Gravity Redux: Measuring
International Trade Costs with Panel Data

Dennis Novy

No 861

WARWICK ECONOMIC RESEARCH PAPERS

DEPARTMENT OF ECONOMICS

THE UNIVERSITY OF
WARWICK

Gravity Redux: Measuring International Trade Costs with Panel Data*

Dennis Novy†
University of Warwick

July 2009

Abstract

Barriers to international trade are known to be large but due to data limitations it is hard to measure them directly for a large number of countries over many years. To address this problem I derive a micro-founded measure of bilateral trade costs that indirectly infers trade frictions from observable trade data. I show that this trade cost measure is consistent with a broad range of leading trade theories including Ricardian and heterogeneous firms models. The measure implies that U.S. trade costs with major trading partners declined on average by about 40 percent between 1970 and 2000, with Mexico and Canada experiencing the biggest reductions.

JEL classification: F10, F15
Keywords: Trade Costs, Gravity, Multilateral Resistance, Ricardian Trade, Heterogeneous Firms, Panel Data

*I am grateful to participants at the 2007 NBER Summer Institute, in particular James Harrigan, David Hummels, Nuno Limão and Peter Neary. I am also grateful to Iwan Barankay, Jeffrey Bergstrand, Natalie Chen, Alejandro Cuñat, Robert Feenstra, David Jacks, Chris Meissner, Niko Wolf, Adrian Wood, seminar participants at Oxford University, the University of Western Ontario and at the European Trade Study Group. I gratefully acknowledge research support from the Economic and Social Research Council, Grant RES-000-22-3112.

†Department of Economics, University of Warwick, Coventry CV4 7AL, United Kingdom. d.novy@warwick.ac.uk and http://www2.warwick.ac.uk/fac/soc/economics/staff/faculty/novy/

1 Introduction

International trade has grown enormously over the last few decades, and almost every country trades considerably more today than thirty or forty years ago. One reason for this increase in trade has undoubtedly been the decline in international trade costs, for example the decline in transportation costs and tariffs. But which countries have experienced the fastest declines in trade costs, and how big are the remaining barriers? These questions are important for understanding what impedes globalization, yet we know surprisingly little about the barriers that prevent international market integration.

This paper sheds light on these issues by developing a way of measuring the barriers to international trade. Specifically, I derive a micro-founded measure of aggregate bilateral trade costs that I obtain from the *gravity* equation. As a workhorse model of international trade, the gravity equation relates bilateral trade to countries' economic size and bilateral trade costs. It has been used for many decades to explain the extent of bilateral trade flows, and it has one of the strongest empirical track records in economics. The core idea of the paper is to analytically solve a theoretical gravity equation for the trade cost parameters that capture the barriers to international trade. The resulting solution expresses the trade cost parameters as a function of observable trade data and thus provides a micro-founded measure of bilateral trade costs that can be tracked over time. The measure is useful in practice because it is easy to implement empirically with readily available data.

The advantage of this trade cost measure is that it captures a wide range of trade cost components. These include transportation costs and tariffs but also other components that can be difficult to observe such as language barriers, informational costs and bureaucratic red tape.[1] While it would be desirable to collect direct data on individual trade cost components at different points in time and add them up to obtain a summary measure of trade costs, this is hardly possible in practice due to severe data limitations. The trade cost measure derived in this paper avoids this problem by providing researchers with a gauge of comprehensive international trade costs that is easy to construct. It can be helpful not only for studying international trade but also for other applications that require a time-varying measure of bilateral market integration.

The approach taken in this paper has a strong theoretical foundation. I show that inferring trade costs indirectly from trade data is consistent with a large variety of leading international trade models. Specifically, I derive the trade cost measure from the well-known gravity model by Anderson and van Wincoop (2003), the Ricardian model by Eaton and Kortum (2002) as well as the heterogeneous firms models by Chaney (2008) and Melitz and Ottaviano (2008). Although these models make fundamentally different assumptions about the driving forces behind international trade, they have in common that they yield gravity equations in general equilibrium.[2] I exploit this similarity and demonstrate that all these models lead to an isomorphic trade cost

[1]For example, Anderson and Marcouiller (2002) highlight hidden transaction costs due to poor security. Portes and Rey (2005) identify costs of international information transmission.

[2]On the generality of the gravity equation also see Grossman (1998), Feenstra, Markusen and Rose (2001), Evenett and Keller (2002) and Feenstra (2004). Since the trade cost measure is derived from the gravity equation, it can be interpreted as a 'gravity residual' that compares actual trade flows to those predicted by the gravity equation for a hypothetical frictionless world. In that sense its nature is related to the literature on missing trade that juxtaposes actual and predicted trade flows (see Trefler, 1995).

measure. The intuition is that gravity equations are basic expenditure equations that indicate how consumers allocate spending across countries under the constraints of trade barriers. The motivation for purchasing foreign goods could be that they are either inherently different from domestic goods as in an Armington world, or they are produced relatively more efficiently as in a Ricardian world. I show formally that for the purpose of measuring international trade costs, it does not matter *why* consumers choose to spend money on foreign goods.

As an illustration, I take the trade cost measure to the data and compute U.S. bilateral trade costs for a number of major trading partners. First, I find that the level of trade costs in the year 2000, expressed as a tariff equivalent, is lowest for Canada at 25 percent, followed by Mexico at 33 percent. But trade costs are considerably higher for Japan and the UK at over 60 percent. While these levels are consistent with comprehensive ballpark figures in the literature, for example those reported by Anderson and van Wincoop (2004), they have the advantage of being country-pair specific. Second, I find that over the period 1970-2000, U.S. trade costs declined by about 40 percent on average, consistent with improvements in transportation and communication technology. But coinciding with the formation of NAFTA, the decline in trade costs was considerably steeper for Canada and Mexico.

There are two differences between the trade cost measure derived in this paper and traditional gravity estimation. First, as I infer aggregate trade costs indirectly from observable trade data, there is no need to assume any particular trade cost function. In contrast, every estimated gravity regression implicitly assumes such a function by relying on trade cost proxies such as geographical distance as explanatory variables. A potential problem with that approach is that many trade cost components such as non-tariff barriers might be omitted because it is hard to find empirical proxies for them. The trade cost measure in this paper avoids this problem because it captures a comprehensive set of trade barriers. As a result, the trade cost levels reported above exceed the numbers associated with individual components such as freight rates because those only represent a subset of overall trade costs. The second difference is that many typical trade cost proxies such as distance do not vary over time. A static trade cost function is therefore ill-suited to capture the variation of trade costs over time.[3] However, the measure derived in this paper is a function of time-varying observable trade data and thus allows researchers to trace changes in bilateral trade costs over time.

Finally, I use the gravity framework to examine the driving forces behind the strong growth of international trade over the last decades. I decompose the growth of bilateral trade into three distinct contributions – the growth of income, the decline of bilateral trade barriers and the decline of multilateral barriers, or *multilateral resistance* as coined by Anderson and van Wincoop (2003). I find that income growth explains the majority of U.S. trade growth over the period 1970-2000. The decline of bilateral trade barriers is the second biggest contribution but this contribution varies considerably across trading partners. For example, the decline of bilateral trade barriers is about twice as important for explaining the growth of trade with Mexico as it is for explaining the growth of trade with Japan. My results are consistent with those of Baier and Bergstrand (2001) who argue that two-thirds of the growth in trade amongst OECD countries between 1958 and 1988 can be explained by the growth of income. But the innovation of my decomposition is to explicitly account for the role of multilateral resistance. As

[3] For example, Anderson and van Wincoop (2003) only consider trade costs in cross-sectional data for the year 1993.

I obtain an analytical solution for the unobservable multilateral resistance variables, I can relate them to observable trade data. Previously it has been either impossible or very cumbersome to solve for multilateral resistance.

An alternative approach to measuring trade costs in the literature is to consider price differences across borders. This is motivated by the idea that arbitrage will eliminate price differences in the absence of international trade costs. While this approach is in principle promising, it is plagued by the difficulty of getting reliable price data on comparable goods in different countries. Another approach attempts to measure trade costs directly (see Anderson and van Wincoop, 2004, for a survey). Limão and Venables (2001) employ data on the cost of shipping a standard 40-foot container from Baltimore, Maryland, to various destinations in the world, showing that transport costs are significantly increased by poor infrastructure and adverse geographic features such as being landlocked. Hummels (2007) examines the costs of ocean shipping and air transportation. Kee, Nicita and Olarreaga (2009) propose a trade restrictiveness index that is based on observable tariff and non-tariff barriers. They show that tariffs alone are a poor indicator of trade restrictiveness since non-tariff barriers also provide a considerable degree of trade protection. I view such direct measures as complements to indirect measures that are inferred from trade flows. Direct measures have the advantage of being more precise on the particular trade cost components that they capture. But the direct approach is often restricted by data limitations and by the fact that many trade cost components are unobservable.

The gravity framework by Anderson and van Wincoop (2003) has attracted a lot of attention in the literature. Baier and Bergstrand (2009) show that in gravity applications the nonlinear multilateral resistance terms can be approximated by a log-linear Taylor-series expansion. Instead of an approximation my approach yields an analytical solution for the multilateral resistance terms that is easy to implement. Furthermore, Balistreri and Hillberry (2007) argue that Anderson and van Wincoop's (2003) solution of the border puzzle critically hinges on the assumption of bilateral trade cost symmetry. In contrast, I do not constrain bilateral trade costs to be symmetric and instead focus on the average of bilateral trade barriers in both directions. This approach accommodates underlying trade cost asymmetries.

Although I derive the trade cost measure from a wide range of leading trade models, the derivation that is based on the Anderson and van Wincoop (2003) framework is related to the 'freeness of trade' measure in the New Economic Geography literature. The freeness measure captures the inverse of trade costs so that a high value corresponds to low trade barriers (see Fujita, Krugman and Venables, 1999; Head and Ries, 2001; Baldwin, Forslid, Martin, Ottaviano and Robert-Nicoud, 2003; Head and Mayer, 2004). My paper adds to this literature by pointing out the direct link to the Anderson and van Wincoop (2003) gravity framework and by relating unobservable multilateral resistance variables to observable data. In addition, it provides the more general insight that the trade cost measure can be derived from model classes that are not typically considered in that literature.

The paper is organized as follows. In Section 2, I derive the micro-founded trade cost measure, showing that it is consistent with a wide range of leading trade models. In Section 3, I present U.S. bilateral trade costs for a number of major trading partners. In Section 4, I decompose the growth of bilateral trade into the growth of income and the decline of trade barriers. Section 5 provides a discussion of the results and a number of robustness checks. Section 6 concludes.

2 Trade Costs in General Equilibrium

In this section, I derive the micro-founded measure of bilateral trade costs. I base the derivation on the well-known Anderson and van Wincoop (2003) model. This is one of the most parsimonious trade models, which makes the derivation particularly intuitive. But in fact, the trade cost measure does not hinge on that particular model. To demonstrate that it is valid more generally I also show how the trade cost measure can be derived from two different types of trade models – the Ricardian model by Eaton and Kortum (2002) as well as the heterogeneous firms models by Chaney (2008) and Melitz and Ottaviano (2008).

2.1 Trade Costs in Anderson and van Wincoop (2003)

Anderson and van Wincoop (2003) develop a multi-country general equilibrium model of international trade. Each country is endowed with a single good that is differentiated from those produced by other countries. Optimizing individual consumers enjoy consuming a large variety of domestic and foreign goods. Their preferences are assumed to be identical across countries and are captured by constant elasticity of substitution utility.

As the key element in their model, Anderson and van Wincoop (2003) introduce exogenous bilateral trade costs. When a good is shipped from country i to j, bilateral variable transportation costs and other variable trade barriers drive up the cost of each unit shipped. As a result of trade costs, goods prices differ across countries. Specifically, if p_i is the net supply price of the good originating in country i, then $p_{ij} = p_i t_{ij}$ is the price of this good faced by consumers in country j, where $t_{ij} \geq 1$ is the gross bilateral trade cost factor (one plus the tariff equivalent).[4]

Based on this framework Anderson and van Wincoop (2003) derive a micro-founded gravity equation with trade costs:

$$x_{ij} = \frac{y_i y_j}{y^W} \left(\frac{t_{ij}}{\Pi_i P_j} \right)^{1-\sigma} \tag{1}$$

x_{ij} denotes nominal exports from i to j, y_i is nominal income of country i and y^W is world income defined as $y^W \equiv \sum_j y_j$. $\sigma > 1$ is the elasticity of substitution across goods. Π_i and P_j are country i's and country j's price indices.

The gravity equation implies that all else being equal, bigger countries trade more with each other. Bilateral trade costs t_{ij} decrease bilateral trade but they have to be measured against the price indices Π_i and P_j. Anderson and van Wincoop (2003) call these price indices *multilateral resistance* variables because they include trade costs with all other partners and can be interpreted as average trade costs. Their exact expressions are given by

$$\Pi_i^{1-\sigma} = \sum_j P_j^{\sigma-1} \theta_j t_{ij}^{1-\sigma} \quad \forall i \tag{2}$$

$$P_j^{1-\sigma} = \sum_i \Pi_i^{\sigma-1} \theta_i t_{ij}^{1-\sigma} \quad \forall j \tag{3}$$

where θ_j is the world income share of country j defined as $\theta_j \equiv y_j / y^W$. Π_i is the *outward*

[4] Modeling trade costs in this way is consistent with the iceberg formulation that portrays trade costs as if an iceberg were shipped across the ocean and partly melted in transit (e.g., Samuelson, 1954, and Krugman, 1980).

multilateral resistance variable as it includes bilateral trade costs t_{ij} summed over and weighted by all destination countries j, whereas P_j is the *inward* multilateral resistance variable as it includes bilateral trade costs t_{ij} summed over and weighted by all origin countries i. Thus, an important insight from gravity equation (1) is that bilateral trade flows depend not only on the bilateral trade barrier but also on the multilateral trade barriers of the two countries involved.

2.1.1 The Link between Multilateral Resistance and Intranational Trade

Since direct measures for appropriately averaged trade costs are generally not available, it is difficult to find expressions for the multilateral resistance variables in equations (2) and (3). Anderson and van Wincoop (2003) assume that bilateral trade costs are a function of two particular trade cost proxies – a border barrier and geographical distance. In particular, they assume the trade cost function $t_{ij} = b_{ij}d_{ij}^{\rho}$ where b_{ij} is a border-related indicator variable, d_{ij} is bilateral distance and ρ is the distance elasticity. In addition, they simplify the model by assuming that bilateral trade costs are symmetric (i.e., $t_{ij} = t_{ji}$). Under the symmetry assumption it follows that outward and inward multilateral resistance are the same (i.e., $\Pi_i = P_i$). Thus, conditioning on these additional assumptions, Anderson and van Wincoop (2003) find an implicit solution for multilateral resistance based on (2) and (3).

There are a number of drawbacks associated with the additional assumptions.[5] First, the chosen trade cost function might be misspecified. Its functional form might be incorrect and it might omit important trade cost determinants such as tariffs. Second, bilateral trade costs might be asymmetric, for example if one country imposes higher tariffs than the other. Third, in practice trade barriers are time-varying, for example when countries phase out tariffs. Time-invariant trade cost proxies such as distance are therefore hardly useful in capturing trade cost changes over time.[6]

In what follows, I propose a method that helps to overcome these drawbacks by deriving an *analytical* solution for multilateral resistance variables. This method does not rely on any particular trade cost function and it does not impose trade cost symmetry. Instead, trade costs are inferred from time-varying trade data that are readily observable.

Intuitively, my method makes use of the insight that a change in bilateral trade barriers does not only affect *international* trade but also *intranational* trade. For example, suppose that country i's trade barriers with all other countries fall. In that case, some of the goods that country i used to consume domestically, i.e., intranationally, are now shipped to foreign countries. It is therefore not only the extent of international trade that depends on trade barriers with the rest of the world but also the extent of intranational trade.

This can be seen formally by using gravity equation (1) to find an expression for country i's intranational trade

$$x_{ii} = \frac{y_i y_i}{y^W}\left(\frac{t_{ii}}{\Pi_i P_i}\right)^{1-\sigma} \tag{4}$$

where t_{ii} represents intranational trade costs, for example domestic transportation costs. Equa-

[5] Anderson and van Wincoop (2003, p. 180) provide a brief discussion on this point.

[6] Combes and Lafourcade (2005) show that although distance is a good proxy for transport costs in cross-sectional data, it is of very limited use for time series data.

tion (4) can be solved for the product of outward and inward multilateral resistance as

$$\Pi_i P_i = \left(\frac{x_{ii}/y_i}{y_i/y^W}\right)^{\frac{1}{(\sigma-1)}} t_{ii} \tag{5}$$

As an example suppose two countries i and j face the same domestic trade costs $t_{ii} = t_{jj}$ and are of the same size $y_i = y_j$ but country i is a more closed economy, that is, $x_{ii} > x_{jj}$. It follows directly from (5) that multilateral resistance is higher for country i ($\Pi_i P_i > \Pi_j P_j$). Equation (5) implies that for given t_{ii} it is easy to measure the change in multilateral resistance over time as it does not depend on time-invariant trade cost proxies such as distance.

2.1.2 A Micro-Founded Measure of Trade Costs

The explicit solution for the multilateral resistance variables can be exploited to solve the model for bilateral trade costs. Gravity equation (1) contains the product of outward multilateral resistance of one country and inward multilateral resistance of another country, $\Pi_i P_j$, whereas equation (5) provides a solution for $\Pi_i P_i$. It is therefore useful to multiply gravity equation (1) by the corresponding gravity equation for trade flows in the opposite direction, x_{ji}, to obtain a bidirectional gravity equation that contains both countries' outward and inward multilateral resistance variables:

$$x_{ij}x_{ji} = \left(\frac{y_i y_j}{y^W}\right)^2 \left(\frac{t_{ij}t_{ji}}{\Pi_i P_i \Pi_j P_j}\right)^{1-\sigma} \tag{6}$$

Substituting the solution from equation (5) yields

$$x_{ij}x_{ji} = x_{ii}x_{jj} \left(\frac{t_{ii}t_{jj}}{t_{ij}t_{ji}}\right)^{\sigma-1} \tag{7}$$

The size variable in gravity equation (7) is not total income $y_i y_j$ as in traditional gravity equations but intranational trade $x_{ii}x_{jj}$. Intranational trade does not only control for the countries' economic size, but according to equation (5) it is also directly linked to multilateral resistance. (7) can be rearranged as

$$\frac{t_{ij}t_{ji}}{t_{ii}t_{jj}} = \left(\frac{x_{ii}x_{jj}}{x_{ij}x_{ji}}\right)^{\frac{1}{\sigma-1}}$$

As shipping costs between i and j can be asymmetric ($t_{ij} \neq t_{ji}$) and as domestic trade costs can differ across countries ($t_{ii} \neq t_{jj}$), it is useful to take the geometric mean of the barriers in both directions. It is also useful to deduct one to get an expression for the tariff equivalent. I denote the resulting trade cost measure as τ_{ij}:

$$\tau_{ij} \equiv \left(\frac{t_{ij}t_{ji}}{t_{ii}t_{jj}}\right)^{\frac{1}{2}} - 1 = \left(\frac{x_{ii}x_{jj}}{x_{ij}x_{ji}}\right)^{\frac{1}{2(\sigma-1)}} - 1 \tag{8}$$

τ_{ij} measures bilateral trade costs $t_{ij}t_{ji}$ relative to domestic trade costs $t_{ii}t_{jj}$. It therefore does not impose frictionless domestic trade and captures what makes international trade more costly over and above domestic trade.[7]

The intuition behind τ_{ij} is straightforward. If bilateral trade flows $x_{ij}x_{ji}$ increase relative to domestic trade flows $x_{ii}x_{jj}$, it must have become easier for the two countries to trade with each other. This is captured by a decrease in τ_{ij}, and vice versa. The measure thus captures trade costs in an indirect way by inferring them from observable trade flows. Since these trade flows vary over time, trade costs τ_{ij} can be computed not only for cross-sectional data but also for time series and panel data. This is an advantage over the procedure adopted by Anderson and van Wincoop (2003) who only use cross-sectional data.

It is important to stress that bilateral barriers might be asymmetric ($t_{ij} \neq t_{ji}$) and that bilateral trade flows might be unbalanced ($x_{ij} \neq x_{ji}$). τ_{ij} indicates the geometric average of the bilateral trade barriers in both directions.

Finally, the model above and thus the trade cost measure τ_{ij} can also be motivated by a Heckscher-Ohlin setting. Deardorff (1998) argues that whenever there are bilateral trade barriers, the Heckscher-Ohlin model cannot have factor price equalization between two countries that trade with each other. If factor prices were equalized, prices would also be equalized and neither country could overcome the trade barriers. In a world with a large number of goods and few factors it is therefore likely that one country will be the lowest-cost producer and that trade in a Heckscher-Ohlin world resembles trade in an Armington world.

2.2 Trade Costs in a Ricardian Model

Whereas the Anderson and van Wincoop (2003) model is a demand-side model that takes production as exogenous, the Ricardian model by Eaton and Kortum (2002) emphasizes the supply side. Each country can potentially produce every single good on the global range of goods but there will be only one lowest-cost producer who serves all other countries, provided that the cross-country price differential exceeds variable bilateral trade costs t_{ij}. Eaton and Kortum (2002) thus introduce an extensive margin of trade.

Productivity in each country is drawn from a Fréchet distribution. The parameter T_i determines the average absolute productivity advantage of country i, with a high T_i denoting high overall productivity. The parameter $\vartheta > 1$ governs the variation of the productivity distribution and is treated as common across countries, with a low ϑ denoting much variation and thus much scope for comparative advantage. The model yields a gravity-like equation for aggregate trade flows. It is given by

$$x_{ij} = \frac{T_i \left(c_i t_{ij}\right)^{-\vartheta}}{\sum_{i=1}^{J} T_i \left(c_i t_{ij}\right)^{-\vartheta}} y_j \tag{9}$$

where c_i denotes the input cost in country i and y_j is total expenditure of destination country j.

Since c_i and T_i are generally unknown, it is not possible to isolate the individual trade cost parameter t_{ij} from equation (9) in terms of observable variables. However, following the same

[7] τ_{ij} can also be interpreted as a measure of the international component of trade costs net of distribution trade costs in the destination country. Formally, suppose total gross shipping costs t_{ij} can be decomposed into gross shipping costs up to the border of j, denoted by t_{ij}^*, times the gross shipping costs within j, denoted by t_{jj}, where t_{jj} does not depend on the origin of shipment. It follows $t_{ij} = t_{ij}^* t_{jj}$ and $t_{ji} = t_{ji}^* t_{ii}$ so that $\tau_{ij} = \sqrt{t_{ij}^* t_{ji}^*} - 1$.

approach as in equation (8) I can relate the combination of bilateral and domestic trade cost parameters to the ratio of domestic trade, $x_{ii}x_{jj}$, over bilateral trade, $x_{ij}x_{ji}$. This yields

$$\tau_{ij}^{EK} = \left(\frac{t_{ij}t_{ji}}{t_{ii}t_{jj}}\right)^{\frac{1}{2}} - 1 = \left(\frac{x_{ii}x_{jj}}{x_{ij}x_{ji}}\right)^{\frac{1}{2\vartheta}} - 1 \tag{10}$$

The trade cost measure τ_{ij}^{EK} is thus isomorphic to τ_{ij} in equation (8) with ϑ corresponding to $\sigma - 1$. The Ricardian model therefore implies virtually the same trade cost measure. Since trade is driven by comparative advantage, the sensitivity of the implied trade costs τ_{ij}^{EK} to trade flows depends on the heterogeneity in countries' relative productivities, determined by ϑ. But in Anderson and van Wincoop's (2003) consumption-based model, where trade is driven by love of variety, the sensitivity depends on the degree of production differentiation, determined by σ.[8]

2.3 Trade Costs in Heterogeneous Firms Models

Turning to an additional class of models, I consider the trade theories with heterogeneous firms by Chaney (2008) and Melitz and Ottaviano (2008). Firms have different levels of productivity, depending on their draws from a Pareto distribution with shape parameter γ.

Chaney (2008) builds on the seminal paper by Melitz (2003) where each firm produces a unique product but faces bilateral fixed costs of exporting, f_{ij}. He derives the following aggregate gravity equation:

$$x_{ij} = \mu \frac{y_i y_j}{y^W} \left(\frac{w_i t_{ij}}{\lambda_j}\right)^{-\gamma} (f_{ij})^{-\left(\frac{\gamma}{\sigma-1}-1\right)} \tag{11}$$

where μ is the weight of differentiated goods in the consumer's utility function, w_i is workers' productivity in country i and λ_j is a remoteness variable akin to multilateral resistance.[9] Once again, I can relate the combination of bilateral and domestic trade cost parameters to the ratio of domestic and bilateral trade flows to obtain

$$\tau_{ij}^{Ch} = \left(\frac{t_{ij}t_{ji}}{t_{ii}t_{jj}}\right)^{\frac{1}{2}} \left(\frac{f_{ij}f_{ji}}{f_{ii}f_{jj}}\right)^{\frac{1}{2}\left(\frac{1}{\sigma-1}-\frac{1}{\gamma}\right)} - 1 = \left(\frac{x_{ii}x_{jj}}{x_{ij}x_{ji}}\right)^{\frac{1}{2\gamma}} - 1 \tag{12}$$

The trade cost measure τ_{ij}^{Ch} captures both variable and fixed trade costs. Its sensitivity to trade flows depends on the productivity distribution parameter γ that governs the entry and exit of firms into export markets.[10]

[8] See Eaton and Kortum (2002, footnote 20) for more details on the similarities between the Ricardian model and theories based on the Armington assumption.

[9] The gravity equation implicitly assumes that the economy can be modeled as having only one sector of differentiated products. This can easily be extended to multiple sectors.

[10] For the case of non-zero trade flows, the heterogeneous firms model by Helpman, Melitz and Rubinstein (2008) is consistent with the same trade cost measure, that is, $\tau_{ij}^{HMR} = \tau_{ij}^{Ch}$. In their notation, non-zero trade flows imply $V_{ij} > 0$. Additional assumptions to obtain this result are: the existence of positive fixed costs for domestic sale, $f_{ii} > 0$, the possibility of positive domestic variable trade costs, $t_{ii} \geq 1$, and, as in Appendix II of their paper, no upper bound in the support of the productivity distribution, $a_L = 0$.

Melitz and Ottaviano (2008) use non-CES preferences that give rise to endogenous markups. Heterogeneous firms face sunk costs of market entry f_E that can be interpreted as product development and production start-up costs. When exporting, the firms only face variable costs and no fixed costs of exporting. They yield the following gravity equation:

$$x_{ij} = \frac{1}{2\delta(\gamma+2)} N_i^E \psi_i L_j \left(c_j^d\right)^{\gamma+2} (t_{ij})^{-\gamma} \tag{13}$$

where δ is a parameter from the utility function that indicates the degree of product differentiation. N_i^E is the number of entrants in country i. ψ_i is an index of comparative advantage in technology. L_j denotes the number of consumers in country j. c_j^d is the marginal cost cut-off above which domestic firms in country j do not produce. As above, the only bilateral variable in equation (13) is the trade cost factor t_{ij}. All other variables are country-specific and therefore drop out when the ratio of domestic to bilateral trade flows is considered. Thus,

$$\tau_{ij}^{MO} = \left(\frac{t_{ij}t_{ji}}{t_{ii}t_{jj}}\right)^{\frac{1}{2}} - 1 = \left(\frac{x_{ii}x_{jj}}{x_{ij}x_{ji}}\right)^{\frac{1}{2\gamma}} - 1 \tag{14}$$

The trade cost measure τ_{ij}^{MO} is exactly the same function of observable trade flows as τ_{ij}^{Ch}. The difference in interpretation is that fixed costs do not enter τ_{ij}^{MO} because firms only face variable costs of exporting.

2.4 Summary

The four measures τ_{ij}, τ_{ij}^{EK}, τ_{ij}^{Ch} and τ_{ij}^{MO} have in common that they scale the ratio of domestic over bilateral trade flows by parameters that indicate a particular form of heterogeneity. A low σ in equation (8) indicates a high degree of differentiation across products ; a low ϑ in equation (10) indicates a high variation of productivity; and a low γ in equations (12) and (14) indicates a high degree of firm heterogeneity.

All four measures imply that higher heterogeneity corresponds to higher trade frictions.[11] The intuition is that higher heterogeneity provides a larger incentive to trade. If heterogeneity is high but international trade flows are small, it must be the case that international integration is impeded by large trade barriers.

3 U.S. Trade Costs

As an illustration of the trade cost measure τ_{ij} derived in the previous section, I compute U.S. bilateral trade costs for a number of major trading partners. I focus on how these bilateral trade costs have evolved over time using annual data for 1970-2000.

All bilateral aggregate trade data are taken from the IMF Direction of Trade Statistics (DOTS) and denominated in U.S. dollars. Data for intranational trade x_{ii} are not directly available but can be constructed following the approach by Shang-Jin Wei (1996). Due to market clearing intranational trade can be expressed as total income minus total exports, $x_{ii} =$

[11] This is true if the ratio of domestic over bilateral trade is larger than one, which is generically the case in the data.

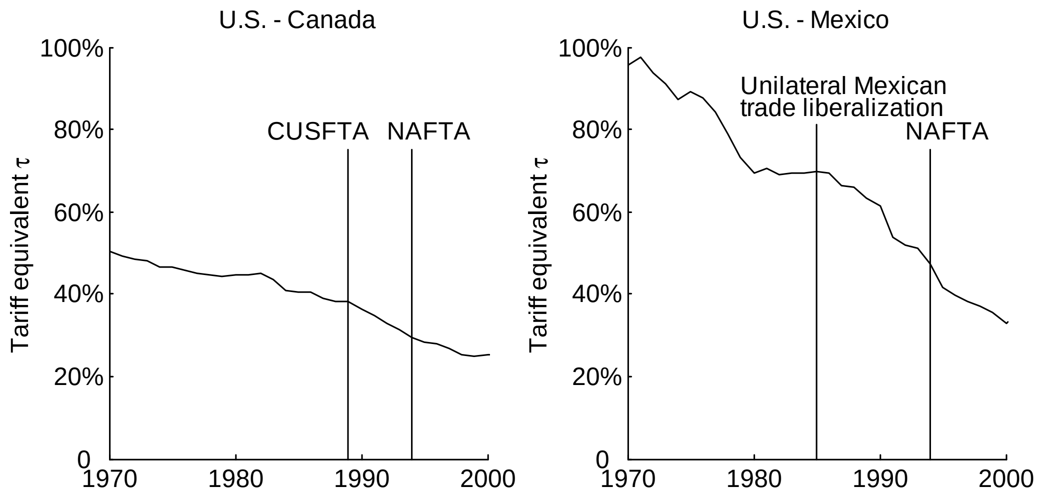

Figure 1: U.S. bilateral trade costs with Canada and Mexico.

$y_i - x_i$, where total exports x_i are defined as the sum of all exports from country i, $x_i \equiv \sum_{j \neq i} x_{ij}$.[12] However, GDP data are not suitable as income y_i because they are based on value added, whereas the trade data are reported as gross shipments. Moreover, GDP data include services that are not covered by the trade data.[13] To get the gross shipment counterpart of GDP excluding services I follow Wei (1996) in constructing y_i as total goods production based on the OECD's Structural Analysis (STAN) database.[14] The production data are converted into U.S. dollars by the period average exchange rate taken from the IMF International Financial Statistics (IFS).

Since the trade cost measure can be derived from various models (see equations 8, 10, 12 and 14), it potentially depends on different parameters, namely the elasticity of substitution σ, the Fréchet parameter ϑ and the Pareto parameter γ. Anderson and van Wincoop (2004) survey estimates of σ and conclude that it typically falls in the range of 5 to 10. Eaton and Kortum (2002) report their baseline estimate for ϑ as 8.3.[15] Helpman, Melitz and Yeaple (2004, Figure 3) estimate $\gamma - (\sigma - 1)$ to be around unity, which implies $\gamma \approx \sigma$. Chaney (2008) estimates $\gamma / (\sigma - 1)$ as roughly equal to 2, which suggests a higher value for γ, but Del Gatto, Mion and Ottaviano (2007) estimate magnitudes of γ that are lower. Given these estimates I proceed by following Anderson and van Wincoop (2004) in setting $\sigma = 8$, which corresponds to $\vartheta, \gamma = 7$.[16] This can be seen as a ballpark parameter value suitable for aggregate trade flows. As I discuss in Section 5, the overall results are not sensitive to this particular value.

Figure 1 illustrates U.S. bilateral trade costs with its two biggest trading partners, Canada and Mexico. U.S. trade costs fell dramatically with Mexico (from 96 to 33 percent) and also with

[12]See equation (8) in Anderson and van Wincoop (2003).

[13]Anderson (1979) acknowledges nontradable services and models the spending on tradables as ϕy_i, where ϕ is the fraction of total income spent on tradables. But ϕy_i would still be based on value added.

[14]Wei (1996) uses production data for agriculture, mining and total manufacturing. Also see Nitsch (2000).

[15]This estimate is based on trade data and falls in the middle of the range of estimates based on other data. They estimate $\vartheta = 12.9$ based on price data and $\vartheta = 3.6$ based on wage data.

[16]The exponent of the ratio of domestic to bilateral trade flows in equation (8) is $1/(2(\sigma - 1))$, which corresponds to $1/(2\vartheta)$ and $1/(2\gamma)$ in equations (10), (12) and (14).

Canada (from 50 to 25 percent). The U.S. experienced a clear downward trend in trade costs with both its neighbors already prior to the North American Free Trade Agreement (NAFTA, effective from 1994), the Canada-U.S. Free Trade Agreement (CUSFTA, effective from 1989) and unilateral Mexican trade liberalization (from 1985).[17]

It is important to stress that these numbers represent bilateral relative to domestic trade costs. For example, take the result that U.S.-Canadian trade costs are 25 percent in the year 2000. Suppose that a particular good produced in either the U.S. or Canada costs $10.00 at the factory gate and abstract from possible fixed costs of exporting.[18] Also suppose that domestic wholesale and retail distribution costs are 55 percent (t_{ii}=1.55), which is the representative domestic distribution cost across OECD countries as reported by Anderson and van Wincoop (2004). A domestic consumer could therefore buy the product for $15.50, whereas a consumer abroad would have to pay $19.40 ($t_{ij}$=1.94=1.55*1.25). This example illustrates that the absolute domestic trade costs ($5.50=$15.50-$10.00) can be substantially bigger than the absolute cost of crossing the border ($3.90=$19.40-$15.50). Of course, this particular example is based on an aggregate average and should be interpreted as such. In practice, trade costs can vary considerably across goods. For instance, perishable goods are more likely to be transported by air freight instead of less expensive truck or ocean shipping.

Table 1 reports the levels and the percentage decline in U.S. bilateral trade costs between 1970 and 2000 with its six biggest export markets as of 2000. In descending order these are Canada, Mexico, Japan, the UK, Germany and Korea.[19] The decline has been most dramatic with Mexico and Canada and has been sizeable with Korea, the UK, Germany and Japan. The trade-weighted average of U.S. trade costs declined by 44 percent between 1970 and 2000, corresponding to an annualized decline of 1.9 percent per year.[20] Its 2000 level stands at 42 percent.

The magnitudes of the bilateral trade costs in Table 1 are entirely consistent with cross-sectional evidence from the literature. For the year 1993 Anderson and van Wincoop (2004) report a 46 percent tariff equivalent of overall U.S.-Canadian trade costs, compared to 31 percent in Figure 1.[21] The reason why the number reported by Anderson and van Wincoop (2004) is somewhat higher is that they use GDP data as opposed to production data to compute trade costs. In fact, when using GDP data I obtain U.S.-Canadian trade costs of 47 percent for 1993, almost exactly the 46 percent value reported by Anderson and van Wincoop (2004).[22] But GDP data tend to overstate the extent of intranational trade and thus the level of trade costs

[17]As pointed out earlier, τ_{ij} is related to the 'freeness of trade' measure in the New Economic Geography literature, see Fujita, Krugman and Venables (1999). For a plot of the inverse freeness measure in a two-country model, see Figure 2 in Head and Ries (2001).

[18]In equation (12) this would mean $f_{ij} = f \;\; \forall \;\; i, j$ so that the fixed costs drop out of the expression for τ_{ij}^{Ch}.

[19]These six countries are those for which the 2000 share of U.S. exports exceeded 3 percent. Between 1970 and 2000 their combined share of U.S. exports fluctuated between 43 and 58 percent.

[20]$x = -0.019$ is the solution to $42 = 74^*(1 + x)^{30}$.

[21]Anderson and van Wincoop (2004) calculate the tariff equivalent as the trade-weighted average barrier for trade between U.S. states and Canadian provinces relative to the trade-weighted average barrier for trade within the United States and Canada, using a trade cost function that includes a border-related dummy variable and distance.

[22]For $\sigma = 5$ and $\sigma = 10$ Anderson and van Wincoop (2004, Table 7) report 1993 U.S.-Canadian trade cost tariff equivalents of 91 and 35 percent, respectively. The corresponding numbers based on (8) are 97 and 35 percent when using GDP data and 61 and 24 percent when using production data.

Table 1: U.S. Bilateral Trade Costs

Partner country	*Tariff equivalent τ in %* **1970**	**2000**	*Percentage change*
CANADA	50	25	−50
GERMANY	95	70	−26
JAPAN	85	65	−24
KOREA	107	70	−35
MEXICO	96	33	−66
UK	95	63	−34
Plain average	88	54	−38
Trade-weighted average	74	42	−44

All numbers are in percent and rounded off to integers.
Countries listed are the six biggest U.S. export markets as of 2000.
Computations based on equation (8).

because they include services.[23] I therefore prefer to follow Wei (1996) in using merchandise production data to match the trade data more accurately. Eaton and Kortum (2002) report bilateral tariff equivalents based on data for 19 OECD countries in 1990. For countries that are 750-1500 miles apart, an elasticity of substitution of $\sigma = 8$ implies a trade cost range of 58-78 percent, consistent with the magnitudes in Table 1.

It is important to point out that the trade cost measure τ_{ij} captures not only trade costs in the narrow sense of transportation costs and tariffs but also trade cost components such as language barriers and currency barriers. In their survey of trade costs, Anderson and van Wincoop (2004) show that such non-tariff barriers are substantial. They suggest that U.S. transport costs on their own constitute a tariff equivalent of only 10.7 percent on average, a value which is substantially lower than the numbers in Table 1. Likewise, world average c.i.f./f.o.b. ratios reported by the IMF only stand around 3 percent for the year 2000. Kee, Nicita and Olarreaga (2009) compute trade restrictiveness indices that are based on tariffs and non-tariff barriers such as import quotas, subsidies and antidumping duties. The tariff equivalent of the U.S. trade restrictiveness index is 29 percent, which is also slightly below the U.S. average in Table 1.

In summary, the trade cost measure τ_{ij} can be constructed for individual country pairs with minimal data requirements. Its main advantage over previous trade cost measures is that it can be easily tracked over time since it does not depend on time-invariant trade cost proxies such as geographical distance. Figure 1 and Table 1 demonstrate that inferred trade costs are large but generally experienced a substantial decline between 1970 and 2000. They exhibit considerable heterogeneity across country pairs that would be masked by a one-fits-all measure of trade costs.[24]

[23] Specifically, intranational trade is given by $x_{ii} = y_i - x_i$. As GDP data include services and as the service share of GDP has continually grown, the use of GDP data for y_i overstates x_{ii} compared to the use of production data despite the fact that imported intermediate goods are included in the trade data (see Helliwell, 2005). Novy (2007) develops a trade cost model with nontradable goods, showing that only the tradable part of output enters the model's micro-founded gravity equation.

[24] For a comparison of the period 1950-2000 to the period 1870-1913 see Jacks, Meissner and Novy (2008).

4 Decomposing the Growth of Trade

Bilateral trade has grown strongly between most countries in recent decades. It is an important question whether this increase in trade is simply the result of secular economic growth or whether the increase can be related to reductions in trade frictions. The gravity equation together with the trade cost measure τ_{ij} provide a simple analytical framework to address this question. I will use the gravity model by Anderson and van Wincoop (2003) for the exposition, but I refer to the Technical Appendix where I show that the growth of trade can be similarly decomposed by using the other gravity equations described in Section 2.

As the first step I take the natural logarithm and then the first difference of equation (6). This yields

$$\Delta \ln (x_{ij}x_{ji}) = 2\Delta \ln \left(\frac{y_i y_j}{y^W}\right) + (1-\sigma)\Delta \ln (t_{ij}t_{ji}) - (1-\sigma)\Delta \ln (\Pi_i P_i \Pi_j P_j) \tag{15}$$

Equation (15) relates the growth of bilateral trade, $\Delta \ln (x_{ij}x_{ji})$, to three driving forces: the growth of the two countries' economies relative to world output, changes in bilateral trade costs, $\Delta \ln (t_{ij}t_{ji})$, and changes in the two countries' multilateral trade barriers, $\Delta \ln (\Pi_i P_i \Pi_j P_j)$. The bilateral trade cost factors $t_{ij}t_{ji}$ are unknown. But we know from equation (8) that the trade cost measure τ_{ij} provides an expression for $t_{ij}t_{ji}$ relative to domestic trade costs $t_{ii}t_{jj}$ as a function of observable trade flows. I therefore substitute τ_{ij} into equation (15) to obtain

$$\Delta \ln (x_{ij}x_{ji}) = 2\Delta \ln \left(\frac{y_i y_j}{y^W}\right) + 2(1-\sigma)\Delta \ln (1+\tau_{ij}) - 2(1-\sigma)\Delta \ln (\Phi_i \Phi_j)$$

where Φ_i is shorthand for country i's multilateral resistance relative to domestic trade costs,

$$\Phi_i = \left(\frac{\Pi_i P_i}{t_{ii}}\right)^{\frac{1}{2}}$$

Finally, I divide by the left-hand side to arrive at the following bilateral decomposition equation:

$$100\% = \underbrace{\frac{2\Delta \ln \left(\frac{y_i y_j}{y^W}\right)}{\Delta \ln (x_{ij}x_{ji})}}_{\text{(a)}} + \underbrace{\frac{2(1-\sigma)\Delta \ln (1+\tau_{ij})}{\Delta \ln (x_{ij}x_{ji})}}_{\text{(b)}} - \underbrace{\frac{2(1-\sigma)\Delta \ln (\Phi_i \Phi_j)}{\Delta \ln (x_{ij}x_{ji})}}_{\text{(c)}} \tag{16}$$

Equation (16) decomposes the growth of bilateral trade into three contributions: (a) the contribution of income growth, (b) the contribution of the decline in bilateral trade costs, and (c) the contribution of the decline in multilateral resistance.[25] For example, if all bilateral trade barriers were constant over time, then contribution (b) would be zero and the

[25] Baier and Bergstrand (2001) further decompose the product of incomes, $y_i y_j$, into income shares and the sum of incomes. Define the bilateral income share as $s_i = y_i/(y_i + y_j)$. It follows $y_i y_j = s_i s_j (y_i + y_j)^2$ and thus $\Delta \ln (y_i y_j) = \Delta \ln (s_i s_j) + 2\Delta \ln (y_i + y_j)$. $\Delta \ln (s_i s_j)$ could then be interpreted as the contribution of income convergence. Also see Helpman (1987), Hummels and Levinsohn (1995) and Debaere (2005). However, after controlling for tariff cuts and transport cost reductions Baier and Bergstrand (2001) find virtually no effect of income convergence on trade growth.

growth of trade would be driven by the growth of income. But if bilateral trade costs fall (i.e., $\Delta \ln(1+\tau_{ij}) < 0$), then contribution (b) becomes positive.[26] If multilateral trade barriers fall (i.e., $\Delta \ln(\Phi_i \Phi_j) < 0$), then contribution (c) becomes negative. This negative contribution can be interpreted as a trade diversion effect. If trade barriers with other countries fall, trade with those countries increases but bilateral trade between i and j decreases.

To be clear about the approach, I do not estimate equation (16). Instead, I decompose the growth of bilateral trade conditional on the theoretical gravity framework. Contribution (a) is given by the data. Contribution (b) is also given by the data through equation (8). Likewise, contribution (c) is given by the solution for multilateral resistance in equation (5).[27] The purpose is to uncover whether trade growth is mainly associated with income growth, declining bilateral trade barriers or changes in multilateral barriers.

As I show in the Technical Appendix, decomposition equations very similar to equation (16) can be derived from the models by Eaton and Kortum (2002), Chaney (2008) and Melitz and Ottaviano (2008). The quantitative contributions of income growth (a), declining bilateral trade costs (b) and multilateral factors (c) turn out exactly the same. But the interpretation of components (b) and (c) slightly differs from model to model. For example, in the heterogeneous firms model by Chaney (2008) components (b) and (c) capture not only variable trade costs but also fixed trade costs.

4.1 Decomposing the Growth of U.S. Trade

I apply equation (16) to decompose the growth of U.S. bilateral trade. As in Table 1, I consider the six biggest U.S. export markets as of 2000. Table 2 reports the decomposition results.

Table 2 shows that for the period 1970 to 2000 the growth of income can explain more than half of the growth of U.S. bilateral trade. Income growth can explain almost all of the trade growth with Korea (92.3 percent) but only just over 50 percent with Mexico and the UK. The decline of bilateral trade costs on average provides the second most important contribution to the growth of bilateral trade. This contribution is biggest for Mexico (57.4 percent) and smallest for Japan (28.3 percent).

The decline of multilateral trade barriers diverts trade away from the U.S. Take the example of Korea. Korean trade barriers with other countries dropped considerably over time so that the diversion effect is relatively strong for Korea (-25.8 percent). The decline in multilateral resistance partially offsets the effect of declining bilateral trade costs so that the overall role of trade costs ($33.5 - 25.8 = 7.7$ percent) is modest compared to other countries in the sample.

The multilateral resistance effect is actually slightly positive for the UK ($+0.3$ percent). This means that on average multilateral trade barriers for the UK increased over time, making trade with the U.S. relatively more attractive. This result is particular to the UK as a major former colonial power since the UK's traditionally strong trade relationships with former colonies such

[26] Recall $\sigma > 1$. To be precise, a fall in bilateral trade costs also leads to a slight fall in $\Phi_i \Phi_j$ because multilateral resistance is a weighted average of all bilateral trade costs, see equations (2) and (3). Since the fall in $\Phi_i \Phi_j$ works against the effect of falling bilateral trade costs, contribution (b) in principle overstates their effect but in practice the overstatement is negligible.

[27] Equation (8) implies $2(1-\sigma)\Delta \ln(1+\tau_{ij}) = \Delta \ln(x_{ij}x_{ji}) - \Delta \ln(x_{ii}x_{jj})$. Equation (5) implies $2(1-\sigma)\Delta \ln(\Phi_i \Phi_j) = \Delta \ln\left(\frac{y_i/y^W}{x_{ii}/y_i}\right) + \Delta \ln\left(\frac{y_j/y^W}{x_{jj}/y_j}\right)$. Note that the decomposition does not depend on the value of the elasticity of substitution σ even if it changes over time.

Table 2: Decomposing the Growth of U.S. Bilateral Trade

Partner country	*Growth in trade*	Contribution of the growth in income	Contribution of the decline in bilateral trade costs	Contribution of the decline in multilateral resistance	Total
CANADA	609	65.3	+ 42.3	− 7.6	= 100
GERMANY	526	67.1	+ 36.4	− 3.5	= 100
JAPAN	580	79.3	+ 28.3	− 7.6	= 100
KOREA	832	92.3	+ 33.5	− 25.8	= 100
MEXICO	944	54.8	+ 57.4	− 12.2	= 100
UK	578	55.9	+ 43.8	+ 0.3	= 100

Growth between 1970 and 2000. All numbers in percent.

Countries listed are the six biggest U.S. export markets as of 2000.

Computations based on equation (16). Also see the Technical Appendix.

as Australia and New Zealand became weaker over time.[28]

In summary, Table 2 demonstrates that income growth is the biggest driving force behind the increase in bilateral U.S. trade. This result is consistent with the findings of Baier and Bergstrand (2001) who argue that two-thirds of the growth in trade amongst OECD countries between 1958 and 1988 can be explained by the growth of income.[29] But the innovation of decomposing the growth of trade with equation (16) is to explicitly take multilateral trade barriers into account. They are important because in general equilibrium, the trade flows between any two countries are affected both by bilateral and multilateral trade barriers.[30]

5 Discussion

A comprehensive trade cost measure The trade cost measure in equation (8) is comprehensive since it captures a wide range of trade cost components such as transportation costs and tariffs, but also components that are not directly observable such as the costs associated with language barriers and red tape. It should therefore be regarded as an upper bound that captures all trade cost elements that make international trade more costly over and above domestic trade. Instead, direct measures of specific trade cost components can be seen as a lower bound of trade costs, for example international transportation costs reported by Hummels (2007). As discussed in Section 3, U.S. transport costs correspond to a tariff equivalent of around 10 percent on average, which is roughly a quarter of the average trade cost measure for the U.S. in 2000 in Table 1. Average c.i.f./f.o.b. ratios are typically even lower. The trade restrictiveness indices by Kee, Nicita and Olarreaga (2009), which capture both tariff and non-tariff barriers,

[28] Novy (2007) shows that the trade-enhancing effect of a former colonial relationship was strong in 1970 but gradually tapered off thereafter, becoming insignificant by the year 2000. Also see Head, Mayer and Ries (2008).

[29] Whalley and Xin (2009) calibrate a general equilibrium model of world trade. For a sample of both OECD and non-OECD countries they find that income growth explains 76 percent of the growth of international trade between 1975 and 2004. This finding suggests that trade barrier reductions might have been less important for explaining the trade growth of non-OECD countries.

[30] Another difference is that Baier and Bergstrand (2001) only consider tariffs and transportation costs, whereas trade costs here are more broadly defined to include informational, institutional and nontariff barriers to trade.

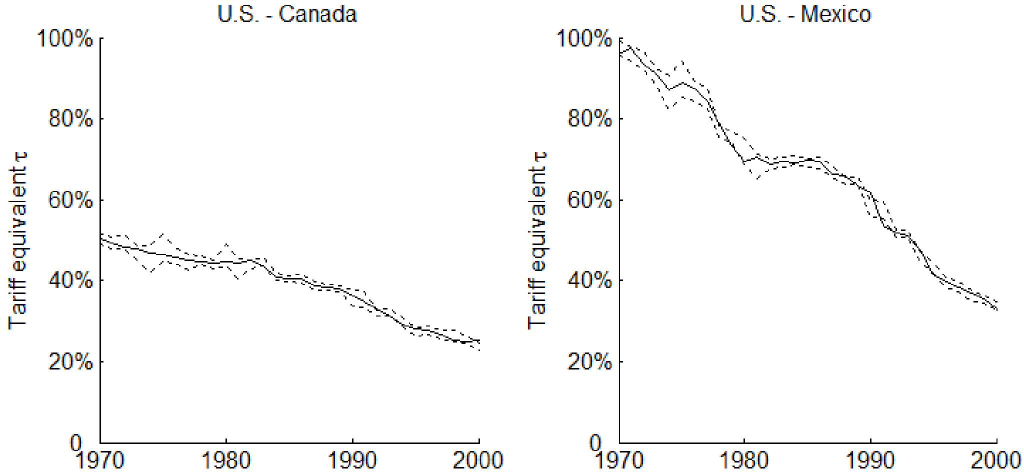

Figure 2: U.S. bilateral trade costs with 99 percent confidence intervals.

stand at 29 percent for the U.S., slightly lower than the average in Table 1.

Measurement error The trade cost measure τ_{ij} is computed based on equation (8) by plugging in the trade data for $x_{ij}x_{ji}$ and $x_{ii}x_{jj}$. Thus, trade costs are inferred without allowing for any stochastic elements. One potential concern with this approach is that the trade data might be subject to measurement error. In particular, suppose that the observed trade flow x_{ij} is a function of the true trade flow x_{ij}^* and an additive measurement error u_{ij} such that $\ln(x_{ij}) = \ln(x_{ij}^*) + u_{ij}$. This measurement error might contaminate the trade cost measure.

To address this concern I divide gravity equation (7) by domestic trade flows $x_{ii}x_{jj}$ and estimate the following loglinear regression:

$$\ln\left(\frac{x_{ij}x_{ji}}{x_{ii}x_{jj}}\right) = \beta \ln\left(\frac{t_{ii}t_{jj}}{t_{ij}t_{ji}}\right) + \alpha_t + \varepsilon_{ij} \tag{17}$$

where α_t are annual time dummies and ε_{ij} is a composite error term given by $\varepsilon_{ij} = u_{ij} + u_{ji} - u_{ii} - u_{jj}$. Since the trade cost parameters are unobservable, I instead substitute country pair fixed effects α_{ij}. The country pair fixed effects are allowed to vary over time to reflect changes in trade costs. As annual fixed effects would leave no degrees of freedom, I choose biennial country pair fixed effects instead. The sample includes the U.S. as well as the countries listed in Table 1 from 1970-2000.[31] The regression yields a very high R^2 (=0.996) with the large majority of fixed effects tightly estimated (p-value < 0.01).

As the final step, I generate predicted values of the dependent variable from the estimated coefficients, and I use the predicted values to construct a predicted trade cost measure $\widehat{\tau}_{ij}$ based on equation (8). $\widehat{\tau}_{ij}$ is supposed to strip out measurement error by construction since it does not include the regression residual that corresponds to ε_{ij}. Figure 2 plots the 'raw' trade cost

[31] There are 651 observations (21 country pairs times 31 years). Standard errors are robust and clustered around country pairs. The last subperiod comprises three instead of two years (1998-2000). Other subperiod lengths, say, quinquennial or decadal, would be possible but would not affect the results qualitatively.

measure τ_{ij} as in Figure 1 (solid lines) as well as the 99 percent confidence intervals (dotted lines) that correspond to the predicted measure $\widehat{\tau}_{ij}$.[32] The intervals are somewhat wider for the 1970s and early 1980s, which suggests lower data quality in that period. Overall, the raw trade cost measure tends to fall within the confidence intervals and it therefore seems unlikely that τ_{ij} is significantly distorted by measurement error.

As an additional check, I rerun regression (17), replacing the country pair fixed effects by standard trade cost proxies. I use the log of bilateral distance, an adjacency dummy, a common language dummy, a joint colonial history dummy as well as country fixed effects to capture the domestic trade cost parameters t_{ii} and t_{jj}.[33] A major problem with this specification is that the explanatory variables are time-invariant and thus not able to capture trade cost changes over time.[34] Instead, the setup imposes a common time trend governed by the annual time dummies α_t. As a result, the predicted trade cost measure fails to pick up pair-specific time trends. For example, it fails to match the relatively strong decline in U.S.-Mexican trade costs during the 1990s that coincides with the establishment of NAFTA. This mismatch could potentially be remedied by time-varying and country-pair specific explanatory variables such as bilateral freight rates but unfortunately, such data are difficult to obtain for a panel.

Income elasticities The trade cost measure is derived from gravity equations that have a unit income elasticity.[35] Although this is a standard feature of gravity models, empirical researchers sometimes estimate income elasticities that deviate from unity, for example Santos Silva and Tenreyro (2006).

Despite the lack of a clear theoretical foundation, suppose the income elasticity in gravity equations (1), (9) and (11) is $\nu \neq 1$ with $\nu > 0$. It is easy to show that the trade cost measure τ_{ij} is unaffected. The contribution of declining bilateral trade costs in decomposition equation (16) therefore also remains the same. But the contribution of income growth would increase if $\nu > 1$ and decrease if $\nu < 1$, and the contribution of declining multilateral resistance would change in the opposite direction by exactly the same extent.

Sensitivity to parameter values The trade cost measure can be derived from different underlying models and therefore potentially depends on different parameters, namely the elasticity of substitution σ, the Fréchet parameter ϑ and the Pareto parameter γ. Although estimates of these parameters usually fall within certain ranges, there is probably no consensus in the literature as to their precise values (see the discussion in Section 3). It turns out that the *levels* of the trade cost measure τ_{ij} are quite sensitive to the chosen parameter values.[36] The *changes* of the trade cost measure over time, however, are hardly affected. In fact, as pointed out in Section 4 and the Technical Appendix, the decomposition of the growth of trade in Table 2 is not affected by parameter values at all.

[32] The confidence intervals are calculated with the delta method. To keep the graph clear, the predicted measure $\widehat{\tau}_{ij}$ is not plotted. It would be located in the middle of the intervals.

[33] As in standard gravity regressions, these trade cost proxies are highly significant. The R^2 is 0.88.

[34] Another potential problem is specification error. The functional form of the implied trade cost function is arbitrary. For a discussion see Anderson and van Wincoop (2004, Section 3.3).

[35] In the case of gravity equation (13) there is a unit elasticity with respect to the number of entering firms in the origin country and the number of consumers in the destination country.

[36] This is true more generally. For example, Anderson and van Wincoop (2004) show that levels of trade cost estimates are typically sensitive to the value of σ.

As $\sigma - 1$ corresponds to ϑ and γ, I will focus the discussion on one single parameter, σ.[16] The trade cost levels reported in Table 1 and Figure 1 are based on $\sigma = 8$, which is in the middle of the common empirical range of 5 to 10 for the elasticity of substitution, as surveyed by Anderson and van Wincoop (2004). For $\sigma = 8$ the trade-weighted average of U.S. bilateral trade costs in Table 1 falls from 74 to 42 percent, a decline of 44 percent. In the case of $\sigma = 10$ the trade-weighted average would fall from 54 to 31 percent, a similar decline of 42 percent. In the case of $\sigma = 5$ the trade-weighted average would fall from 167 to 87 percent, a decline of 48 percent. Thus, although the levels are sensitive to the parameter value, the change of the trade cost measure over time is robust.

A higher value of σ in equation (8) implies lower inferred trade costs levels. Intuitively, a higher elasticity of substitution means that goods are less differentiated and consumers are more price-sensitive. The more price-sensitive consumers are, the fewer foreign goods they would buy for a given difference between bilateral and domestic trade costs. In order to match the given empirical trade flows, a higher elasticity of substitution implies that the difference between bilateral and domestic trade costs must be relatively small, that is, τ_{ij} must be relatively low.

Likewise, a higher value of ϑ in equation (10) implies lower inferred trade costs. Here the intuition is that a higher ϑ corresponds to less scope for Ricardian comparative advantage. Thus, consumers have a smaller incentive to trade and implied trade costs must be lower to match the empirical trade flows. A higher value of γ in equations (12) and (14) also implies lower trade costs. A higher γ corresponds to less heterogeneity across firms, which all else being equal would translate into fewer trade flows unless trade costs were lower.

Finally, it might be the case that the elasticity of substitution has changed over time. Following the approach of Feenstra (1994), Broda and Weinstein (2006) estimate elasticities of substitution based on demand and supply relationships for disaggregated U.S. imports. When comparing the period 1972-1988 with 1990-2001, they find that the median elasticity fell marginally. But the difference is not significant for all levels of disaggregation and it is unclear whether there has been a significant change in the elasticity at the aggregate level. If it were the case that the aggregate elasticity fell over time, this would suggest that trade costs have declined less quickly than indicated in Table 1. But quantitatively, this effect would probably not be large.[37]

Home bias in preferences It is conceivable that consumers predominantly consume domestic goods not because of trade barriers that impede the import of foreign goods but simply because of an inherent home bias in preferences. It is straightforward to incorporate a home bias in preferences into the models outlined in Section 3. Their effect would be observationally equivalent to lower domestic trade barriers.[38] Since the trade cost measure τ_{ij} captures bilateral relative to domestic trade barriers, a home bias in preferences would correspond to inferred trade cost levels that are higher than the 'true' underlying levels. Home bias would thus lead to an overestimate of levels. The change of inferred trade costs over time, however, does not depend on home bias. This reinforces the view that changes in the trade cost measure tend to be more instructive than its levels.

[37] According to Broda and Weinstein (2006, Table IV) the median elasticity fell from 3.7 to 3.1 at the 7-digit level, from 2.8 to 2.7 at the 5-digit level and from 2.5 to 2.2 at the 3-digit level.

[38] That is, lower t_{ii} or f_{ii}.

6 Conclusion

This paper develops a measure of international trade costs that varies across country pairs and over time. The measure is micro-founded and infers bilateral trade costs indirectly from trade data based on a workhorse model of international trade – the gravity equation. I show that the measure can be derived from a range of leading trade theories, including the Ricardian model by Eaton and Kortum (2002), the gravity framework by Anderson and van Wincoop (2003) as well as the heterogeneous firms models by Chaney (2008) and Melitz and Ottaviano (2008). The trade cost measure is a function of observable trade data and can therefore be implemented easily with time series and panel data to track the changes of trade costs over time. This approach obviates the need to impose specific trade cost functions that rely on trade cost proxies such as distance.

In an empirical application I compute U.S. bilateral trade costs for a number of major trading partners. I find that trade costs on average declined by about 40 percent between 1970 and 2000. The decline of U.S. trade costs has been particularly strong with its neighbors Mexico and Canada. I also examine the reasons behind the strong growth of U.S. bilateral trade over that period. I find that income growth is the single most important driving factor. Declines in bilateral trade costs are in second place but quantitatively also play a substantial role.

References

Anderson, J., 1979. A Theoretical Foundation for the Gravity Equation. *American Economic Review* 69, pp. 106-116.

Anderson, J., Marcouiller, D., 2002. Insecurity and the Pattern of Trade: An Empirical Investigation. *Review of Economics and Statistics* 84, pp. 342-352.

Anderson, J., van Wincoop, E., 2003. Gravity with Gravitas: A Solution to the Border Puzzle. *American Economic Review* 93, pp. 170-192.

Anderson, J., van Wincoop, E., 2004. Trade Costs. *Journal of Economic Literature* 42, pp. 691-751.

Baier, S., Bergstrand, J., 2001. The Growth of World Trade: Tariffs, Transport Costs, and Income Similarity. *Journal of International Economics* 53, pp. 1-27.

Baier, S., Bergstrand, J., 2009. Bonus Vetus OLS: A Simple Method for Approximating International Trade-Cost Effects Using the Gravity Equation. *Journal of International Economics* 77, pp. 77-85.

Baldwin, R., Forslid, R., Martin, P., Ottaviano, G., Robert-Nicoud, F., 2003. The Core-Periphery Model: Key Features and Effects. In: The Monopolistic Competition Revolution in Retrospect, edited by S. Brakman and B. Heijdra, Cambridge University Press.

Balistreri, E., Hillberry, R., 2007. Structural Estimation and the Border Puzzle. *Journal of International Economics* 72, pp. 451-463.

Broda, C., Weinstein, D., 2006. Globalization and the Gains from Variety. *Quarterly Journal of Economics* 121, pp. 541-585.

Chaney, T., 2008. Distorted Gravity: The Intensive and Extensive Margins of International Trade. *American Economic Review* 98, pp. 1707-1721.

Chen, N., Novy, D., 2008. International Trade Integration: A Disaggregated Approach. CEPR Discussion Paper #7103.

Combes, P., Lafourcade, M., 2005. Transport Costs: Measures, Determinants, and Regional Policy Implications for France. *Journal of Economic Geography* 5, pp. 319-349.

Deardorff, A., 1998. Determinants of Bilateral Trade: Does Gravity Work in a Neoclassical World? In: The Regionalization of the World Economy, edited by J. Frankel, University of Chicago Press, pp. 7-22.

Debaere, P., 2005. Monopolistic Competition and Trade, Revisited: Testing the Model without Testing for Gravity. *Journal of International Economics* 66, pp. 249-266.

Del Gatto, M., Mion, G., Ottaviano, G., 2007. Trade Integration, Firm Selection and the Costs of Non-Europe. Mimeo.

Eaton, J., Kortum, S., 2002. Technology, Geography and Trade. *Econometrica* 70, pp. 1741-1779.

Evenett, S., Keller, W., 2002. On Theories Explaining the Success of the Gravity Equation. *Journal of Political Economy* 110, pp. 281-316.

Feenstra, R., 1994. New Product Varieties and the Measurement of International Prices. *American Economic Review* 84, pp. 157-177.

Feenstra, R., 2004. Advanced International Trade – Theory and Evidence. Princeton University Press.

Feenstra, R., Markusen, J., Rose, A., 2001. Using the Gravity Equation to Differentiate among Alternative Theories of Trade. *Canadian Journal of Economics* 34, pp. 430-447.

Fujita, M., Krugman, P., Venables, A., 1999. The Spatial Economy – Cities, Regions, and International Trade. MIT Press.

Grossman, G., 1998. Comment. In: The Regionalization of the World Economy, edited by J. Frankel, University of Chicago Press, pp. 29-31.

Head, K., Mayer, T., 2004. The Empirics of Agglomeration and Trade. In: Handbook of Regional and Urban Economics: Cities and Geography, Volume 4, North-Holland.

Head, K., Mayer, T., Ries, J., 2008. The Erosion of Colonial Trade Linkages after Independence. CEPR Discussion Paper #6951.

Head, K., Ries, J., 2001. Increasing Returns versus National Product Differentiation as an Explanation for the Pattern of U.S.-Canada Trade. *American Economic Review* 91, pp. 858-876.

Helliwell, J., 2005. Borders, Common Currencies, Trade, and Welfare: What Can We Learn from the Evidence? *Bank of Canada Review*, Spring, pp. 19-33.

Helpman, E., 1987. Imperfect Competition and International Trade: Evidence from Fourteen Industrial Countries. *Journal of the Japanese and International Economies* 1, pp. 62-81.

Helpman, E., Melitz, M., Rubinstein, Y., 2008. Estimating Trade Flows: Trading Partners and Trading Volumes. *Quarterly Journal of Economics* 123, pp. 441-487.

Helpman, E., Melitz, M., Yeaple, S., 2004. Export versus FDI with Heterogeneous Firms. *American Economic Review* 94, pp. 300-316.

Hummels, D., 2007. Transportation Costs and International Trade in the Second Era of Globalization. *Journal of Economic Perspectives* 21, pp. 131-154.

Hummels, D., Levinsohn, J., 1995. Monopolistic Competition and International Trade: Reconsidering the Evidence. *Quarterly Journal of Economics* 110, pp. 799-836.

Jacks, D., Meissner, C., Novy, D., 2008. Trade Costs, 1870-2000. *American Economic Review* 98, Papers & Proceedings, pp. 529-534.

Kee, H., Nicita, A., Olarreaga, M., 2009. Estimating Trade Restrictiveness Indices. *Economic Journal* 119, pp. 172-199.

Krugman, P., 1980. Scale Economies, Product Differentiation and the Pattern of Trade. *American Economic Review* 70, pp. 950-959.

Limão, N., Venables, A., 2001. Infrastructure, Geographical Disadvantage, Transport Costs and Trade. *World Bank Economic Review* 15, pp. 451-479.

Melitz, M., 2003. The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity. *Econometrica* 71, pp. 1695-1725.

Melitz, M., Ottaviano, G., 2008. Market Size, Trade, and Productivity. *Review of Economic Studies* 75, pp. 295-316.

Nitsch, V., 2000. National Borders and International Trade: Evidence from the European Union. *Canadian Journal of Economics* 33, pp. 1091-1105.

Novy, D., 2007. Is the Iceberg Melting Less Quickly? International Trade Costs after World War II. Mimeo, University of Warwick.

Portes, R., Rey, H., 2005. The Determinants of Cross-Border Equity Flows. *Journal of International Economics* 65, pp. 269-296.

Samuelson, P., 1954. The Transfer Problem and Transport Costs: Analysis of Effects of Trade Impediments. *Economic Journal* 64, pp. 264-289.

Santos Silva, J., Tenreyro, S., 2006. The Log of Gravity. *Review of Economics and Statistics* 88, pp. 641-658.

Trefler, D., 1995. The Case of the Missing Trade and Other Mysteries. *American Economic Review* 85, pp. 1029-1046.

Wei, S., 1996. Intra-National versus International Trade: How Stubborn are Nations in Global Integration? NBER Working Paper #5531.

Whalley, J., Xin, X., 2009. Regionalization, Changes in Home Bias, and the Growth of World Trade. *Journal of Policy Modeling*, forthcoming.

Technical Appendix: Decomposing the Growth of Trade

This appendix derives decomposition equations based on the models by Eaton and Kortum (2002), Chaney (2008) and Melitz and Ottaviano (2008). These decomposition equations correspond to equation (16), which is based on the model by Anderson and van Wincoop (2003). The main result is that the decomposition results in Table 2 are consistent with all these models.

Decomposition Based on Eaton and Kortum (2002)

Eaton and Kortum (2002) rewrite gravity equation (9) as

$$x_{ij} = y_i y_j \frac{\left(\frac{t_{ij}}{P_j}\right)^{-\vartheta}}{\sum_{j=1}^{J} \left(\frac{t_{ij}}{P_j}\right)^{-\vartheta} y_j}$$

where P_j is the CES price index in country j and y_i are total sales of exporter i defined as $y_i \equiv \sum_{j=1}^{J} x_{ij}$. Multiplying and dividing the right-hand side by world income y^W yields

$$x_{ij} = \frac{y_i y_j}{y^W} \left(\frac{t_{ij}}{\Pi_i^{EK} P_j}\right)^{-\vartheta} \tag{18}$$

where $\left(\Pi_i^{EK}\right)^{-\vartheta} \equiv \sum_{j=1}^{J} P_j^{\vartheta} \theta_j t_{ij}^{-\vartheta}$ has the same structure as the outward multilateral resistance variable Π_i in equation (2), with $\sigma - 1$ replaced by ϑ. Gravity equation (18) and gravity equation (1) are thus isomorphic and the decomposition equation can be derived as outlined in Section 4. It follows as

$$100\% = \underbrace{\frac{2\Delta \ln \left(\frac{y_i y_j}{y^W}\right)}{\Delta \ln \left(x_{ij} x_{ji}\right)}}_{\text{(a)}} + \underbrace{\frac{-2\vartheta \Delta \ln \left(1 + \tau_{ij}^{EK}\right)}{\Delta \ln \left(x_{ij} x_{ji}\right)}}_{\text{(b)}} - \underbrace{\frac{-2\vartheta \Delta \ln \left(\Phi_i^{EK} \Phi_j^{EK}\right)}{\Delta \ln \left(x_{ij} x_{ji}\right)}}_{\text{(c)}} \tag{19}$$

where

$$\Phi_i^{EK} = \left(\frac{\Pi_i^{EK} P_i}{t_{ii}}\right)^{\frac{1}{2}}$$

Note that the decomposition in equation (19) does not depend on the value of ϑ even if ϑ changes over time. Contribution (a) is given by the data. Contribution (b) is also given by the data through equation (10), i.e., $-2\vartheta \Delta \ln \left(1 + \tau_{ij}^{EK}\right) = \Delta \ln \left(x_{ij} x_{ji}\right) - \Delta \ln \left(x_{ii} x_{jj}\right)$. Contribution (c) is the multilateral residual. The quantitative results are therefore the same as in Table 2.

Decomposition Based on Chaney (2008)

Gravity equation (11) implies that the product of bilateral trade flows is given by

$$x_{ij} x_{ji} = \left(\mu \frac{y_i y_j}{y^W}\right)^2 \left(\frac{w_i w_j t_{ij} t_{ji}}{\lambda_i \lambda_j}\right)^{-\gamma} \left(f_{ij} f_{ji}\right)^{-\left(\frac{\gamma}{\sigma - 1} - 1\right)}$$

Taking the natural logarithm and the first difference leads to

$$\Delta \ln \left(x_{ij}x_{ji}\right) = 2\Delta \ln \left(\frac{y_i y_j}{y^W}\right) - 2\gamma \Delta \ln \left(1+\tau_{ij}^{Ch}\right) + 2\gamma \Delta \ln \left(\Phi_i^{Ch}\Phi_j^{Ch}\right)$$

where τ_{ij}^{Ch} is substituted from equation (12) and where

$$\Phi_i^{Ch} = \left(\frac{\mu^{\frac{1}{\gamma}}\lambda_i}{w_i t_{ii}\left(f_{ii}\right)^{\frac{1}{\sigma-1}-\frac{1}{\gamma}}}\right)^{\frac{1}{2}}$$

Φ_i^{Ch} captures multilateral resistance λ_i relative to variable and fixed domestic trade costs, as well as domestic productivity w_i and the preference weight μ consumers put on the differentiated goods sector. The decomposition equation follows as

$$100\% = \underbrace{\frac{2\Delta \ln \left(\frac{y_i y_j}{y^W}\right)}{\Delta \ln \left(x_{ij}x_{ji}\right)}}_{\text{(a)}} + \underbrace{\frac{-2\gamma \Delta \ln \left(1+\tau_{ij}^{Ch}\right)}{\Delta \ln \left(x_{ij}x_{ji}\right)}}_{\text{(b)}} - \underbrace{\frac{-2\gamma \Delta \ln \left(\Phi_i^{Ch}\Phi_j^{Ch}\right)}{\Delta \ln \left(x_{ij}x_{ji}\right)}}_{\text{(c)}} \tag{20}$$

Note that the decomposition in equation (20) does not depend on the value of γ even if γ changes over time. Contribution (a) is given by the data. Contribution (b) is also given by the data through equation (12), i.e., $-2\gamma\Delta \ln \left(1+\tau_{ij}^{Ch}\right) = \Delta \ln \left(x_{ij}x_{ji}\right) - \Delta \ln \left(x_{ii}x_{jj}\right)$. Contribution (c) is the multilateral residual whose precise interpretation rests on the elements captured by Φ_i^{Ch}. The quantitative results are therefore the same as in Table 2.

Decomposition Based on Melitz and Ottaviano (2008)

Gravity equation (13) can be rewritten as

$$x_{ij} = \frac{y_i y_j}{y^W}\left(t_{ij}\right)^{-\gamma}\frac{1}{2\delta(\gamma+2)}\frac{N_i^E}{y_i/y^W}\psi_i\frac{L_j}{y_j}\left(c_j^d\right)^{\gamma+2}$$

so that the product of bilateral trade flows can be expressed as

$$x_{ij}x_{ji} = \left(\frac{y_i y_j}{y^W}\right)^2\left(t_{ij}t_{ji}\right)^{-\gamma}\left(\frac{1}{2\delta(\gamma+2)}\right)^2\frac{N_i^E}{y_i/y^W}\frac{N_j^E}{y_j/y^W}\psi_i\psi_j\frac{L_i}{y_i}\frac{L_j}{y_j}\left(c_i^d c_j^d\right)^{\gamma+2}$$

Taking the natural logarithm and the first difference leads to

$$\Delta \ln \left(x_{ij}x_{ji}\right) = 2\Delta \ln \left(\frac{y_i y_j}{y^W}\right) - 2\gamma \Delta \ln \left(1+\tau_{ij}^{MO}\right) + 2\gamma \Delta \ln \left(\Phi_i^{MO}\Phi_j^{MO}\right)$$

where τ_{ij}^{MO} is substituted from equation (14) and where

$$\Phi_i^{MO} = \left(\frac{\left(\frac{N_i^E}{y_i/y^W} \psi_i \frac{L_i}{y_i} \left(c_i^d \right)^{\gamma+2} \right)^{\frac{1}{\gamma}}}{\left(2\delta(\gamma+2) \right)^{\frac{1}{\gamma}} t_{ii}} \right)^{\frac{1}{2}}$$

Φ_i^{MO} reflects domestic trade costs t_{ii}, the number of entrants N_i^E in country i relative to its size in the global economy (y_i/y^W), the extent of comparative advantage ψ_i, per-capita income L_i/y_i and the marginal cost cut-off c_i^d above which domestic firms do not produce. Note that both N_i^E and c_i^d depend on the bilateral trade costs between all other countries in the world (see equations A.1 and A.2 in Melitz and Ottaviano, 2008) so that they have a multilateral interpretation.

The decomposition equation follows as

$$100\% = \underbrace{\frac{2\Delta \ln \left(\frac{y_i y_j}{y^W} \right)}{\Delta \ln \left(x_{ij} x_{ji} \right)}}_{\text{(a)}} + \underbrace{\frac{-2\gamma \Delta \ln \left(1 + \tau_{ij}^{MO} \right)}{\Delta \ln \left(x_{ij} x_{ji} \right)}}_{\text{(b)}} - \underbrace{\frac{-2\gamma \Delta \ln \left(\Phi_i^{MO} \Phi_j^{MO} \right)}{\Delta \ln \left(x_{ij} x_{ji} \right)}}_{\text{(c)}} \tag{21}$$

Note that the decomposition in equation (21) does not depend on the value of γ even if γ changes over time. Contribution (a) is given by the data. Contribution (b) is also given by the data through equation (14), i.e., $-2\gamma \Delta \ln \left(1 + \tau_{ij}^{MO} \right) = \Delta \ln \left(x_{ij} x_{ji} \right) - \Delta \ln \left(x_{ii} x_{jj} \right)$. Contribution (c) is the multilateral residual whose precise interpretation rests on the elements captured by Φ_i^{MO}. The quantitative results are therefore the same as in Table 2.