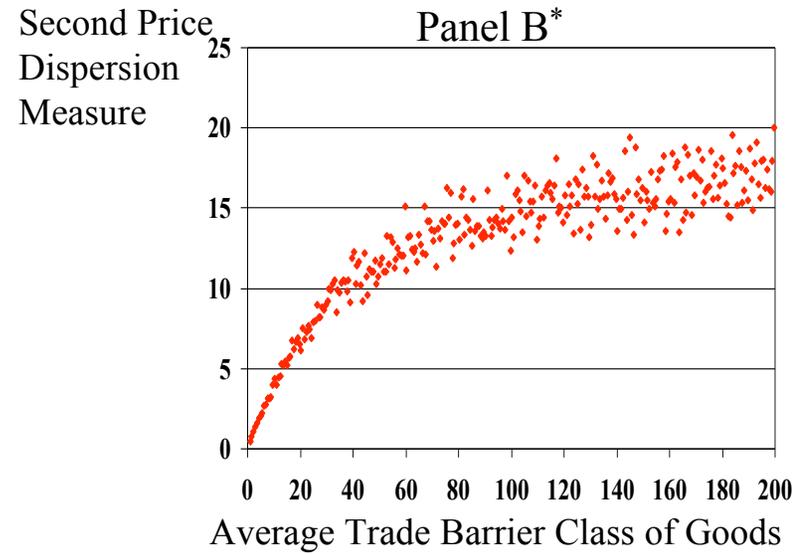
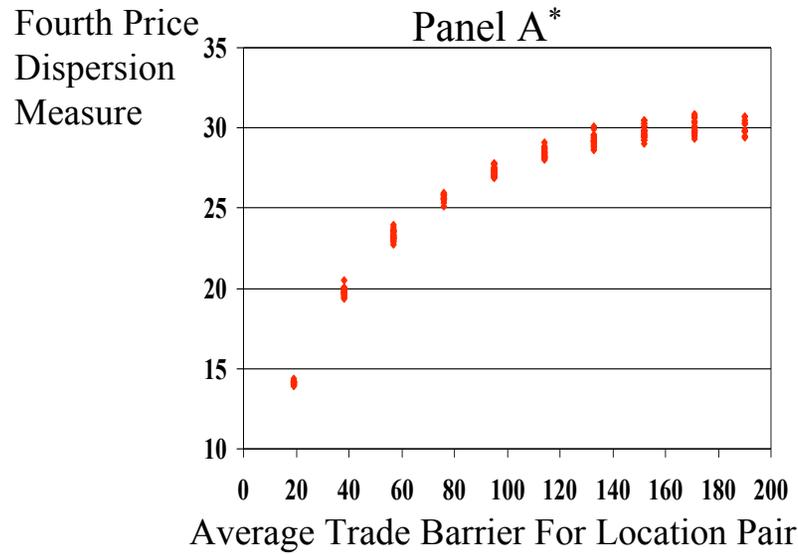


Figure 2 Price Dispersion Across Location Pairs and Goods



* Trade Costs Only Source of Price Dispersion

** Including Other Sources of Price Dispersion