TI 2012-065/3
Tinbergen Institute Discussion Paper

Product Heterogeneity, Intangible Barriers and Distance Decay: The Effect of Multiple Dimensions of Distance on Trade across Different Product Categories

Maureen B.M. Lankhuizen
Thomas de Graaff
*Henri L.F. de Groot**

Faculty of Economics and Business Administration, VU University Amsterdam.

** Tinbergen Institute.*

Product Heterogeneity, Intangible Barriers and Distance Decay: The effect of multiple dimensions of distance on trade across different product categories

Maureen B.M. Lankhuizen[a*] Thomas de Graaff[a]
Henri L.F. de Groot[a,b]

[a] *Department of Spatial Economics, Faculty of Economics and Business Administration VU University, Amsterdam, The Netherlands*

[b] *Tinbergen Institute, Gustav Mahlerplein 117, 1082 MS Amsterdam*

July 9, 2012

Abstract

We empirically examine the heterogeneity in the effects of multiple dimensions of distance on trade across detailed product groups. Using finite mixture modeling on bilateral trade data at the 3-digit SITC level, we endogenously group product categories into an, a priori unknown, number of segments based on estimated coefficients of multiple dimensions of distance in the gravity equation. We find that institutional distance, whether countries belong to the same trade block and especially geographical distance are crucial and distinct factors to classify commodities in homogeneous groups.

Keywords: bilateral trade, gravity models, distance, institutions, product heterogeneity, finite mixture modeling

JEL-classification: F14, F21,F23

*Corresponding author. Tel +31 (0)20 5986088. Email address: m.b.m.lankhuizen@vu.nl.

1 Introduction

This paper examines the heterogeneity in the effects of multiple dimensions of distance on trade across detailed product groups. The role of distance in international trade is widely researched. Despite—or perhaps, because of—various announcements of the "death of distance" (see, e.g., Cairncross 1997, and Friedman 2005), there have been several recent contributions to the literature that illustrate the continued importance of distance (e.g., Disdier and Head, 2008). A common element in these studies is that they extend the dimension of distance beyond mere geographic distance. As such they elaborate on Deardorff's suggestion that actual trade is below the level one would expect on the basis of transport costs alone (Deardorff, 2004). These findings suggest that there are additional costs involved in trade besides transport costs. To illustrate this empirically, Anderson and Marcouiller (2002), Loungani et al. (2002), de Groot et al. (2004), and Guiso et al. (2009) explicitly consider the role of intangible trade barriers such as the quality of governance or institutions in explaining patterns of bilateral trade. In addition, Linders et al. (2005) and Lankhuizen et al. (2011) also consider the role of cultural differences. Both strands of literature find a significant negative effect of intangible trade barriers on bilateral trade flows. On a related but different note, Hummels (2000) demonstrates that the transportation cost component of manufactured goods has actually gone up, thus challenging the common wisdom of falling transportation costs. This explanation builds on the growing demand for timely delivery.[1] The quality of transport (in terms of speed and reliability) rather than cheaper transport is becoming more important. This is reflected, for instance, in the rising share of air cargo in world trade (Hummels et al., 2007) and the spread of advanced just-in-time (JIT) logistics and distribution systems into almost all modes of global manufacturing, retailing and distribution (McCann, 2008).

Most of the empirical contributions so far have focused on total trade flows. Nevertheless, it is quite conceivable that the effects from distance on trade differ between different types of commodities. In this context, Rauch (1999) and Möhlmann et al. (2010) identify—a priori and based on theoretical considerations—different product groups. Rauch (1999) develops a model of international trade within an economic search framework and distinguishes between homogeneous and differentiated products. He argues that differentiated products are traded through networks of traders, customers and suppliers and that search costs present a major

[1] "The growing demand for timely delivery is rooted in (1) the shift in world trade from bulk commodities to inherently time-sensitive complex manufactures, (2) wealthy consumers' preference for precise product characteristics and ability and willingness to pay for fast delivery, and (3) increasingly segmented production chains." (Hummels et al., 2007, page 2).

barrier to trade in differentiated products, whereas distance only increases transport costs for trade in homogeneous products. Multiple dimensions of distance, such as geographical distance, a common language or a shared colonial history are all factors that affect search costs. Estimations for both homogeneous and differentiated products provide empirical support for the hypotheses above. Möhlmann et al. (2010) extend the work by Rauch. They consider more refined product categories, i.e., they estimate gravity equations for the ten 1-digit product groups in the SITC classification. Furthermore, they also include additional cultural and institutional indicators, thus considering intangible barriers to trade in more detail. Their results indicate that there is substantial heterogeneity in the impact of trade barriers for different product groups. Finally, Hinloopen and van Marrewijk (2012) use the factor intensity classification of the International Trade Center. They find that especially technology intensive products display comparative advantages and relatively high trade values.

This paper builds on these studies by considering the effect of multiple dimensions of distance at a disaggregated level of trade. The contribution of the paper is the application of finite mixture modeling in order to endogenously group international trade at the SITC 3-digit level into an, a priori unknown, number of homogeneous segments (see, e.g., Wedel and Kamakura, 2000, and De Graaff et al., 2009). The main advantage of finite mixture modeling over other clustering techniques is that both the number and the constituent elements of the segments are endogenously generated from the data based on the estimated coefficients of multiple dimensions of distance in the gravity equation.

We find that grouping the data into eight segments yields the best result (i.e., in terms of explanatory power). To summarize, we are able to distinguish the following two general groups: (*i*) high geographic distance segments and (*ii*) low geographic distance segments. To illustrate, product groups belonging to machinery and transport equipment (1-digit SITC 7) are included primarily in the former group, whilst the latter comprises mainly bulk goods and crude materials. Still, there is additional heterogeneity within these two broad groups: other dimensions of distance are important as well in order to further distinguish the individual segments.

The finite mixture modeling applied in this paper captures the heterogeneity in distance decay in international trade in a comprehensive way. As a result, our methodological framework encompasses all three previous classifications in a single framework while offering additional insight in the various effects of multiple dimensions of distance.

The remainder of the paper is organized as follows. The next section describes the estimation method in which we employ a gravity model with fixed effects for each country-product group combination (for both importing and exporting countries) to remove all between country and between product group specific variation. In addition, this section treats briefly the finite mixture modeling procedure. Section 3 presents the data. In section 4, we present our baseline results. The subsequent section interprets the segments as found in section 4 based on the parameter estimates. Section 6 provides an additional robustness check where we investigate whether sample selection biases the product group classification. The last section concludes.

2 Estimation strategy

The gravity model is the most widely used spatial interaction model to study a variety of origin-destination flow phenomena, varying from commuting, telecommunication and asset flows, to migration and trade (see, e.g., Fotheringham and O'Kelly, 1989). It is the workhorse model to study patterns of international trade (see, e.g., Deardorff, 1998, and Anderson and van Wincoop, 2003). The basic gravity model postulates that bilateral trade depends on the economic size of the trade partners, which reflects market size and purchasing power, and a variety of measures of economic distance (or proximity) between the countries to reflect trade costs.

In this paper we use the following extended gravity equation to study patterns of bilateral trade for products from product group p:

$$\ln\left(\mathrm{T}_{eip}\right) = \alpha_p + \ln\left(\mathrm{E}_{ip}\right)\beta_p + \ln\left(\mathrm{I}_{jp}\right)\gamma_p + \ln\left(\mathrm{D}_{eip}\right)\delta_p + \epsilon_{eip}, \tag{1}$$

where T_{eip} denotes bilateral trade between exporter e and importer i in products from product group p. E_{ip} and I_{ip} reflects country specific characteristics (including their economic size) for both exporting and importing countries, respectively. D_{eip} measures various dimensions of distance; α_p, β_p, γ_p and δ_p are (vectors of) product group specific coefficients; and ϵ_{eip} denotes an i.i.d. disturbance term.

In line with theoretical concerns about omitted variables in the gravity equation for both exports and imports (Feenstra, 2004), we estimate a model with country-specific fixed effects for both the country of export and the country of import. Given the structure of the model, in our case this entails estimating a model with country-specific fixed effects *for each product group*. Instead of including dummy variables, we perform the familiar within-transformation (see, e.g., Baltagi, 1995) where we demean the left-hand side and each of the right-hand side variables with the mean of the country of origin by product group and the mean of the country of destination by product group. When applying this, e.g., on our endogenous variable, this transformation yields:

$$\ln\left(\widetilde{\mathrm{T}}_{eip}\right) = \ln\left(\mathrm{T}_{eip}\right) - \ln\left(\mathrm{T}_{e\bullet p}\right) - \ln\left(\mathrm{T}_{\bullet ip}\right) + \ln\left(\mathrm{T}_{\bullet\bullet p}\right), \qquad (2)$$

where the $\bullet$ denotes the mean over that particular variable. OLS on this transformed model gives the familiar two-way fixed effects estimator. In this way, country-specific effects, such as GDP and GDP per capita and institutional quality, for both the importer and the exporter country, and product-specific characteristics that determine the size or value of the trade flow between countries, e.g., high-tech versus low-tech goods, are intrinsically controlled for. Moreover, each country's specific comparative advantage (whether relative or absolute) in exporting products from product group p is controlled for as well.

Building on recent insights in the empirical literature on international trade, this paper focuses on multiple dimensions of distance (D_{eip}). The effect of geographic distance on trade is shown to be persistent over time (see, e.g., Disdier and Head, 2008; Linders et al., 2011). This has presented somewhat of a puzzle as the costs of trade associated with transport and communication have fallen in recent decades. In a thought-provoking contribution, Obstfeld and Rogoff (2000) underline the importance of intangible barriers, such as incomplete information barriers, cultural barriers and institutional barriers, in explaining the persistence of 'transactional distance' between countries. The importance of search costs and networks in trade (see, e.g. Rauch, 1999, 2001) illustrates the importance of information costs for patterns of trade. The effect of cultural barriers consists of two aspects, viz. cultural familiarity and cultural distance. Much like other sources of incomplete information, unfamiliarity with foreign cultures leads to search costs and adjustment costs incurred in international interactions. Familiarity with foreign culture is expected to increase if countries share a common language, and to decrease with geographical distance. Apart from that, distance in terms of cultural values and norms, causes barriers related to trust and understanding (Linders et al., 2005). Institutions influence the uncertainty surrounding transactions. The quality of institutions affects expropriation risks, the degree of corruption, the enforceability of private contracts, and hence the security of trade (Anderson and Marcouiller, 2002). Controlling for the quality of the governance environment in both countries, bilateral trade may be hampered more if the distance between governance systems increases.

So, in addition to geographic distance, we specify distance in terms of cultural and institutional distance. Besides a measure of cultural distance, we control for a shared cultural background by including two dummy variables indicating whether countries share a common language or colonial ties. Thirdly, the set of control variables also includes a dummy variable indicating whether or not countries share borders. Finally, as an indicator of policy-induced barriers to trade, we use a dummy variable indicating whether countries belong to the same trade bloc.

Finally, we pool all product groups p. To still allow for heterogeneity in the distance coefficients, δ_p, without loosing too many degrees of freedom, we adopt a finite mixture modeling approach to group trade in different product categories into segments (see, e.g., De Graaff et al., 2009). The finite mixture approach allows us to explicitly model the variation in the effects of multiple dimensions of distance on trade flows across different (3-digit SITC) product groups. The actual number of segments is not known *a priori*. The approach is to expand the number of segments until the log-likelihood no longer improves (De Graaff et al., 2009). To this end, we use the Flexmix package in the software environment 'R' (Leisch, 2004). The Bayesian information criterion (BIC) is used to determine the optimal number of segments. We describe our finite mixture modeling approach in more detail in Appendix A.

Note that in the present context estimating a model with country-by-product specific fixed effects has an important additional advantage: the segmentation of the product groups is based primarily on the (variation in the) dimensions of distance. To be more specific, the segmentation is only based on the heterogeneity within the transformed bilateral trade flows, which are primarily caused by variation in the various dimensions of distance but may—to a lesser extent—be caused by market and location effects as well. For example, commodities that primarily come from Australia or New Zealand, such as particular agricultural goods, face relatively flat distance-decay functions, because of the countries' absolute geographical location relative to the world market.

3 Data

We use the United Nations trade data compiled by Feenstra et al. (2005). The data report bilateral trade flows according to the Standard International Trade Classification (SITC) revision 2 at the 3-digit level.[2] We use data for the year 2000. We use distance in miles between capital cities for geographical distance between countries. The data for the indicators of cultural

[2]The dataset contains only non-zero trade flows. Estimating trade without first estimating whether trade is observed could lead to a sample selection bias (cf. Heckman, 1979): namely, zero flows of the dependent variable, here bilateral trade, may not be random. Nevertheless, using aggregate trade data, Linders and de Groot (2006) show that OLS does not greatly suffer from selection bias. We therefore use a straightforward regression framework. To investigate to what extent sample selection has an impact on the determinants of trade on a more disaggregate level, we analyze a sample with only large trade flows (larger than 1 million U.S. dollars) in Section 6.

distance are from Hofstede (2001). Hofstede (1980, 2001) has developed a set of variables that reflect national cultures in terms of norms and values. These variables are: masculinity (versus femininity); uncertainty avoidance; individualism (versus collectivism); and power distance. Each is constructed on the basis of principal components analysis, and intends to reflect the stance of a distinct set of work-related norms and values in national cultures. Our measure of institutional distance is based on Kaufmann et al. (2005). They have constructed six indicators of perceived institutional quality on the basis of principal components analysis. These indicators are: voice and accountability; political stability; government effectiveness; regulatory quality; rule of law; control of corruption. To measure cultural and institutional distance, we apply an index of distance that was developed for these purposes and first applied by Kogut and Singh (1988).[3] Data on adjacency, common language and colonial ties are from CEPII. The data on trade blocs are obtained from the OECD. Because of computer memory considerations, our calculations are based on a random sample (35 per cent) of the total observations in the Feenstra data set. In total we have 101,743 observations belonging to 237 3-digit SITC product groups. Note that the latter classification is the basis for our segmentation.

4 Estimation results

Using finite mixture modeling, we find that grouping the product groups into eight segments yields the best result (i.e., in terms of explanatory power). Figure 1 illustrates the quality of the segmentation. The figure gives a rootogram of the posterior probabilities of membership for each segment, i.e. the probability that (trade flows for) a certain product group belongs to a specific segment (horizontal axis).[4] For comparison reasons, the vertical axis corresponds to the square root of the number of product groups in each bar. A peak at a probability of one or zero indicates that a segment is well separated from the other segments; significant mass in the middle of the unit interval indicates overlap with other segments (see, e.g., Leisch, 2004). Hence, product groups in the intermediary range of probabilities indicate that they can be allocated to multiple segments. Figure 1 shows that most segments contain product groups in the middle of

[3] The index is defined as:

$$D_{e,i} = \frac{1}{K}\frac{\sum_{k=1}^{K}(I_{e,k} - I_{i,k})^2}{V_k},$$

where $D_{e,i}$ is the measure of distance between country e and country i, K is the number of indicators of culture/institutional quality distinguished (indexed by k), $I_{e,k}$ is country e's score with respect to indicator k, and V_k the variance of indicator k over all countries in the sample.

[4] Product groups with posterior probabilities smaller than 0.0001 are omitted. Usually many product groups in each segment have posterior probabilities close to zero. To avoid having the high count in the corresponding bar obscure the information in other bars of the rootogram, these product groups are omitted (see, e.g., Leisch, 2004).

the interval. Thus, there is some overlap between segments.

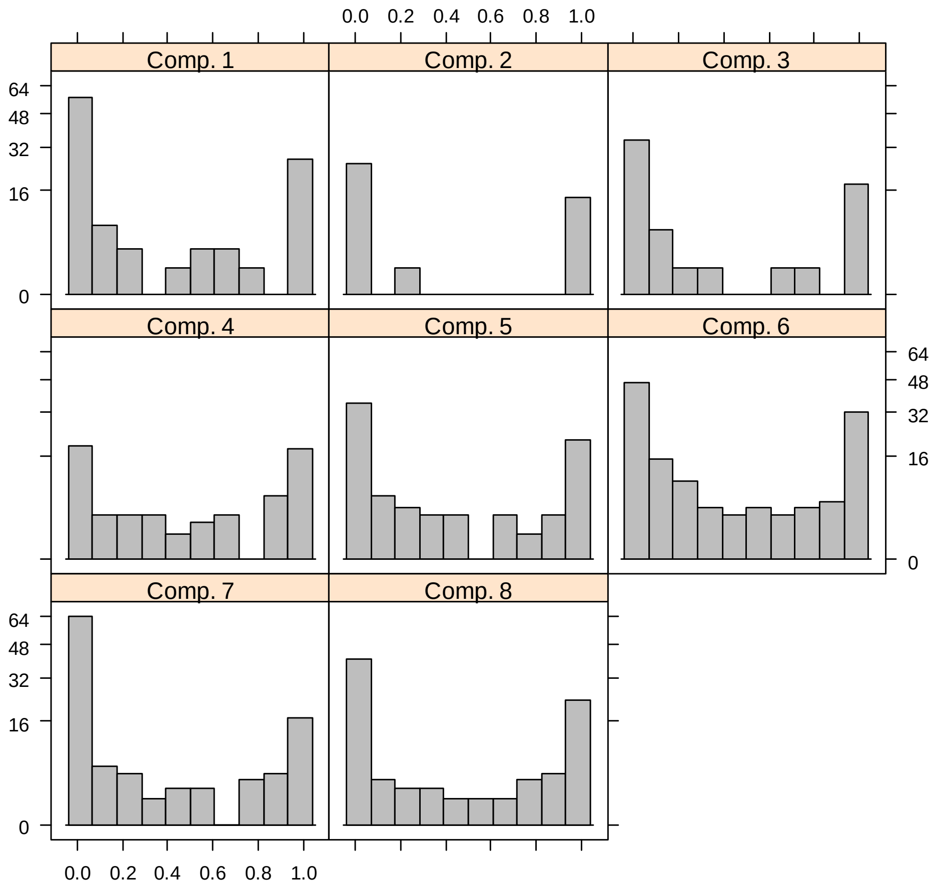

Figure 1 – Rootograms of posterior membership probabilities for each segment

The same information is conveyed by Table 1. The column 'Size' gives the number of trade flow observations actually assigned to each segment. The column 'Posterior' gives the number of trade flows in each segment with a posterior probability exceeding some threshold level (the default is 0.0001). The fact that the number of trade flows in column (2) exceeds the number of trade flows in (1) indicates that some trade flows have membership probabilities exceeding the threshold for more than one segment. So, the difference between the numbers in columns (2) and (3) represents the overlap between segments. The values for the ratio in column (3) indicate that there is some overlap for our data. That is, for some product groups trade may be assigned to one or more segments.[5]

Table 2 gives the regression results for the multiple dimensions of distance in the segments. Overall, coefficients by and large have the a priori theoretically hypothesized sign and are

[5]The quality of the segmentation improves when the sample size increases. Using, e.g., a random sample of 70 per cent of the total observations in the Feenstra data set, the corresponding ratios are approximately 0.6 and higher. Changing the sample size leaves our main conclusions regarding the segments (qualitatively) unaltered.

Table 1 – Summary statistics of the finite mixture modeling results

Segment	Size (1)	Posterior (2)	Ratio (3) = (1)/(2)
1	21,876	47,740	0.46
2	10,319	20,508	0.50
3	10,540	21,025	0.50
4	4,088	9,042	0.45
5	14,355	30,996	0.46
6	17,944	39,99	0.45
7	14,234	36,276	0.39
8	8,387	18,513	0.45

Table 2 – Estimates of multiple dimensions of distance for various segments at the 3-digit SITC level (N = 101,743)

Segment	1	2	3	4	5	6	7	8
ln (distance)	-0.68^{***}	-0.95^{***}	-0.57^{***}	-0.12^{***}	-0.60^{***}	-0.37^{***}	-0.62^{***}	-0.12^{***}
Common language	0.44^{***}	0.28^{***}	0.30^{***}	0.14	0.19^{***}	0.20^{***}	0.38^{***}	0.14
Adjacency	0.54^{***}	0.59^{***}	0.67^{***}	0.52^{***}	0.68^{***}	0.72^{***}	0.86^{***}	0.87^{***}
Colony	0.46^{***}	0.42^{***}	0.37^{***}	0.12	0.33^{***}	0.34^{***}	0.36^{***}	0.30^{***}
Common trade bloc	0.56^{***}	0.59^{***}	0.93^{***}	0.30^{***}	0.44^{***}	0.55^{***}	0.65^{***}	0.66^{***}
Institutional distance	-0.10^{***}	-0.05^{***}	-0.09^{***}	0.06^{***}	-0.05^{***}	-0.04	0.04^{**}	0.02
Cultural distance	-0.10^{***}	-0.08^{***}	0.00	-0.04^{*}	-0.06^{***}	-0.04^{***}	0.00	-0.08^{***}
Intercept	-0.02^{*}	0.00	-0.01	0.00	0.01	0.00	0.00	0.00

Dependent variable: bilateral trade at the 3-digit SITC level. Significance levels : $*$: 10% $**$: 5% $***$: 1%.

statistically significant (apart from the Intercept). Only the finding that institutional distance has a positive and statistically significant coefficient in segments 4 and 7 (and statistically insignificant in segment 8) is at odds with theoretical expectations. At the same time though, segment 4 has relatively low values for all distance parameters. The overall positive coefficient for institutional distance may well be a reflection of the overall low distance decay in this segment. As for segment 7, this segment contains quite some products belonging to 1-digit SITC 6, i.e., manufactured goods classified chiefly by material, and textile wearing apparel (2-digit SITC 84).[6] Institutional proximity, i.e. a similar quality of the governance system in both countries, is less important for this type of products, explaining the positive parameter for institutional distance. The coefficient of cultural distance is statistically insignificant in two segments, including segment 7. We believe that much of the reasoning for the role of institutional distance in segment 7 also applies to cultural distance. The coefficient of cultural distance is negative and statistically significant in the other segments. This is consistent with theoretical expectations.

Ranking the segments in terms of the coefficient of geographic distance, we find that trade in segments 4, 6 and 8 has a relatively low sensitivity to geographic distance, whilst the sensitivity

[6] See the next section for a further interpretation of the segments.

of trade to geographic distance in segments 1, 2, 3, 5 and 7 is relatively high. The segments are further distinguished by the relative importance of other dimensions of distance.

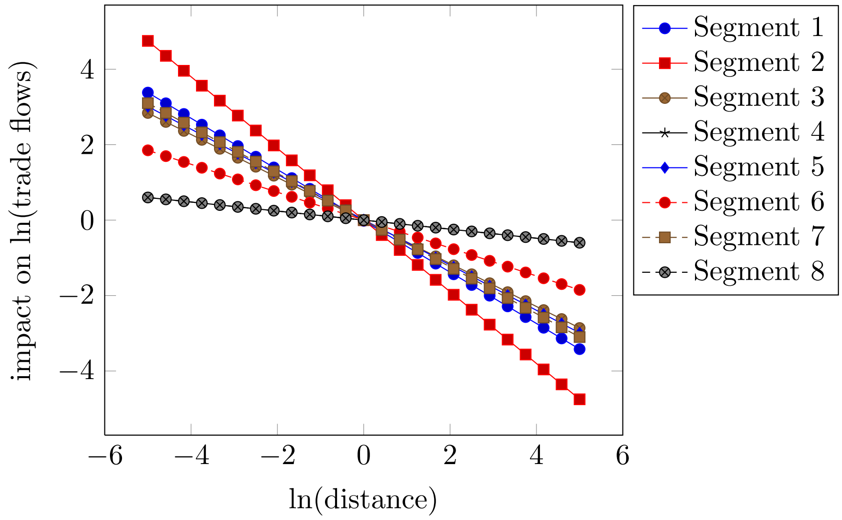

Figure 2 – Fitted distance decay functions for each of the segments

To illustrate the heterogeneity, Figure 2 gives a graphical representation of the segments. The graph presents the impact of the (two-way transformed and logarithmic) geographic distance on the (two-way transformed and logarithmic) trade flows. The graph clearly shows how similar values of geographic distance have a different effect on trade flows in the different segments.

5 Interpretation of the segments

In this section we interpret the different segments based on the parameter estimates. For a complete overview of our classification see Appendix B. Products in *Segment 4* are the least sensitive to distance: the segment has relatively low coefficients for all dimensions of distance, including a significantly positive parameter on institutional distance. The latter indicates that similarity between two countries in terms of the quality of the governance environment is not important for trade. *Segment 8* has similar characteristics, notably with respect to geographic distance. However, adjacency and a common trade bloc are relatively important for trade in segment 8. Thus, trade in this segment is also largely between countries that are part of the same trade bloc and between neighbouring countries despite the fact that the sensitivity to geographic distance of trade is low. We can describe these patterns by looking at the products within the segments. Trade in 'bulk' commodities is concentrated largely in segments 4 and 8. These are essentially all crude materials (1-digit SITC 2), fuels (1-digit SITC 3) and animal and vegetable oils, fats and waxes (1-digit SITC 4). We use the term bulk to refer to the mode in which these products are transported. Transport costs are the predominant trade barrier in these product categories. However, these commodities are easy to transport (in large volumes) and are time-insensitive, explaining the relatively low parameter on geographical distance. Sensitivity to institutional distance is relatively low (see the parameters in Table 2). At the same time, segments 4 and 8 include 'location-dominated' products. These are products that are only produced (*i*) in a limited number of (geographic) areas or (*ii*) by a small number of producers, whereas the goods are consumed worldwide. An example of the first sub-group is rice (042). Rice is grown mainly in Asia but consumed all around the world. Other examples of the first sub-group include coffee and tea (071 and 074), spices (075), tobacco (121) and silk (261). Aircraft and associated equipment and parts (792) and ships, boats and floating structures (793) make up the second sub-group. There are, e.g., only a few manufacturers of commercial aircraft. These companies sell planes to airline companies worldwide. The low sensitivity to geographical distance of these products is an expression of the fact that goods are transported across long distances. This does not exempt strong effects from adjacency or a common trade bloc. For instance, most of Japan's rice produce is grown in Thailand. Similarly, Boeings are likely to be sold foremost in Northern America, whilst the main markets for Airbus are probably in Europe. There is, by and large, no trade in machinery in components 4 and 8. Our analysis suggests that trade in machinery is generally more sensitive to distance.

Trade in segments 1, 2, 3, 5 and 7 is characterized by a relatively high sensitivity to geographic distance. The segments differ in terms of the relative importance of other dimensions of distance. *Segment 2* has the highest overall sensitivity of trade to geographic distance (−0.95). Trade in this segment deceases more rapidly when geographic distance between two countries increases. Hence, relative proximity to buyers, both final customers (in the case of household equipment (775)) and industrial buyers (in the case of, e.g., iron and steel bars (673), copper (682) and aluminium (684)), is important.

Whilst the high sensitivity to geographic distance is the predominant characteristic of trade in segment 2, trade in *segment 3* is also characterized by a particularly high importance of the trade bloc variable. An important group of products in segment 3 are 'high-tech' products.[7] These include computers (752), electrical parts (759, 776), motor vehicles (713, 781, 782), and television sets and telecom equipment (761, 764).[8] Production and consumption of these products

[7] Our use of the term high-tech is broader than the definition of high-tech and high-tech/medium-high-tech (see, e.g., OECD).

[8] Other high-tech product groups are included in segment 1, e.g., pharmaceuticals (541) and electrical machinery (778).

are largely concentrated in highly developed, industrialized countries (North America, Europe, Japan and South Korea), and parts of Asia (Taiwan, Hong Kong, India (Bangalore, Mumbai) and China (Shanghai, Beijing) and Latin America (Brazil). Hence, there is an a priori limit to the number of actual trading distances in these types of products. Besides, these industries are generally characterized by the presence of multinational enterprises (MNEs). These companies actively seek to reduce geographical distance by locating plants close to (main) markets: a number of countries at a relatively short range may then be serviced through export platforms. Even if MNEs offshore parts of the production chain for reasons of efficiency (vertical FDI), it is likely to be in vicinity of the main markets (e.g., assembly of electronics and cars in Eastern Europe and Mexico). FDI may also explain the importance of the trade bloc variable in this segment: MNEs locate production in one location and then serve other countries within the trade bloc so as to jump tariff barriers (see, e.g., Lankhuizen et al., 2011).

Segments 1, 5 and 7 have, by and large, comparable parameters of geographic distance. *Segment 7* differs from the other two segments because of its low sensitivity to institutional distance. The coefficient is positive and statistically significant. The coefficient of cultural distance is statistically insignificant in this segment. In contrast to segments 1 and 5, segment 7 contains relatively little chemical products and machinery, but more textile wearing apparel (2-digit SITC 84).[9] Institutional and cultural proximity is less important for the latter type of products.[10] *Segment 1* also includes some high-tech product groups, such as medicine and pharmaceuticals (541) and electrical machinery and apparatus (778). This explains why trade in segment 1 is more sensitive to most dimensions of distance than *segment 5* (with the exception of adjacency). The parameter of institutional distance is particularly high in segment 1. Coordination and transaction costs are lower in countries that are close in terms of the quality of the governance system. This seems particularly important for more complex products.[11] Furthermore, trade in segment 1 is relatively sensitive to a common language and colonial ties.

Segments 5 and 6 largely represent the 'other' segments in our analysis, with the former belonging to the segments with a relatively high sensitivity to geographic distance and the latter to the segments with a relatively low sensitivity to geographic distance. In particular, trade in *segment 6* is more sensitive to the various dimensions of distance than trade in segment 4. On the one hand, segment 6 contains edibles that are more perishable than edibles in segment

[9] All three segments contain quite some products belonging to 1-digit SITC 6, i.e., manufactured goods classified chiefly by material.

[10] This might reflect the relative long trade distance between typical exporting countries, such as Brasil and China, and typical importing countries—usually located in the United States and Europe.

[11] The same applies to segment 3, the other main segment with high-tech products.

4 (i.e., meat, butter and fish versus cereals, tea and spices).[12] On the other hand, segment 6 contains more machinery and transport equipment. Nevertheless, segment 6 also contains 'bulk' products, i.e., products belonging to 1-digit SITC 2, SITC 3 and SITC 4. This explains the relatively low overall sensitivity to distance of trade in this segment.

5.1 Discussion

The main characteristics of the segments we found are summarized in Table 3.

Table 3 – Characteristics of segments—summary

Segments	Sensitivity to geographic distance	Other dimensions of distance and characteristics
1	High	High sensitivity to institutional distance, language and colony
2	High	Highest overall sensitivity to geographic distance
3	High	High sensitivity to trade bloc and institutional distance
5	High	Lowest sensitivity to trade bloc
7	High	Low sensitivity to institutional distance
4	Low	Lowest overall sensitivity to distance
6	Low	Relatively low sensitivity to most dimensions of distance
8	Low	High sensitivity to adjacency and trade bloc

Our framework illustrates how different dimensions of distance affect trade across different product groups differently. In this manner, it provides a richer picture of the heterogeneity in distance decay in international trade than previously introduced classifications used by Rauch (1999), Möhlmann et al. (2010) and Hinloopen and van Marrewijk (2012).

Rauch (1999) identifies three product groups: homogeneous products comprise products traded on an organized exchange and reference-priced articles; the third group consists of differentiated goods. The network theory of trade hypothesizes that search costs are most important for the pattern of trade in differentiated products and least important for organized-exchange products. Our framework encompasses Rauch's classification and extends this. For instance, differentiated goods in the tradition of Rauch, with a high sensitivity to particularly institutional distance, in our framework are split across four segments: 1, 2, 3 and 5. Homogeneous products are further divided across segments 4, 5 and 7. This indicates that within the groups

[12] Segment 6 contains the preserved versions of fresh meat, fish and fruit that are included in segment 3. It makes sense that the overall sensitivity to distance in the former is lower. At the same time, other edibles that are sensitive to time, e.g., milk and cream (022) and eggs (025), are included in segments with a high relative sensitivity to geographic distance (segments 1, 3, 5 and 7). Hence, our mixture modeling analysis captures these differences well.

defined by Rauch, there is additional heterogeneity that is captured by our finite mixture modeling approach.

The same argument applies to the framework of Hinloopen and van Marrewijk (2012). They classify trade by technology-intensity. Our framework also distinguishes segments based on the type of products traded. Yet, even within product categories, our analysis distinguishes differences with respect to the sensitivity to distance. For instance, high-tech products are divided across two segments (1 and 3) based on differences in the parameters of the multiple dimensions of distance.

Finally, our analysis also differs from Möhlmann et al. (2010) in that we do not specify the number of segments ex ante. Möhlmann et al. estimate gravity equations for product groups according to the SITC (1-digit) classification. Instead, we endogenously group product categories into an, a priori unknown, number of segments based on the parameters of multiple dimensions of distance in the gravity equation. The result is that segments cut across predefined SITC groups. This indicates that there is a large amount of heterogeneity within SITC groups in terms of sensitivity to distance. We conclude that the finite mixture modeling captures the heterogeneity in distance decay in international trade in a comprehensive way.

6 Robustness

Linders and de Groot (2006) show that sample selection (in their case zero flows) of aggregate trade flows might have a limited impact on the coefficients. To test the robustness of our classification on more disaggregated trade flows regarding sample selection bias, we therefore repeated the finite mixture procedure using only 'important trade flows'. All trade flows with a value of bilateral trade below 1 million U.S. dollars were dropped from the sample. The ensuing sample still includes bilateral trade flows for all 237 3-digit SITC codes. So, both our main analysis and the sensitivity analysis are based on (a sample of) the same product categories.

Once again, a specification of the model with eight segments yields the best result in terms of the BIC. The segments are relatively well segmented (see Table 4). Closer inspection of the segments reveals that approximately 60 per cent of the product categories are clustered within the same segment as with a sample including all trade flows. This indicates that the clustering is rather robust.[13]

[13]There is some natural variability in the segmentation anyway. Given that there is some overlap between segments, some product groups may be even assigned to different segments when repeating the analysis with different starting values.

Table 4 – Summary statistics of the finite mixture modeling results used in the sensitivity analysis

Segment	Size (1)	Posterior (2)	Ratio (3)=(1)/(2
1	5,924	7,589	0.78
2	13,244	21,848	0.61
3	14,047	18,208	0.77
4	8,939	10,413	0.86
5	23,153	50,917	0.46
6	39,357	49,870	0.79
7	21,043	43,928	0.48
8	28,515	39,064	0.73

Moreover, the parameter estimates are smaller in the sample with important trade flows. Some coefficients are no longer statistically significant. This shows that the segmentation seems to be robust to the segmentation procedure, but that the estimation itself is less robust. The latter obviously depends on the fact that we have only included the largest trade flows. These results are available upon request.

7 Conclusion

This paper applies finite mixture modeling in order to endogenously group international trade at the SITC 3-digit level into an, *a priori* unknown, number of segments. We find that grouping the data into eight segments yields the best result. We distinguish the following two general groups: trade flows that are sensitive to (*i*) high geographic distance and to (*ii*) low geographic distance. As an example, product groups belonging to machinery and transport equipment (1-digit SITC 7) are included primarily in the former group, whilst the latter comprises mainly bulk goods and crude materials. Still, there is additional heterogeneity within these two broad groups: other, less tangible, dimensions of distance are important as well in order to further distinguish the individual segments. We find that institutional distance and whether both countries belong to the same trade block are particularly important in this respect.

The main contribution of this paper is that our framework provides a richer picture of the heterogeneity in distance decay in international trade than previously introduced classifications used by Rauch (1999), Möhlmann et al. (2010) and Hinloopen and van Marrewijk (2012). Our framework encompasses all these three classifications in a single framework while offering additional insight in the various effects of multiple dimensions of distance. Moreover, we find that that there is a large amount of heterogeneity *within* SITC groups in terms of sensitivity to multiple dimensions of distance, even at the 3-digit level. We thus infer that estimating at a more aggregate level—e.g., at a 1-digit level—might give misleading estimations. This is especially important when imputing missing data with estimated trade flows.

The insights of recent studies including this one that are looking into the heterogenous impact of different dimensions of trade across different product groups is also of clear potential relevance for future policy implications. It underlines that care is needed in developing policies targeted at fostering trade. The success of particular policy measures will depend heavily on the impact on the various dimensions of distance and their importance for the country and sector that is targeted. One size fits all policy measures will not work. Increased research efforts along the lines pursued in this paper can help in enhancing our understanding of trade determinants in all their complexities and heterogeneities and contribute to better evidence based trade policies.

A Segmenting product groups with finite mixture modeling

This paper uses a finite mixture approach to divide the product groups at a disaggregated level into an, *a priori* unknown, number of segments (we follow here the notation of Leisch, 2004).

In total we have 101,743 bilateral trade flows belonging to 237 3-digit SITC groups. Thus, we have 237 product groups that we want to segment, where product group p consists of N_p observations. Assume that observations on $\ln\left(\widetilde{\mathrm{T}}_{eip}\right)$ arise from a population that is a mixture of S segments in proportions $\pi_1, \ldots, \pi_S$, where we do not know in advance from which segment observations on $\ln\left(\widetilde{\mathrm{T}}_{eip}\right)$ arise. Then, the conditional density function of $\ln\left(\widetilde{\mathrm{T}}_{eip}\right)$ (where in our case $\ln\left(\widetilde{\mathrm{T}}_{eip}\right)$ denote the logarithmic bilateral trade flows by 3-digit SITC) can be decomposed into its various segments as follows:

$$f(\ln\left(\widetilde{\mathrm{T}}_{eip}\right) \mid \ln\left(\widetilde{\mathrm{D}}_{eip}\right), \delta) = \sum_{s=1}^{S} \pi_s f_s(\ln\left(\widetilde{\mathrm{T}}_{eip}\right) \mid \ln\left(\widetilde{\mathrm{D}}_{eip}\right), \delta_s), \quad (3)$$

where $\pi_s \geq 0$, $\sum_{s=1}^{S} \pi_s = 1$, $\ln\left(\widetilde{\mathrm{D}}_{eip}\right)$ is the matrix of variables that measures various dimensions of distance and δ_s is the vector of parameters specific for each segment s.

The log-likelihood of (3) is estimated by applying the expectation maximization (EM) algorithm of Dempster et al. (1977). The first step is the expectation (E) step, which computes the expected value of the complete log-likelihood function with respect to the segments s. This is given by:

$$\ln L = \sum_{p=1}^{P} \sum_{n=1}^{N_p} \ln f(\ln\left(\widetilde{\mathrm{T}}_{eip}\right) \mid \ln\left(\widetilde{\mathrm{D}}_{eip}\right), \delta), \quad (4)$$

where $\sum_{p=1}^{P} N_p/N = 1$. The posterior probability that product group p belongs to segment s is given by:

$$\hat{\pi}_{ps} = \frac{\pi_s \prod_{n=1}^{N_p} f(\ln\left(\widetilde{\mathrm{T}}_{eip}\right) \mid \ln\left(\widetilde{\mathrm{D}}_{eip}\right), \delta_s)}{\sum_{s=1}^{S} \pi_s \prod_{n=1}^{N_p} (\ln\left(\widetilde{\mathrm{T}}_{eip}\right) \mid \ln\left(\widetilde{\mathrm{D}}_{eip}\right), \delta_s)}. \tag{5}$$

We can now derive the probability of segment s which can be inserted in (3) as:

$$\hat{\pi}_s = \frac{1}{P} \sum_{p=1}^{P} \hat{\pi}_{ps}. \tag{6}$$

Thus, $\hat{\pi}_s$ are estimated using current values of the model parameters and can be inserted in (3).

In the maximization (M) step, the expected value of the complete log-likelihood function (3) is maximized with respect to the model parameters using the posterior probabilities as weights. This maximization step is performed sequentially (see van Dijk et al., 2007) as follows:

$$\max_{\delta_s} \sum_{p=1}^{P} \hat{\pi}_s \ln f(\ln\left(\widetilde{\mathrm{T}}_{eip}\right) \mid \ln\left(\widetilde{\mathrm{D}}_{eip}\right), \delta_s), \tag{7}$$

Both steps E and M are now iteratively applied until convergence occurs (Leisch, 2004).

B Product group classification

In Table 5 we list the segments that follow from the finite mixture modeling with (a random sample from) the set of all trade flows in terms of their constituent product categories.

Table 5 – Segments and corresponding product categories

Segment 1	Segment 2
Chocolate & other food preptions containing cocoa	Condensation, polycondensation & polyaddition prod.
Edible products and preparations n.e.s.	Polymerization and copolymerization products
Synthetic fibres suitable for spinning	Paper and paperboard
Alcohols, phenols, phenol-alcohols, & their derivat.	Paper and paperboard, cut to size or shape
Carboxylic acids, & their anhydrides, halides, etc.	Textile yarn
Pigments, paints, varnishes & related materials	Iron and steel bars, rods, angles, shapes & sections
Medicinal and pharmaceutical products	Universals, plates and sheets, of iron or steel
Perfumery, cosmetics and toilet preparations	Copper
Soap, cleansing and polishing preparations	Aluminium
Miscellaneous chemical products, n.e.s.	Structures & parts of struc.; iron, steel, aluminium
Rubber tyres, tyre cases, etc.for wheels	Equipment for distributing electricity
Veneers, plywood, improved or reconstituted wood	Household type, elect.& non-electrical equipment

Textil.fabrics, woven, oth.than cotton/man-made fibr	Parts & accessories
Knitted or crocheted fabrics	Articles of materials described in division 58
Glass	
Tubes, pipes and fittings, of iron or steel	
Metal containers for storage and transport	
Tools for use in hand or in machines	
Household equipment of base metal, n.e.s.	
Manufactures of base metal, n.e.s.	
Civil engineering & contractors plant and parts	
Mach.& equipment specialized for particular ind.	
Mach.tools for working metal or met.carb., parts	
Heating & cooling equipment and parts	
Pumps & compressors, fans & blowers, centrifuges	
Mechanical handling equip.and parts	
Non-electric parts and accessories of machines	
Electric power machinery and parts thereof	
Electrical machinery and apparatus, n.e.s.	
Trailers & other vehicles, not motorized	
Sanitary, plumbing, heating, lighting fixtures	
Printed matter	
Musical instruments, parts and accessories	
Other miscellaneous manufactured articles	

Segment 3	Segment 4
Live animals chiefly for food	Barley, unmilled
Meat, edible meat offals, fresh, chilled or frozen	Cereals, unmilled (no wheat, rice, barley or maize)
Crustaceans and molluscs, fresh, chilled, frozen etc.	Other cereal meals and flours
Fruit & nuts (not including oil nuts), fresh or dried	Tea and mate
Tobacco manufactured	Spices
Wood, simply worked, and railway sleepers of wood	Tobacco, unmanufactured; tobacco refuse
Petroleum products, refined	Furskins, raw (including astrakhan, caracul, etc.)
Hydrocarbons nes, & their halogen.& etc.derivatives	Fuel wood (excluding wood waste) and wood charcoal
Ingots and other primary forms, of iron or steel	Silk
Internal combustion piston engines & parts	Jute & other textile bast fibres, nes, raw/processed
Automatic data processing machines & units thereof	Vegetable textile fibres and waste of such fibres
Parts of and accessories suitable	Other man-made fibres suitabl.for spinning & waste
Television receivers	Fertilizers, crude
Telecommunications equipment and parts	Sulphur and unroasted iron pyrites
Thermionic, cold & photo-cathode valves, tubes, parts	Natural abrasives, n.e.s (including industrial diamonds)
Passenger motor cars, for transport of pass.& goods	Briquettes; coke and semi-coke of coal, lignite/peat
Motor vehicles for transport of goods/materials	Animal oils and fats

Outer garments, women's, of textile fabrics	Dyeing & tanning extracts; synthetic tanning materials
Special transactions & commod., not class.to kind	Explosives and pyrotechnic products
	Wool and other animal hair (excluding wool tops)
	Other artificial resins and plastic materials
	Manufactures of leather/of composition leather nes
	Furskins, tanned/dressed, pieces/cuttings of furskin
	Tin
	Uranium depleted in u235 & thorium, & their alloys
	Miscell.non-ferrous base metals employ.in metallgy
	Cinematograph film, exposed-developed, neg.or pos.
	Animals, live, n.e.s., including zoo-animals
	Armoured fighting vehicles, arms of war & ammunit.
	Coin (other than gold) not being legal tender

Segment 5	Segment 6
Eggs and yolks, fresh, dried or otherwise preserved	Meat & edible offals, salted, in brine, dried/smoked
Sugar confectionery and other sugar preparations	Butter
Synthetic rubber latex synthetic rubber reclaimed	Fish, dried, salted or in brine smoked fish
Old clothing and other old textile articles; rags	Fish, crustaceans and molluscs, prepared or preserved
Other crude minerals	Meal and flour of wheat and flour of meslines
Other inorganic chemicals	Vegetables, roots & tubers, prepared/preserved, n.e.s.
Synth. organic dyestuffs, natural indigo & colour lakes	Fruit, preserved, and fruit preparations
Regenerated cellulose; cellulose nitrate, etc.	Cocoa
Disinfectants, insecticides, fungicides, weed killers	Margarine and shortening
Starches, inulin & wheat gluten; albuminoidal subst.	Non alcoholic beverages, n.e.s.
Materials of rubber (e.g., pastes, plates, sheets, etc)	Alcoholic beverages
Articles of rubber, n.e.s.	Hides and skins (except furskins), raw
Tulle, lace, embroidery, ribbons, & other small wares	Oils seeds and oleaginous fruit, whole or broken
657: Special textile fabrics and related products	Cork, natural, raw & waste (including in blocks/sheets)
Floor coverings, etc.	Pulp and waste paper
Clay construct.materials & refractory constr.mater	Cotton
Mineral manufactures, n.e.s	Crude animal materials, n.e.s.
Glassware	Crude vegetable materials, n.e.s.
Iron/steel wire, wheth/not coated, but not insulated	Residual petroleum products, nes.& related materials
Wire products and fencing grills	Electric current
Nails, screws, nuts, bolts etc.of iron, steel, copper	Animal & vegetable oils and fats, processed & waxes
Agricultural machinery and parts	Nitrogen-function compounds
Textile & leather machinery and parts	Other organic chemicals
Food processing machines and parts	Inorganic chemical elements, oxides & halogen salts
Metal working machinery and parts	Essential oils, perfume and flavour materials
Pumps for liquids, liq.elevators and parts	Cork manufactures

Other non-electrical mach.tools, apparatus & parts	Pottery
Medical instruments and appliances	Pig iron, spiegeleisen, sponge iron, iron or steel
Office and stationery supplies, n.e.s.	Rails and railway track construction material
	Iron & steel castings, forgings & stampings; rough
	Nickel
	Cutlery
	Steam & other vapour generating boilers & parts
	Steam & other vapour power units, steam engines
	Other power generating machinery and parts
	Tractors fitted or not with power take-offs, etc.
	Paper & pulp mill mach., mach for manuf.of paper
	Printing & bookbinding mach.and parts
	Office machines
	Electric apparatus for medical purposes, (radiolog)
	Motorcycles, motor scooters, invalid carriages
	Clothing accessories of textile fabrics
	Optical instruments and apparatus
	Meters and counters, n.e.s.
	Measuring, checking, analysing instruments
	Photographic apparatus and equipment, n.e.s.
	Photographic & cinematographic supplies
	Optical goods, n.e.s.
	Watches and clocks
	Baby carriages, toys, games and sporting goods
	Jewellery, goldsmiths and other art. of precious m.

Segment 7	Segment 8
Meat & edible offals, prep./pres., fish extracts	Wheat (including spelt) and meslin, unmilled
Milk and cream	Rice
Cheese and curd	Maize (corn), unmilled
Fish, fresh (live or dead), chilled or frozen	Sugar and honey
Cereal prepar. & prepar. of flour of fruits or vegetables	Coffee and coffee substitutes
Vegetables, fresh, chilled, frozen/preserved; roots, tubers	Oil seeds and oleaginous fruit, whole or broken
Feed.stuff for animals (not including unmilled cereals)	Natural rubber latex; nat.rubber & sim.nat. gums
Other wood in the rough or roughly squared	Pulpwood (including chips and wood waste)
Stone, sand and gravel	Iron ore and concentrates
Non-ferrous base metal waste and scrap, n.e.s.	Organo-inorganic and heterocyclic compounds
Leather	Radio-active and associated materials
Wood manufactures, n.e.s.	Fertilizers, manufactured
Cotton fabrics, woven	Pearls, precious & semi-prec.stones, unwork./worked
Fabrics, woven, of man-made fibres	Silver, platinum & oth.metals of the platinum group

Made-up articles, wholly/chiefly of text.materials	Lead
Lime, cement, and fabricated construction materials	Engines & motors, non-electric
Zinc	Radio-broadcast receivers
Rotating electric plant and parts	Road motor vehicles, n.e.s.
Gramophones, dictating, sound recorders etc	Waste and scrap metal of iron or steel
Elect.app.such as switches, relays, fuses, plugs etc.	Ores and concentrates of uranium and thorium
Furniture and parts thereof	Ores and concentrates of base metals, n.e.s.
Outer garments, men's, of textile fabrics	Ores & concentrates of precious metals; waste, scrap
Under garments of textile fabrics	Coal, lignite and peat
Outer garments and other articles, knitted	Petrol & crude oils obtained from bituminous minerals
Under garments, knitted or crocheted	Gas, natural and manufactured
Art.of apparel & clothing accessories, no textile	Fixed vegetable oils, soft, crude, refined/purified
Footwear	Other fixed vegetable oils, fluid or solid, crude
	Organo-inorganic and heterocyclic compounds
	Radio-active and associated materials
	Fertilizers, manufactured
	Pearls, precious & semi-prec.stones, unwork./worked
	Silver, platinum & oth.metals of the platinum group
	Lead
	Engines & motors, non-electric
	Radio-broadcast receivers
	Road motor vehicles, n.e.s.
	Railway vehicles & associated equipment
	Aircraft & associated equipment and parts
	Ships, boats and floating structures
	Travel goods, handbags, brief-cases, purses, sheaths
	Works of art, collectors pieces & antiques
	Gold, non-monetary

References

J. E. Anderson and D. Marcouiller. Insecurity and the pattern of trade: An empirical investigation. *Review of Economics and Statistics*, 84:342–352, 2002.

J. E. Anderson and E. van Wincoop. Gravity with gravitas: A solution to the border puzzle. *American Economic Review*, 93:170–192, 2003.

B. H. Baltagi. *Econometric Anysis of Panel Data*. Wiley, Chichester, 1995.

F Cairncross. *The death of distance: how the communications revolution will change our lives.* Orion Publishing, London, 1997.

T. De Graaff, J. Boter, and J. Rouwendal. On spatial differences in the attractiveness of Dutch museums. *Environment and Planning A*, 41:2778–2797, 2009.

H. L. F. de Groot, G. M. Linders, P. Rietveld, and U. Subramanian. The institutional determinants of bilateral trade patterns. *Kyklos*, 57:103–123, 2004.

A. V. Deardorff. Determinants of bilateral trade: does gravity work in a neoclassical world? In J. Frankel, editor, *The regionalization of the world economy*, chapter 1, pages 7–32. University of Chicago Press, 1998.

A. V. Deardorff. Local comparative advantage: trade costs and the pattern of trade. Research seminar in international economics discussion paper, University of Michigan, no. 500, Ann Arbor, 2004.

A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society B*, 39:1–38, 1977.

A. C. Disdier and K. Head. The puzzling persistence of the distance effect on bilateral trade. *Review of Economics and Statistics*, 90:37–48, 2008.

R. C. Feenstra. *Advanced international trade: theory and evidence.* Princeton University Press, Princeton, 2004.

R. C. Feenstra, R. E. Lipsey, H. Deng, A. C. Ma, and H. Mo. World trade flows: 1962–2000. NBER working paper, no. 11040, Cambridge, MA., 2005.

A. S. Fotheringham and M. E. O'Kelly. *Spatial interaction models: formulations and applications.* Kluwer Academic, Dordrecht, 1989.

Th. L. Friedman. *The world is flat: a brief history of the twenty-first century.* Farrar, Straus and Giroux, New York, 2005.

L. Guiso, P. Sapienza, and L. Zingales. Cultural Biases in Economic Exchange? *Quarterly Journal of Economics*, 124:1095–1131, 2009.

J. J. Heckman. Sample Selection Bias As A Specification Error. *Econometrica*, 47:153–161, 1979.

J. Hinloopen and C. van Marrewijk. Power laws and comparative advantage. *Applied Economics*, 44:1483–1507, 2012.

G. Hofstede. *Culture's Consequences: International Differences in Work-related values*. Sage Publications, Beverly Hills, CA, 1980.

G. Hofstede. *Culture's consequences: comparing values, behaviors, instititions and organizations across nations*. Sage Publications, London, 2001.

D. Hummels. Have international transportation costs declined? Purdue university, mimeo, 2000.

D. Hummels, P. Minor, M. Reisman, and E. Endean. Calculating tariff equivalents for time in trade. USAID/Nathan Associates, 2007.

D. Kaufmann, A. Kraay, and M. Mastruzzi. Governance matters iv: governance indicators for 1996–2004. World Bank policy research paper, no. 3630, Washington DC, 2005.

B. Kogut and H. Singh. The effect of national culture on the choice of entry mode. *Journal of International Business Studies*, 19:411–432, 1988.

M. B. M. Lankhuizen, H. L. F. de Groot, and G. M. Linders. The trade-off between foreign direct investment and exports: the role of multiple dimensions. *The world economy*, 34:1395–1416, 2011.

F. Leisch. Flexmix: A general framework for finite mixture models and latent class regression in R. *Journal of Statistical Software*, 11:1–18, 2004.

G. M. Linders and H. L. F. de Groot. Estimation of the gravity equation in the presence of zero flows. Tinbergen Discussion Paper 06-072/3, Amsterdam-Rotterdam, 2006.

G. M. Linders, A. H. J. Slangen, H. L. F. de Groot, and S. Beugelsdijk. Cultural and institutional determinants of bilateral trade flows. Tinbergen Institute discussion paper, no. 2005-074/3, Amsterdam-Rotterdam, 2005.

G. M. Linders, H. L. F. de Groot, R. J. G. M. Florax, and P. Nijkamp. Persistent distance decay effects in the space-economy. In M. Jovanovic, editor, *International handbook of economic integration*, chapter 14, pages 324–350. Edward-Elgar, 2011.

P. McCann. Globalization and economic geography: the world is curved, not flat. *Cambridge Journal of Regions Economy and Society*, 1:351–370, 2008.

J. Möhlmann, S. Ederveen, H. L. F. de Groot, and G. M. Linders. Intangible barriers to international trade: a sectoral approach. In P. A. G. van Bergeijk and S. Brakman, editors, *The gravity model in international trade*, chapter 8, pages 224–251. Cambridge University Press, 2010.

M. Obstfeld and K. Rogoff. The six major puzzles in international macroeconomics: is there a common cause? NBER Macroeconomics Annual, no. 15, the MIT press, Cambridge, MA, 2000.

J. E. Rauch. Networks versus markets in international trade. *Journal of International Economics*, 48:7–35, 1999.

J. E. Rauch. Business and social networks in international trade. *Journal of Economic Literature*, 39:1177–1203, 2001.

B van Dijk, P. H. Franses, R. Paap, and D. J. C. van Dijk. Modeling regional house prices. Econometric Institute Report, 2007-55, Erasmus University Rotterdam, 2007.

M. Wedel and W. Kamakura. *Market Segmentation: Conceptual and Methodological Foundations*. Kluwer Academic Publishers, Boston, 2000.