

# The Gravity Equation and the Interdependency of Trade Costs and International Trade

Doctoral Dissertation

Stephan Rudolph

Faculty of Business Management and Economics

Technische Universität Dresden

Referees: Prof. Dr. Udo Broll, Prof. Dr. Georg Hirte

January 10, 2011

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>The Gravity Equation: Theory and Application</b>	<b>8</b>
2.1	Theoretical Concepts . . . . .	9
2.1.1	Demand Side Models . . . . .	11
2.1.2	A Supply Side Model . . . . .	16
2.1.3	Gravity Equations Derived from New-new Trade Theory . . . . .	17
2.2	Empirical Concepts . . . . .	19
2.2.1	The Basic Specification . . . . .	19
2.2.2	Dealing with Unobservable Country Characteristics: Fixed Effects .	20
2.2.3	Dealing with Zero-Trade-Flows: Non-linear Estimators . . . . .	23
2.2.4	Dealing with Endogeneity and Simultaneity: Instrumental Variables	25
2.3	Conclusions . . . . .	27
<b>3</b>	<b>A Theory of Endogenous Trade Costs</b>	<b>28</b>

## CONTENTS

---

3.1	A Theory Based Gravity Equation . . . . .	31
3.2	A Micro-founded Form of Trade Costs . . . . .	36
3.3	Economies of Scale in International Commodity Trade . . . . .	37
3.4	Interaction between the Gravity Equation and Trade Cost Function . . . . .	40
3.5	Conclusion . . . . .	43
<b>4</b>	<b>Estimation</b>	<b>45</b>
4.1	Econometric Model . . . . .	47
4.2	Data . . . . .	49
4.3	Results . . . . .	53
4.4	Conclusion . . . . .	61
<b>5</b>	<b>Computing Multilateral Resistances</b>	<b>63</b>
5.1	Background . . . . .	65
5.2	Computing Bilateral Trade Costs . . . . .	67
5.3	Computing the Multilateral Resistances . . . . .	68
5.3.1	The Equation System and its Solution . . . . .	68
5.3.2	An Illustrative Example . . . . .	70
5.3.3	Multilateral Resistances of the 23 OECD Countries . . . . .	76

## CONTENTS

---

5.4	Conclusion . . . . .	82
<b>6</b>	<b>Estimation with Multilateral Resistances</b>	<b>84</b>
6.1	Econometric Model and Data . . . . .	85
6.2	Empirical Results . . . . .	87
6.3	Conclusion . . . . .	92
<b>7</b>	<b>Summary</b>	<b>94</b>
<b>A</b>	<b>Derivation of the Gravity Equation (Anderson and van Wincoop)</b>	<b>100</b>
<b>B</b>	<b>Derivation of the Index for Bilateral Trade Costs</b>	<b>104</b>
<b>C</b>	<b>Multilateral Resistances as a Linear Equation System</b>	<b>106</b>
	<b>Bibliography</b>	<b>108</b>

# List of Figures

1.1	Regional Structure of the World Trade . . . . .	3
1.2	Organization of the Study . . . . .	7
2.1	Distribution of Country Pairs Based on Direction of Trade . . . . .	24
3.1	Plot of the Bias Term with a Given Value for the Elasticity of Substitution	42
4.1	Trade Cost Index of Germany . . . . .	53
4.2	Residual-Analysis of the Standard Gravity Specification . . . . .	55
4.3	Residual-Analysis of the Theory-Based Gravity Equation . . . . .	60
4.4	Residual-Analysis of the Theory-Based Trade Cost Equation . . . . .	60
5.1	Simulation Results of the Three Country Example (Upward Approximation)	76
5.2	Simulation Results of the Three Country Example (Downward Approximation) . . . . .	77
5.3	Countries with Low Multilateral Resistances . . . . .	81

## LIST OF FIGURES

---

5.4	Countries with High Multilateral Resistances . . . . .	81
5.5	Multilateral Resistances of Australia, Japan, Korea, and the U.S. . . . .	82
6.1	Residual-Analysis of the Standard Gravity Specification . . . . .	90
6.2	Residual-Analysis of the Standard Gravity Specification . . . . .	90

# List of Tables

- 4.1 Summary statistics of the OECD data set . . . . . 54
- 4.2 Results of the Standard Specification . . . . . 56
- 4.3 Results of the Specification with Trade Cost Endogeneity . . . . . 57
  
- 5.1 Assumed Data for the Numerical Example . . . . . 72
- 5.2 Simulation of the Numerical Example . . . . . 75
- 5.3 Amplitude and Convergence of the Simulation . . . . . 78
- 5.4 Descriptive statistics of  $\Delta_1 = P_{i(100)} - P_{i(99)}$  . . . . . 78
- 5.5 Summary statistics of the Multilateral Resistances . . . . . 80
  
- 6.1 Summary statistics of the OECD data set . . . . . 86
- 6.2 Results of the Standard Specification . . . . . 89
- 6.3 Results of the Simultaneous Equation Model . . . . . 91

# Chapter 1

## Introduction



The gravity equation is probably the most important tool in international economics to explain and estimate trade flows. In its simplest form, it states that the exports between any two given countries (say  $i$  and  $j$ ) are a multiplicative function of these countries' economic size, as measured by GDP, and their bilateral trade costs:

$$\text{Exports}_{ij} = \frac{\text{GDP}_i \times \text{GDP}_j}{\text{Trade Costs}_{ij}}. \quad (1.1)$$

The idea goes back to the pioneering works of Tinbergen (1962) and Pöyhönen (1963), who developed gravity equations independently of each other. The name comes from the similarity of equation (1.1) to Isaac Newton's law of gravity where the attraction force between two physical bodies equals the product of their masses divided by the squared distance between the bodies.<sup>1</sup>

Figure 1.1 illustrates the idea behind the gravity equation with some data. It plots the export values among world regions. For example, the tallest column in the corner symbolizes the exports from European countries to European countries. With a value of \$4,243 billion (USD) they account for one third of world trade.<sup>2</sup> On the one hand, it can be seen from this figure that trade activities within regions that contain relatively large economies (Europe, Asia and North America) are very high. On the other hand, regions with rich countries seem to attract trade with regions containing poor countries. For example, the trade between Europe and Africa is higher than the trade inside Africa. Thus, there are obviously two contrary effects on trade:

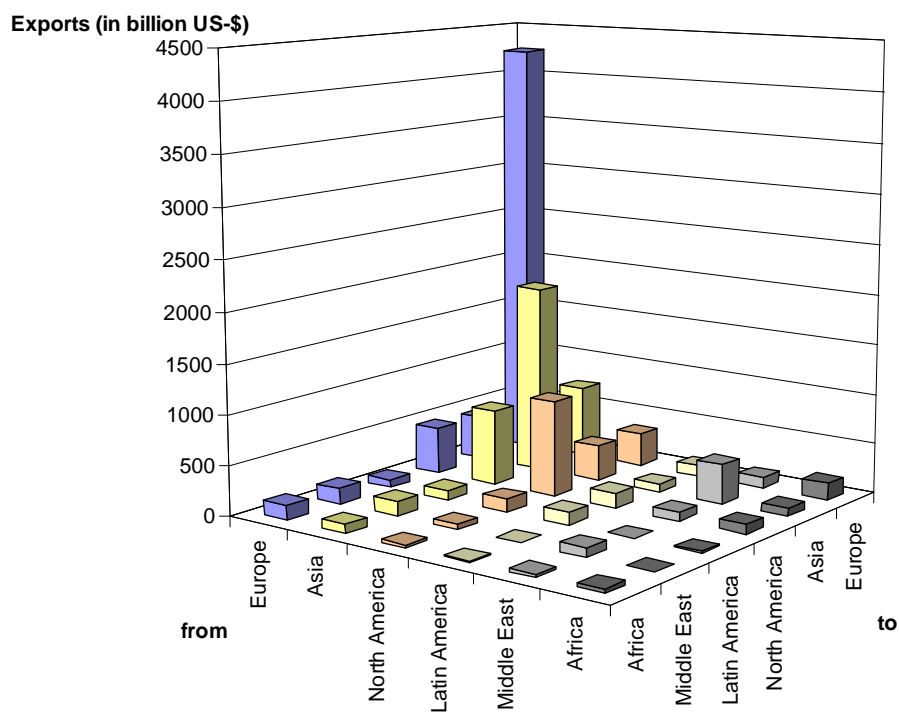
- economic size increases trade flows and
- distance or trade costs decrease trade flows.

---

<sup>1</sup> The explicit formulation of Newton's physical gravity equation to explain the attraction force between two bodies  $i$  and  $j$  is:

$$\text{Attraction Force}_{ij} = \frac{\text{Mass}_i \times \text{Mass}_j}{\text{Distance}_{ij}^2}.$$

<sup>2</sup> Europe has a much higher export value than North America because it consists of multiple large economies. For North America the USA is treated as one unit. If the individual, American states were treated separately, such that "interstate trade" were recognized, the value for North American exports might be closer in magnitude to that of Europe.



**Figure 1.1:** Regional Structure of the World Trade.  
*Source: WTO International Trade Statistics (2008).*

Gravity equation (1.1) becomes estimable after log-linearizing and parameterizing it. Bilateral export and GDP data are broadly available in several databases. Trade costs are not directly measurable and are therefore usually proxied by geographic distance and a set of further proxy variables like: access to the sea, common border, common language, membership in a certain group of countries, and others. The trade cost proxies can be subdivided into *geographical* and *political* variables. Geographical properties of a country can hardly be changed by policymakers. If a country is located on a small, remote island in the ocean or between many industrialized countries on a continent, if it has access to the sea or not, if it is small or large, must be taken as exogenously given. Yet, policymakers can influence trade costs through tariff rates, currency unions, free trade agreements, membership in certain country groups and many other measures.

The estimated effects of these policy driven trade cost proxies are frequently used in the consultation of policymakers. For example, there are numerous studies about the role of the WTO in fostering trade. This question was raised by Rose (2004) who found no clear evidence that GATT/WTO members have more trade activities. However, later studies criticize the Rose study and find that a GATT/WTO membership significantly improves trade (e.g. Subramanian and Wei, 2007; Tomz, Goldstein, and Rivers, 2007; Herz and Wagner, 2007).<sup>3</sup> In a similar way, the trade effects of currency unions (Rose, 2000; Rose and van Wincoop, 2001; Nitsch, 2002),<sup>4</sup> the Group of Eight (Nitsch, 2007a), the EU-enlargement of 2004 (Fuchs and Wohlrabe, 2008), borders between countries and states (for USA-Canada-trade McCallum, 1995; Nitsch, 2002, for Eastern-Western German trade), state visits (Nitsch, 2007b), terrorism (Nitsch and Schumacher, 2004), mega events like the Olympics (Rose and Spiegel, 2009) and many other possible influences on international trade are studied, only to give a few examples of the many possibilities to use the gravity equation.

However, if the gravity equation is important for political decisions, it is very important to

---

<sup>3</sup> Furthermore, there are studies using the gravity equation to compute trade potentials of a WTO accession for certain countries, e.g. Babetskaia-Kukharchuk and Maurel (2004) for Russia.

<sup>4</sup> See also Baldwin (2006) for a critical review of this literature.

achieve reliable results from its empirical application. Thus, it is necessary to employ the gravity equation using a theoretically and empirically proper methodology. One important discussion addresses the implausibly high measures for the impact of trade cost proxies on exports which frequently appear, especially in older works. McCallum (1995) discovered the effect of the US-Canadian border. He concluded that this border increases trade between a certain Canadian province and another Canadian province by a factor 22 (2,200 percent) compared to trade between the same Canadian province and a US-state of the same economic size and the same distance like the other province. Anderson and van Wincoop (2003) were able to show that this extremely high and unexpected result is wrong because certain theoretical considerations and empirical implications were not respected in McCallum's study.<sup>5</sup> They argue that the gravity equation in its basic form (1.1) is misleading and that trade costs must more exactly be related to the countries' overall trade costs with the rest of the world. They derive an index measuring these overall trade costs and call it *multilateral resistances* to trade (as bilateral trade costs can be interpreted as bilateral trade resistances).

The aim of this study is to contribute to the discussion about the suitability of the gravity equation's empirical applications. The basic idea is that trade costs between two countries could additionally depend on the exports between these two countries and not only on the (more or less) exogenous proxy variables for trade costs, as they are normally used. Shipping goods from one country to another, overcoming geographic, political and cultural borders, requires an infrastructure which is likely to yield economies of scale. Per-unit trade costs of sending a small amount of a certain good to another country are likely to be more expensive than per-unit trade costs of sending a large amount. In this study, I show that ignoring this *reverse causality problem* might overestimate the effects of the right-hand-side variables on trade if these effects are interpreted as direct effects. Since trade costs are not directly measurable, I will use a novel index of comprehensive trade costs (Novy, 2007) to estimate a simultaneous system, first of a gravity equation and second of a trade cost equation. I find that these scale effects appear and that ignoring

---

<sup>5</sup> See also Baldwin and Taglioni (2006) for a comprehensive discussion of studies using the gravity equations to estimate the Euro's trade effect with respect to theoretical and empirical adequacy.

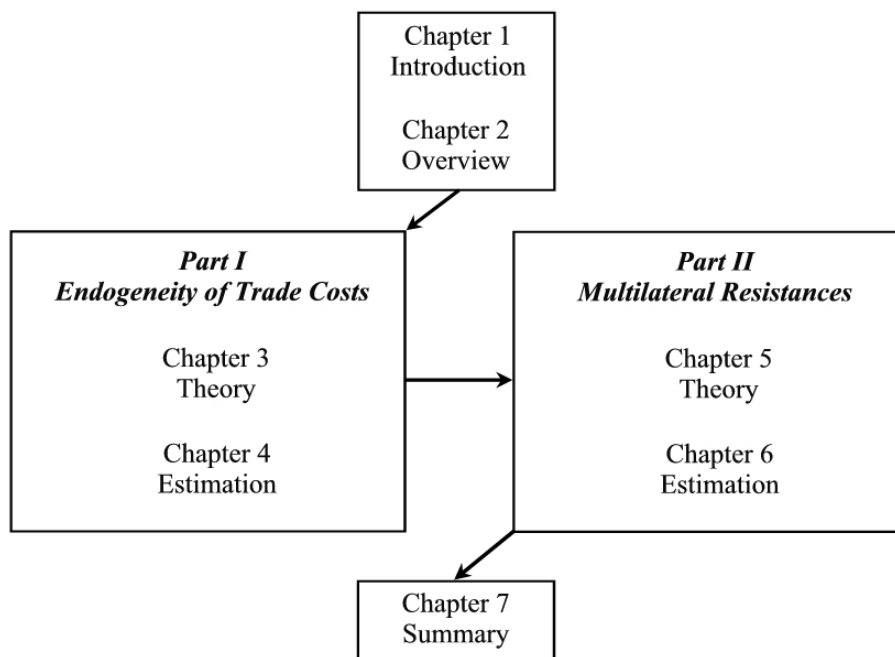
the reverse causality correction systematically increases the estimated coefficients.

A further contribution of this study is in its use of the comprehensive trade cost index to compute multilateral resistances of countries to trade, introduced in the trend-setting work by Anderson and van Wincoop (2003). These multilateral resistances are necessary to retrieve unbiased results from empirical gravity equations. They are defined as a weighted summation over all countries' trade costs from a certain country's view and can be interpreted as a country's (adjusted) trade costs with all other countries. The index contains measurable shares of the countries' GDPs relative to the world's GDP as well as unmeasurable trade costs. Because of this unmeasurable component and because of the complexity and mutual interaction of the countries' multilateral resistance term, they are usually controlled by country or country-pair dummies. In this study, I shall demonstrate a new way to solve the complex equation system of multilateral resistances and compute them for a set of 23 OECD countries.

The study is structured as follows. Chapter 2 overviews the basic literature dealing with the gravity equation. It introduces the most important theoretical foundations of the gravity equation which appear to be consistent with all three branches of the economic theory on international trade: the *classical/neo-classical theory* (Ricardo and Heckscher-Ohlin), the *new trade theory* (Krugman, 1979), and the *new-new trade theory* (Melitz, 2003).<sup>6</sup> Subsequently it shows how to deal with three problems which frequently arise: first, the treatment of unobservable country-specific effects with country and country-pair dummies; second, the treatment of zero trade flows with non-linear estimators; and third, the treatment of endogeneity or simultaneity with instrumental variables (IV regression). After this introduction of the "state of the art", chapter 3 introduces a simple theory of endogenous trade costs. Replacing the trade cost measure in a theory-based gravity equation (in this study the gravity equation by Anderson and van Wincoop, 2003) makes it possible to show that the direct effect of a right-hand side variable on exports is biased upward. Chapter 4 tests the persistence of trade cost-related economies of scale and checks

---

<sup>6</sup> For an introduction into the classical/neoclassical trade theory and the new trade theory see the textbook by Krugman and Obstfeld (2009). For a more detailed description see Feenstra (2004). The new-new trade theory has not yet gained access into standard textbooks on international economics.



**Figure 1.2:** Structure of the Study.

the bias of the estimated parameters using a simultaneous equation system (consisting of a gravity function and a trade cost function) with a comprehensive measure for trade costs.

Chapter 5 introduces a method to solve the complex system of multilateral resistances with a numerical program. As a result, values of the multilateral resistances of 23 OECD countries are computed. Thus, all right-hand side variables of the Anderson/van Wincoop gravity equation are available and therefore this equation becomes estimable. Chapter 6 presents the results of estimating the gravity equation in the traditional way and in a simultaneous equation model. Chapter 7 summarizes the outcomes of this study.

Chapters 3 and 4 can be seen as one major section of the study introducing and testing a theory of endogenous trade costs. Chapters 5 and 6 build another major section, solving for multilateral resistances and directly estimating a fully theory-based gravity equation. The single chapters of this study were written in a way such that each chapter can be read independently from each other. Figure 1.2 illustrates the structure of this study.

## Chapter 2

# The Gravity Equation: Theory and Application

Much work has been done, since Tinbergen (1962) and Pöyhönen (1963) presented the seminal idea that the economic size of two countries and the distance between these two countries could explain their bilateral trade volume. In these first studies, bilateral trade flows were simply regressed on the given countries' respective incomes and the geographic distance between the countries as a proxy variable for trade costs using ordinary least-squares estimators.<sup>1</sup> The gravity equation became a useful instrument to study the effects of trade barriers, especially policy-driven trade barriers, on exports.<sup>2</sup> After rising criticism that the gravity equation was a purely intuitive and not theoretically founded empirical tool, Anderson (1979) was one of the first who developed a theoretical framework to derive the gravity equation in its essential form. Over the last 20 years, the gravity equation was derived from several trade models and appeared to conform with the different branches of trade theory. These theoretical considerations as well as the technical progress in econometrics have helped to justify and improve the application of gravity equations as an instrument to measure the determinants of international trade. This chapter provides an overview about theoretical foundations of the gravity equation and the requirements for preferably unbiased empirical results.

### 2.1 Theoretical Concepts

The current theory of international trade rests on three different foundations. The first foundation is the *classical/neo-classical theory* with its well known Ricardo model (comparative differences between countries in technology) and Heckscher-Ohlin model (comparative differences between countries in factor endowment). This theory traces interna-

---

<sup>1</sup> Tinbergen (1962) additionally included trade cost control dummies for the European Community and BeNeLux (Belgium, Netherlands and Luxemburg). His PhD-student Linnemann (1966) was the first who extended the gravity equation with per capita income and more trade cost control variables in his dissertation. He also tried to derive the gravity equation from a Walrasian general equilibrium model. However, too many explanatory variables appear for trade flows in his model such that a simple reduction to the gravity equation is not feasible.

<sup>2</sup> Examples for such early studies are Aitken (1973), who studied the trade effects of EEC, EFTA, and Aitken and Obutelewicz (1976), who studied the trade effects of colonial links between EEC-countries and former colonies.



tional trade flows back to comparative advantages between countries. The statement of this theory is that countries specialize in goods in whose production they have comparative advantages, and sell these goods in international markets against goods in whose production they have comparative disadvantages. For example, if France is relatively better in producing wine and Germany in producing beer, France specializes in wine and exports a share of the production to Germany while Germany specializes in beer and exports a share of the production to France.

The second foundation – the *new trade theory* by Krugman (1979) – criticizes these classical/neo-classical models because their basis for trade is that countries are (comparatively) different, although the highest volume of trade appears between very similar countries (for example, among EU-countries). Furthermore, these similar countries trade goods from the same sectors, meaning that France and Germany, for example, export *and* import the same goods (like cars) from each other rather than completely different goods. He argues that the integration of markets makes it possible to realize economies of scale. They lower per-unit costs of production and therefore prices. Market integration also increases the available number of product varieties. If consumers value the availability of many differentiated products (“love of variety”) and each available good gains lower prices from economies of scale, this is a source for gains from trade between similar countries.

The third and most recent foundation – the *new-new trade theory* – goes back to Melitz (2003). His central argument is that it is not countries but firms that export goods to foreign countries and that these firms are heterogeneous: it depends on their productivity, whether they will export or not (the so-called “extensive margin”) and how much a certain firm will export (i.e. the “intensive margin”). A large and highly productive firm is more likely to overcome trade barriers and export a substantial volume to another country than a small and less productive firm which perhaps cannot even overcome the trade barriers to any foreign market and therefore does not export at all.

This section reviews the theoretical literature on the gravity equation with respect to these different foundations of trade theory. It starts with the pioneering demand side

model by Anderson (1979). This model is the basis of several further models which have in common that trade flows are determined by the demand side. It builds on the assumption of complete specialization of countries. Deardorff (1998) noted that – under certain conditions presented in this section – the foundation is consistent with the classical/neo-classical as well as with the new trade theory. Eaton and Kortum (2002) presented a supply side model where countries have access to different levels of technology to produce a continual set of goods and derive a gravity-like equation. The setup is compatible to a Ricardian trade model with a continuum of goods. Very recent literature derives the gravity equation directly from the new-new trade theory (Chaney, 2008; Melitz and Ottaviani, 2008).

### 2.1.1 Demand Side Models

To derive the gravity equation, demand side models act on the assumption that countries are exogenously endowed with a certain supply of goods. The flows of these goods from one country to another (the trade flows) are thus driven by the demand of the target country.

The first models which derive the gravity equation in its characteristic multiplicative form were presented by Anderson (1979). In the simplest formulation he assumes that each country ( $i, j \in \{1, \dots, C\}$ , where  $C$  is the number of all countries) is endowed with a certain GDP:  $Y_i$  and  $Y_j$ , respectively. GDP is hereby assumed to be the endowment of each country with a certain differentiated, tradeable good which is characteristic for the respective country. Imagine for example that each country is endowed with a combined bundle of many goods and consumers distinguish between these bundles by the label “made in  $i$ ” or “made in  $j$ ”. In this simplest form of the model, trade costs are ignored. The preferences are assumed to be identical over all countries and represented by a Cobb-Douglas utility function. Therefore, each consumer worldwide will spend the same fraction of income on a certain (country-differentiated) good. Consequently, the spending on exports from country  $i$  in country  $j$ ,  $X_{ij}$ , can be expressed as the share of country  $j$ ’s

GDP multiplied with the income share  $s_i$  that consumers spend for the composite good of country  $i$ :<sup>3</sup>

$$X_{ij} = s_i \cdot Y_j. \quad (2.1)$$

In a general equilibrium, all exports of country  $i$  in all countries  $j$  (including the “intra-country” exports from country  $i$  to country  $i$  itself,  $X_{ii}$ ) must equal the GDP, or in other words, the sales must equal the income:

$$Y_i = s_i \cdot \sum_j Y_j = s_i \cdot Y_w, \quad (2.2)$$

with  $Y_w = \sum_j Y_j$  being the world’s GDP. Combining equation (2.1) and (2.2) yields the probably simplest form of the gravity equation:

$$X_{ij} = \frac{Y_i \cdot Y_j}{Y_w}, \quad (2.3)$$

where the world GDP  $Y_w$  can be treated as a constant. This gravity equation is thus able to explain bilateral trade flows by the respective countries’ GDPs in a multiplicative form.

However, this simplistic form is not yet able to account for trade costs. Anderson (1979) provides several augmentations of this simple model. He introduces non-tradeable goods, trade costs, many commodity classes of goods, and more general constant elasticity of substitution (CES) preferences instead of Cobb-Douglas preferences. These augmentations lead to more complex versions of the gravity equation compared to equation (2.3). One result of these augmentations is that the effect of trade costs on exports between two countries increases with the elasticity of substitution between the countries’ specific goods. The intuition is: if a country’s specific good is more substitutable with the specific good of any country exporting into this certain country, the probability that consumers in this country are willing to pay price markups due to trade costs will be lower. They

---

<sup>3</sup> More detailed, this is equivalent to the demand function resulting from the explicit Cobb-Douglas utility function  $U_j = \prod_i c_{ij}^{s_i}$ , with income shares  $s_i$  summing up to one,  $\sum_i s_i = 1$ , and  $c_{ij}$  being the amount of consumption of country  $i$ ’s specific good in country  $j$ . The value of exports between country  $i$  and  $j$  is then  $X_{ij} = p_i \cdot c_{ij}$  where  $p_i$  is the price of country  $i$ ’s good.

can substitute the imported goods more easily by the domestic goods in this case. Furthermore, Anderson (1979) derives a version of the gravity equation using a setup with trade costs and CES preferences. In this version, not only bilateral trade costs affect the exports between countries but rather the ratio of bilateral trade costs to (an adjusted measure of) trade costs with all other countries, as represented by a Dixit-Stiglitz price index over all goods.

Building on this framework, Anderson and van Wincoop (2003) develop a very adaptable version of the gravity equation using the generalization with CES preferences. They show that exports in gravity equations do not only depend on bilateral trade costs but rather on a ratio of bilateral trade costs and the respective two countries' trade costs to all countries as well. The index that measures a country's overall resistance to trade is called *multilateral resistance*. Abbreviating this by "mr", their gravity equation can be displayed as:<sup>4</sup>

$$X_{ij} = \frac{Y_i \cdot Y_j}{Y_w} \times \left( \frac{\text{mr}_i \cdot \text{mr}_j}{\text{trade barriers}_{ij}} \right)^{\text{elasticity of substitution}-1} \quad (2.4)$$

This result is in contrast to the traditional gravity equation (1.1) which only considers bilateral trade barriers. Why should multilateral resistance play a role in explaining bilateral trade flows? Imagine two countries, 1 and 2, trading with each other and country 2 signs a free trade agreement with a third country 3. If this free trade agreement leads to a trade diversion effect, the trade volume between countries 1 and 2 is likely to decline. This is actually the case if trade barriers between countries 1 and 2 and their respective GDPs are unaffected. For the trade between countries 1 and 2, equation (2.4) is able to take the free trade agreement between countries 2 and 3 into account, since this free trade agreement reduces the multilateral resistances – the trade barriers to all countries of the world – of country 2: Trade between 1 and 2 becomes relatively more expensive for country 2.

The model by Anderson (1979) was used to bring the gravity equation in line with the different trade theories. Helpman and Krugman (1985, chapter 8) and Helpman (1987)

---

<sup>4</sup> Since this approach is the conceptual basis for the following chapters, its derivation will be made explicit in chapter 3. For the calculus see appendix A.

conclude that the approach by Anderson (1979) in its simplest form is in accordance with the new trade theory, which states that relatively similar countries trade more actively amongst each other. They derive the following version of a gravity equation, whereby the trade flows inside a group of countries are related to a measure for the dissimilarity of countries – the so-called index of dispersion. They conclude that if the countries of a given group are more similar, trade in this group is higher. For illustration, assume that there is a region  $A$  consisting of two countries  $i$  and  $j$ . The trade volume between these two countries and therefore the trade volume of this region,  $TV_A = X_{ij} + X_{ji}$ , can be used to rewrite (2.3) as:<sup>5</sup>

$$TV_A = 2 \cdot \frac{Y_i \cdot Y_j}{Y_w} = 2 \cdot s_i \cdot s_j \cdot Y_w. \quad (2.5)$$

Let the GDP of this region be  $Y_A = Y_i + Y_j$ , and the share of region  $A$  relative to the world GDP be  $s_A \equiv Y_A/Y_w$ . Furthermore we can denote the GDP-shares of the two countries in region  $A$  as  $s_{iA} \equiv Y_i/Y_A$ , country  $j$  analogously. This makes it possible to rewrite equation (2.5) as:<sup>6</sup>

$$\frac{TV_A}{Y_A} = 2 \cdot s_{iA} \cdot s_{jA} \cdot s_A. \quad (2.6)$$

Since region  $A$  consists of two countries, it must be the case that their shares of the region's GDP sum up to one,  $s_{iA} + s_{jA} = 1$ . Squaring both sides of this condition yields  $s_{iA}^2 + 2s_{iA}s_{jA} + s_{jA}^2 = 1$  and after rearranging:

$$2 \cdot s_{iA} \cdot s_{jA} = 1 - s_{iA}^2 - s_{jA}^2.$$

Substituting this back into equation (2.6) yields:

$$\frac{TV_A}{Y_A} = s_A \cdot \left( 1 - \sum_{i \in A} s_{iA}^2 \right). \quad (2.7)$$

---

<sup>5</sup> Because  $s_i \equiv Y_i/Y_w$ .

<sup>6</sup> The reason is that

$$\frac{TV_A}{Y_A} = 2s_{iA}s_{jA}s_A = 2 \cdot \frac{Y_i}{Y_A} \cdot \frac{Y_j}{Y_A} \cdot \frac{Y_A}{Y_w} = 2 \frac{Y_i \cdot Y_j}{Y_A \cdot Y_w} = 2 \cdot \frac{Y_i}{Y_w} \cdot \frac{Y_j}{Y_w} \cdot \frac{Y_w}{Y_A} = 2 \cdot s_i \cdot s_j \cdot \frac{Y_w}{Y_A}.$$

This is equation (2.5) multiplied by  $1/Y_A$ .

Equation (2.7) shows that the size-adjusted trade volume inside region  $A$  equals the world share of this region's GDP multiplied by the so-called index of dispersion, which is the expression in parentheses. This dispersion index takes higher values if the single income shares  $s_{iA}$  are quite similar and it takes lower values if they are more divergent.<sup>7</sup> Note that equation (2.7) also holds for regions with more than two countries (Helpman, 1987). Also note that region  $A$  can contain all countries of the world as a special case.<sup>8</sup> The import of the literature following Helpman and Krugman (1985) is therefore: Similarity of the countries' sizes in a region raises the trade volume inside this region. This underlines the outcomes of the new trade theory, according to which more comparable economies trade more. Thus, empirical versions of equation (2.7) were used in several empirical studies to test this context.<sup>9</sup>

Bergstrand generalizes the model provided by Anderson (1979) and uses production frontiers with constant elasticities of transformation to model the countries endowments with goods (see Bergstrand, 1985, 1989, 1990).<sup>10</sup> Bergstrand (1985) derives a gravity equation augmented by price indices using a general equilibrium model with differentiated goods and one production factor (labor). These price indices have a similar logic to the multilateral resistances from Anderson and van Wincoop (2003): Considering trade costs in a gravity equation requires their relation to a measure of multilateral prices or trade costs. Bergstrand (1989) extends this framework by an additional factor (capital). This latter model thus combines elements of the Heckscher-Ohlin trade theory (two sectors, two factors) with elements of the new trade theory (monopolistic competition and product differentiation between the firms of each sector). The result is a gravity equation

---

<sup>7</sup> For example, in a two-country set, this index takes a value of  $0.5 (= 1 - 0.5^2 - 0.5^2)$  if both countries have the same size and therefore a 50% share of region  $A$ 's GDP, respectively. If the countries' sizes are completely different, so that hypothetically one country has a 100% share of the region's output and the other country 0%, the index takes the value  $0 (= 1 - 1^2 - 0^2)$ .

<sup>8</sup> In this special case,  $Y_A = Y_w$ ,  $s_A = 1$  and  $s_{iA} = s_i$ .

<sup>9</sup> See Helpman (1987) for OECD-countries, as well as Hummels and Levinsohn (1995) and Debaere (2005), who find contradictions to the new trade theory in an extension for non-OECD-countries. See also Feenstra (2004, chapter 5, pp. 146 ff.) for an overview of this literature.

<sup>10</sup> Although the modeling of production frontiers has characteristics of a supply side model, the production frontier is not directly derived from factor endowments or production functions. Therefore, Bergstrand's models should be taken rather as demand side models.

augmented by price indices as well as capital-labor ratios. This approach is a theoretical reason for many empirical studies to include per capita incomes into the gravity equation to control for capital-labor endowments.<sup>11</sup> In Bergstrand (1990) he uses this hybrid model to study the prevalence of intra-industry trade.

Deardorff (1998) uses the model proposed by Anderson (1979) and gives an intuitive explanation that the assumed specialization of each country in a particular good is also consistent with classical trade theory, especially with the Heckscher-Ohlin model under certain circumstances (many goods and fewer factors). His argument is that complete specialization results from both the Ricardo model as well as the Heckscher Ohlin model (with many goods and fewer factors) if factor prices are not equalized due to trade costs. He concludes with emphatic respect to Helpman (1987) that it is dangerous to use the gravity equation to legitimize the success of a particular trade theory because it is compatible with many strands of the different trade theories.<sup>12</sup>

### 2.1.2 A Supply Side Model

Eaton and Kortum (2002) pursue an alternative setup compared to the demand side models following Anderson (1979). They derive a gravity like equation from a Ricardian trade model with a continuum of goods. The common assumption of the demand side models is that the production values of goods are exogenously given and that the distribution over countries results from the consumer preferences. By contrast, in the model by Eaton and Kortum (2002), the distribution of goods between countries is driven by technology differences on the supply side rather than by consumer preferences. In their model, each country can potentially produce each single good in a continual range of goods. However, only the country with the lowest comparative production costs (inclusive of trade costs) for

---

<sup>11</sup> Note that the inclusion of per capita incomes into empirical gravity equations to control for comparative differences was already introduced by Linnemann (1966).

<sup>12</sup> Notably, Evenett and Keller (2002) use structural implications from the new trade theory model and the Heckscher-Ohlin model (as they are outlined in Helpman and Krugman, 1985, chapters 7 and 8) to develop a gravity equation that can distinguish between the different trade structures.

a certain good will provide all other countries with this good. Production costs depend on productivity, which is drawn from a probabilistic country-specific Fréchet function. This distribution function is determined by two parameters:

1. the average productivity level for each respective country  $i$ , which varies over the countries,
2. the productivity differences between all countries, which takes the same value in each country's distribution function as it describes a property of the whole (model) world.

The resulting equation appears to be quite similar to the results of the advanced versions of the models following Anderson (1979), especially Anderson and van Wincoop (2003).<sup>13</sup> In their equation, the effect of trade costs on exports is not explained by the elasticity of substitution of the traded goods but by the parameter for the productivity differential between countries. Lower values of this parameter indicate higher productivity differences between the countries and thus greater opportunities for comparative advantages. More comparative advantages diminish the negative effects of trade barriers in their equation. Moreover, trade costs are related to an input price measure over all destination countries, which is a comparable result to the multilateral resistance approach.

### 2.1.3 Gravity Equations Derived from New-new Trade Theory

Recently, the so-called new-new trade theory based on Melitz (2003) has attracted a lot of interest. This theory emphasizes that, above all, firm characteristics and not country characteristics lead to trade. One important feature of this literature is that there occurs

1. an *intensive margin* for trade that measures the export value of the heterogeneous firms,

---

<sup>13</sup>See also Anderson and van Wincoop (2004, p. 709 f.) for a discussion of this model.



2. an *extensive margin* that measures how many firms will be productive enough to be able to export to other countries.

Chaney (2008) expands this theory to derive a gravity equation. In his model, there is one (reference) sector producing a homogeneous good under constant returns to scale and a continuum of sectors, each producing a differentiated good under increasing returns to scale. The firms in the sectors with increasing returns draw their productivity from a sector-specific Pareto distribution, with a parameter determining the degree of firm heterogeneity in each respective sector.<sup>14</sup> Trade costs are modeled with a fixed component for each pair of countries. Chaney (2008) derives a gravity equation wherein the effect of trade costs on exports increases with the elasticity of substitution. This finding is consistent with the findings of the demand side literature. In his gravity equation, trade costs are also relative to a remoteness index comparable with multilateral resistances. Chaney's innovation is that the effect of trade costs on exports decreases in firm heterogeneity, measured by the parameter of the Pareto distribution, and thus does not only depend on the elasticity of substitution.

Melitz and Ottaviani (2008) have developed a gravity equation, using a heterogeneous firms model, replacing the assumption of a constant elasticity of substitution by quasi-linear preferences. Consequently, the markups of firms are no longer exogenously given but more generally depend on market size and integration. Similar to Chaney (2008), in their gravity equation the effect of trade costs on exports is determined by the firm-heterogeneity parameter. Contrarily, in their framework, trade costs are not related to a direct measure of remoteness to other countries. Rather, they are related to comparative advantages in technology of the exporting country and the intensity of competition (in terms of a marginal cost cut-off) of the importing country.

---

<sup>14</sup>The distribution function for productivity in terms of output per unit of labor is equal for each country, meaning that all countries have access to the same technology. Country differences enter the model via differences in the countries' reference wages from the homogeneous good sector that produces with constant returns to scale.

## 2.2 Empirical Concepts

The insight that the gravity equation is obviously consistent with the theory of international trade has effectuated a renaissance of applying the gravity equation to estimate the effects of trade barriers on the trade volumes between countries. New developments in econometric modeling have made it possible to improve the explanatory power of these applications.

This section reviews important results of the econometric literature. It focuses on three econometrical problems that are likely to appear in empirical applications. These problems are also relevant in the remainder of this study. I shall introduce the basic empirical specification of the gravity equation and discuss solutions for the following problems:

1. unobservable country characteristics,
2. the presence of zero trade flows,
3. endogeneity of right-hand side variables or their interdependency with the left-hand variable (bilateral exports).

### 2.2.1 The Basic Specification

The first studies use cross-section data. Usually, the common estimation strategy in these studies was simply to use OLS for the logarithmized values of the variables.<sup>15</sup> Because of the scarce availability of accurate proxies for trade costs besides geographical distance, it became common to use dummy variables that take the value of 1 if a certain country or country pair satisfies a certain condition (e.g. membership in the EU, landlocked, common

---

<sup>15</sup>With the notable exception of Pöyhönen (1963), who already used a non-linear estimator in his pioneer work.

border) and 0 else. The empirical form of these gravity equations can be displayed as:

$$\ln X_{ij} = \pi_1 + \pi_2 \ln Y_i + \pi_3 \ln Y_j + \pi_4^k w_{ij}^k + \eta_{ij}, \quad (2.8)$$

where  $X_{ij}$  are exports from country  $i$  to  $j$ ,  $Y_i$  and  $Y_j$  are the respective countries' GDPs,  $w_{ij}^k$  a vector of  $k$  trade cost proxies (like log of distance, and dummies like membership in a certain country group, linguistic and geographical patterns). The  $\pi$ 's are the parameters which are to be estimated and  $\eta_{ij}$  is the disturbance term.

### 2.2.2 Dealing with Unobservable Country Characteristics: Fixed Effects

Mátyás (1997) was one of the first who recognized the requirement of fixed-effect dummies in the empirical application of gravity equations. From an econometrical point of view, one must control for unmeasurable country characteristics by introducing dummy variables for each exporting and importing country.<sup>16</sup> One of these unobservable country characteristics are the multilateral resistances. Their importance was highlighted in the theoretical overview of the previous section with reference to works by Bergstrand (1985) and Anderson and van Wincoop (2003).<sup>17</sup> If the data set has a panel structure and therefore a time dimension in addition, it becomes necessary to include time dummies (e.g. for each year a dummy variable that is 1 if the considered data is from the respective year and 0 else). They control for special circumstances of a certain time unit, e.g. if there was an economic boom or depression. Note that there are plenty of arguments that using panel data yields more advantages than cross-sectional data, since panel data contains

---

<sup>16</sup>It is worth emphasizing that Pöyhönen (1963) already was aware of the presence of specific effects of the exporting and importing countries on trade. But he used his estimate of the model's intercept to disentangle the country characteristics instead of using fixed or random effects as it is common in modern studies.

<sup>17</sup>Bergstrand (1985) uses price indices in his estimation, but they appear to be a weak proxy. Anderson and van Wincoop (2003) use a non-linear program to solve for multilateral resistances, but the computational effort is very high. They conclude that results with country fixed-effect dummies are quite similar and much easier to handle compared to the complex approach.

much more information (see Baltagi, 2005, chapter 1.2, for a discussion).

Mátyás (1997) argues that ignoring these dummies leads to an estimation bias. In Mátyás (1998), he argues that under certain circumstances, especially if the data set contains a large number of countries, it might be better to use a random effects estimator (where country characteristics are seen as random variables and captured by multiple error terms) instead of a fixed-effects approach (where country characteristics are controlled by a separate intercept for each exporting and importing country). However, in most cases fixed-effects estimators are more appropriate, since using the random effects approach requires that there is no correlation between the country characteristic and the regressors (Egger, 2000).<sup>18</sup>

The empirical gravity equation with country and time fixed-effects, as it is suggested by Mátyás (1997), and augmented by a time index  $t$  can be denoted as:

$$\ln X_{ijt} = \pi_1^i + \pi_1^j + \pi_1^t + \pi_2 \ln Y_{it} + \pi_3 \ln Y_{jt} + \pi_4^k w_{ijt}^k + \eta_{ijt}, \quad (2.9)$$

where  $\pi_1^i$ ,  $\pi_1^j$  and  $\pi_1^t$  are the vectors of the exporting country, importing country, and time dummies. This specification is known as a *two-way model*, because it considers country as well as time characteristics. In comparison, a *one-way model* would only control for country characteristics but not for time. A *pooled regression model* would ignore all country and time specific effects, as in the basic empirical gravity equation (2.8).

Other papers starting with Hummels and Levinsohn (1995) pursue a *three-way model*. They replace export and import country dummies by country-pair dummies. Their model

---

<sup>18</sup>Basically, random effects are preferable if there is a large number of individuals in the data (for example household-level data like micro census). Whether random effects or fixed-effects are appropriate is usually tested by Hausman's specification test.

therefore has the form<sup>19</sup>

$$\ln X_{ijt} = \pi_1^{ij} + \pi_1^t + \pi_2 \ln Y_{it} + \pi_3 \ln Y_{jt} + \pi_4^k w_{ijt}^k + \eta_{ijt}. \quad (2.10)$$

Cheng and Wall (1999) as well as Egger and Pfaffermayr (2003) find that the omission of country-pair fixed-effects is likely to cause an estimation bias. Baltagi, Egger, and Pfaffermayr (2003) go one step further and augment the three-way model by country-year interaction terms, which they find to be preferable for panel specifications of the gravity equation.

Although the three-way approach can be expected to have the best fit, its application has a heavy disadvantage. The inclusion of country-pair dummies (or the computationally elegant within-transformation which analyzes the deviations from the variables' averages and yields the same results) eliminates all bilateral variables that are characteristic of a country pair and not varying over time. Among them are variables of very high interest like: distance, common languages or borders, etc. Direct use of the three-way approach makes it impossible to estimate the effects of these variables, because they are captured by the country-pair dummies. Egger (2005) recommends the Hausman-Taylor estimator to solve this problem. The Hausman-Taylor estimator uses information from the error term of a fixed-effects estimator to identify the effects of the time-invariant variables. The technique requires that both time-variant and time-invariant right-hand side variables can be split into two groups, exogenous and endogenous variables. In contrast to the exogenous variables, the endogenous variables are correlated with the error term.

---

<sup>19</sup> According to Egger and Pfaffermayr (2003), who refer to Christensen (1987), this specification is equal to the form

$$\ln X_{ijt} = \pi_1^i + \pi_1^j + \pi_1^{ij} + \pi_1^t + \pi_2 \ln Y_{it} + \pi_3 \ln Y_{jt} + \pi_4^k w_{ijt}^k + \eta_{ijt},$$

and is therefore controlling via “three ways” for country, country-pair and time effects.

### 2.2.3 Dealing with Zero-Trade-Flows: Non-linear Estimators

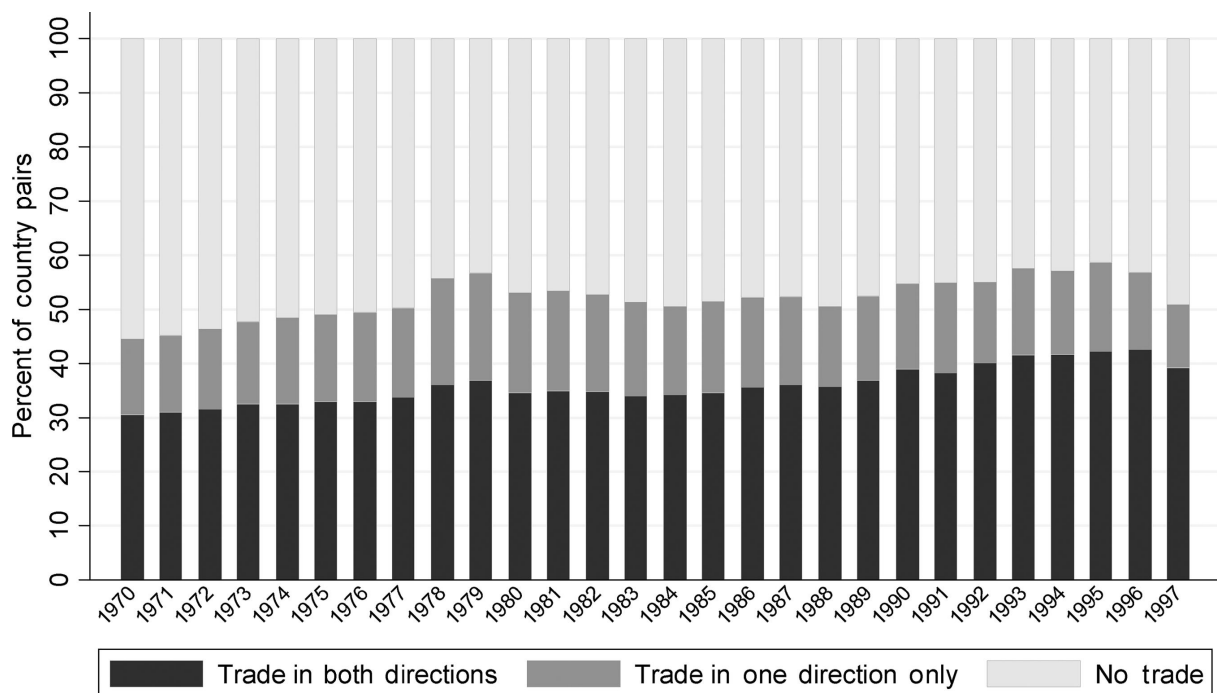
In its theoretically founded form, the gravity equation requires that trade flows occur between each pair of countries. This follows immediately from equation (1.1). As long as the economic size of one or both of the two respective countries is not zero and trade costs are not infinitely high on the right-hand side, the left-hand side can not be zero in this formulation. Yet, in reality there are pairs of countries without any bilateral trade flows. Basically, these countries are small economies in remote regions which do not trade with other small economies in other remote regions. An example could be a Central-African development country with a small insular state in the Pacific Ocean.

The extent of the phenomenon of zero trade flows is illustrated in figure 2.1. The black fractions of the columns show the percentage of country pairs with trade flows in both directions for the years 1970 to 1997 using data for 158 countries. Trade flows in both directions, where country  $i$  exports to country  $j$  and country  $j$  also to country  $i$ , are represented by the black fractions. The gray fractions represent country pairs with trade in only one direction, meaning that country  $i$  exports to country  $j$  but country  $j$  does not export to country  $i$ . The light gray fractions represent the share of country pairs without any trade. Country pairs without any trade have a share of about 50% relative to all country pairs. However, over the three displayed decades the share of country pairs trading in both directions could rise from 30 to roughly 40%.

Helpman, Melitz, and Rubinstein (2008) draw on the new-new trade theory (Melitz, 2003) to derive a theoretical model with firm heterogeneity and trade costs with a fixed component to explain the presence of zero trade flows.<sup>20</sup> Due to firm heterogeneity, it depends on the firm-specific productivities and the country-pair specific trade costs whether a firm exports into a certain other country. Moreover, firm heterogeneity implies that trade flows are not symmetrical, which means that trade flows from  $i$  to  $j$  need not be of the same magnitude as trade flows from  $j$  to  $i$ . Given the presence of firm heterogeneity,

---

<sup>20</sup>Their model is very similar to the approach by Chaney (2008), which was introduced in the previous section of this chapter. But it concentrates on finding an empirically practical way to handle the zero trade flow bias at all, while Chaney (2008) concentrates more on theoretical issues.



**Figure 2.1:** Distribution of Country Pairs Based on Direction of Trade.

*Source: Helpman, Melitz, and Rubinstein (2008), Figure I.*

they argue that export data between two countries are subject to a selection bias since e.g. high productive firms are more likely exporters than low productive firms. Therefore, the sample is not randomly selected. To correct for this selection bias, they apply the Heckman estimator. This method consists of two stages. In the first stage, they estimate the probability that country  $i$  exports to country  $j$ , using a Probit estimator. They note that this is possible without firm-level data and derive the probability employing country-level data. The predicted values are then used in the second stage to estimate a gravity equation extended by the probability that country  $i$  exports to  $j$ . They conduct several robustness checks and conclude that the Heckman method is more consistent and unbiased compared to the other estimation strategies ignoring this problem.

An earlier approach by Silva and Tenreyro (2006) is motivated by the problem that heteroscedasticity is likely to lead to biases using log-linear OLS approach, which is common practice. They apply a non-linear Poisson estimator to control for this bias and note that their strategy is furthermore able to solve the zero trade flow bias. Helpman, Melitz, and Rubinstein (2008, p. 447) explicitly emphasize that the Poisson method yields similar estimates and is consistent with their findings.

Felbermayr and Kohler (2006) use a censored data approach to tackle the problem of zero trade flows. They argue that a transformation of the left-hand side by adding 1 to the export values is necessary, so that the left-hand side becomes  $\ln(1 + X_{ijt})$ . Otherwise, zero trade flows disappear due to logarithmizing and are treated as missing values. Because OLS would yield biased results due to heteroscedasticity using this censored data, they suggest a Tobit estimator to estimate the gravity equation with zero trade flows.

Heterogeneity and zero trade flows are more likely to occur in data sets that contain a sample of rather different countries. In this study I will use a set of relatively similar OECD countries that circumvents these problems. The OECD data set used in the remainder of this study is completely free of zero trade flows.

### **2.2.4 Dealing with Endogeneity and Simultaneity: Instrumental Variables**

Considering the right-hand side of a standard gravity equation, one could argue that many of the regressors are endogenous and therefore depend on variables excluded from the model. In this case, the respective regressor(s) are correlated with the disturbance term. The results for the estimated parameters are biased and inconsistent when this problem is ignored. Furthermore, endogenous right-hand side variables could depend on the left-hand side variable. This reverse effect is known as simultaneity which also leads to biased and inconsistent estimates.

The econometric way to handle endogeneity is an instrument variable (IV) regression. This approach usually consists of two stages, and so it is known as the two stages least-square estimator (2SLS). In the first stage, the endogenous right-hand side variables are regressed on all exogenous variables. In the second stage, the estimated values of the endogenous regressors are used as instruments to run the regression of the gravity equation.<sup>21</sup>

---

<sup>21</sup>This procedure will be applied in chapter 4. Also see chapter 4 for a more detailed methodological description.



Endogeneity and simultaneity in gravity equations have been addressed by several studies. The literature on trade and growth argues that countries' GDPs depend on exports. For instance, Frankel and Romer (1999) run a gravity equation and use the estimated values to compute "constructed trade shares" for each country. These constructed trade shares are used as instruments in a further step to regress per capita income on trade shares, besides population and area. However, they cannot find evidence that controlling for endogeneity improves the results substantially.<sup>22</sup> One reason might be that GDP is a function of the difference between exports and imports rather than pure exports, and this difference is normally quite small compared to a country's GDP. Apart from that, there should not be much simultaneity between GDP and *bilateral* exports since GDP rather depends on multilateral exports. These considerations may help to explain why taking potential endogeneity of GDPs into account has not prevailed so far in the gravity literature.

Another literature branch, represented by Baier and Bergstrand (2007) amongst others, concentrates on the question of whether there is a reverse causality between bilateral trade flows and free trade agreements (FTA). This literature argues that signing a FTA is motivated by the notion that the agreement-member countries tend to have considerable trade flows among each other. Studies addressing this problem basically find evidence for simultaneity between FTAs and exports.

The literature on endogenous right-hand side variables in gravity equations does not focus on the endogeneity of overall trade costs, which is the subject of this study. In chapter 4, a recently developed index for comprehensive trade costs will be employed to analyze the simultaneity between exports and trade costs as a whole.

---

<sup>22</sup> "As a result, the hypothesis that the IV and OLS estimates are equal cannot be rejected" (Frankel and Romer, 1999, p. 388).

## 2.3 Conclusions

In the 1970s and 80s, several economists raised criticisms that the gravity equation does not rely on trade theory and is therefore purely intuitive and atheoretic (see e.g. Deardorff, 1984). The lesson from the different contributions presented so far is that versions of the gravity equation, even with its characteristic log-linear form, can be derived from all three pillars of the trade theory: classical/neo-classical, new trade theory and new-new trade theory. The gravity equation is not only consistent with theory in the sense that it can be justified as an empirical tool. Rather, the gravity equation is the collective result of different ways to model international trade theoretically.

The empirical application of the gravity equation is not without its share of pitfalls. In most cases, the simple regression of the basic specification (2.8) is inappropriate. It has been shown that fixed-effect dummies should be employed to control for unobservable country characteristics. It has also been shown that the presence of zero trade flows and heteroscedasticity, which usually appears in data sets with differing countries (e.g. North-South trade), requires alternative (non-linear) estimation strategies to yield consistent and unbiased results. Finally, it has been stated that endogeneity of regressors and simultaneity between right-hand side and left-hand side variables can be handled using IV estimators.

## Chapter 3

# A Theory of Endogenous Trade Costs

In theoretical foundations of the gravity equation, trade costs are usually assumed to be exogenously given “iceberg-melting-costs”: If a certain good is sent from one country to another, this good loses a *fixed* part of its value (Samuelson, 1954). Iceberg-costs can be interpreted as an *ad valorem* tariff equivalent to trade costs. Using them to model trade costs is very common in theoretical models, because this is quite practicable. In this chapter I shall argue that trade costs should be treated as endogenous from a microeconomic point of view. Moreover, they should depend on trade input prices and the underlying trade volume, since scale effects are likely to appear. My argument is that trade costs per dollar of trade volume are lower if there is more trade between the countries. The intuition is that trade costs should come along with a fixed cost intensive physical and social infrastructure. Thus scale effects should play a role in modeling trade costs. For example, sending a bottle of wine from one country to another should cost less in the presence of an established trade-services infrastructure – itself a result of significant overall trade between the two countries. If trade costs are determined by the trade volume, and average trade costs are falling with trade volume (e.g. due to economies of scale in trade sector), the estimated effects of right-hand side variables from gravity equations might be biased upwards, if they are interpreted as direct effects on exports.

Grossman (1998) criticizes the iceberg-approach in theoretical gravity models as a “technology for shipping tomatoes”. He raised the suspicion that a poorly-designed inclusion of trade costs into gravity frameworks could be a reason for what Obstfeld and Rogoff (2001) later posed as one of their six puzzles<sup>1</sup> of international macroeconomics: the problem that the estimated coefficients of border and distance effects on trade have unexpectedly high values.<sup>2</sup> But important theoretical contributions that help to improve the interpretation of empirical gravity equation outcomes, like those of Anderson and van Wincoop (2003),

---

<sup>1</sup> Note that the puzzle of overly high estimates of gravity coefficients is related to but not the same as the distance puzzle by Disdier and Head (2008), who find rising distance parameters over time which is contra-intuitive to the hypothesis that trade costs were falling over recent decades. See Felbermayr and Kohler (2006, section 2) or Buch, Kleinert, and Toubal (2004) for a possible solution to this kind of distance puzzle.

<sup>2</sup> Probably, the most cited example is McCallum (1995), who estimated that the border between Canada and the United States makes trade between a certain Canadian province and another Canadian province higher by a factor 22 (2,200 percent) than trade between this Canadian province and a U.S.-state of the same economic size and distance.

also use the concept of exogenous iceberg-costs to insert trade costs into their model.<sup>3</sup>

Some new studies are aware of the circumstance that per-unit trade costs might decrease with greater trade volume. Felbermayr and Kohler (2006) introduce a threshold value for trade and argue that countries will not trade if the trade volume is lower than this threshold, because trade requires physical and social infrastructures and maintaining this infrastructure is related to fixed costs. To bear these fixed costs, a minimum volume of trade must persist (see Felbermayr and Kohler, 2006, p. 657 f.). Helpman, Melitz, and Rubinstein (2008), Chaney (2008) as well as Melitz and Ottaviani (2008) model transport costs as iceberg-costs plus an additional fixed markup for shipping one unit from one country to another.<sup>4</sup> However, the introduction of fixed trade costs into these models has the purpose to derive extensive margins for firms to export their products.

My intention is to use scale effects in international trade to identify a duality of a general trade cost function and the gravity function which can lead to simultaneity-biases in empirical frameworks. Because iceberg-costs can be interpreted as exogenously given and constant average costs of trade, they are independent from the underlying exports. Yet, if economies of scale in trade occur, this assumption becomes inadequate: the higher the exports between two countries, the lower the costs should be of sending one (composite) unit of the export volume from one country to another, since economies of scale cause declining average costs. I shall derive this concept from a simple microeconomic model. The consequence of economies of scale in trade is a simultaneity problem in empirical gravity equations. This leads to a bias if the estimated parameters are interpreted as direct effects of the variables on exports. Estimates of traditional gravity equations must be interpreted as overall effects, resulting from a presumably frictionless and immediate interaction between trade costs and the gravity equation. Under certain circumstances, this bias of the direct effects can be a contribution to explain implausibly high estimates for border effects in gravity frameworks.

---

<sup>3</sup> The innovation by Anderson and van Wincoop (2003) is that trade barriers between two countries must be seen as relative to the trade barriers with all other barriers of these two countries. Their approach is the basis for this analysis.

<sup>4</sup> As a result, Chaney (2008) yields an endogeneous elasticity of the exports with respect to trade barriers.

The chapter is structured as follows. Section 3.1 introduces the theoretical derivation of the gravity equation by Anderson and van Wincoop (2003) with trade costs modeled as iceberg-costs. Section 3.2 offers an approach to model trade costs endogenously. If exports are considered as the output of a trade sector, microeconomic theory reveals that the commensurate trade costs depend on input prices and the volume of exports. The presence of economies of scale in this trade sector, which according to several empirical studies may be assumed, leads to decreasing average trade costs in exports. Section 3.3 gives an overview of theoretical arguments to justify why trade should be subject to scale effects. Since, in traditional specifications of the gravity equation, trade cost proxy variables are usually directly inserted into the gravity equation, section 3.4 introduces a more exact modeling by inserting the theory-based trade cost function into the theory-based gravity equation. Further, it shows the resulting bias. Section 3.5 summarizes the implications of this chapter.

## 3.1 A Theory Based Gravity Equation

This section introduces the well known derivation of a theory-based gravity equation developed by Anderson and van Wincoop (2003), which is the theoretical starting point of this study.<sup>5</sup> Consider a world with many countries  $\{1, \dots, C\}$ . The respective GDP of each country is exogenously given. Each country  $i$ 's total production  $Y_i$  can be seen as a specific tradeable good of this country – the so-called Armington assumption (Armington, 1969). The intuition of this assumption is that consumers – to give an example – don't care whether it is a car or an apple, but they care where the commodity has been produced.<sup>6</sup> Consumers over the world are assumed to have the same preferences. An exporting country will be denoted with  $i$ , an importing country with  $j$ .

---

<sup>5</sup> The detailed calculus is documented in appendix A.

<sup>6</sup> This assumption is used for simplicity. Anderson and van Wincoop (2004) show the same results with many goods per country. See also Deardorff (1998) for a discussion of the case of many goods and relaxing the Armington assumption.

Following Anderson and van Wincoop (2003) trade costs enter the model as iceberg-costs. Iceberg-costs are a fixed exogenously given markup (“iceberg-factor”)  $t_{ij}$  to the factory price  $p_i$ , so that the price of the (composite) commodity of country  $i$  paid in country  $j$  is  $p_{ij} = t_{ij} \cdot p_i$ . The price of the commodity from  $i$  is higher in country  $j$  by the factor  $t_{ij}$  due to trade costs. It is assumed that  $t_{ij} > 1$  for all countries  $j \neq i$  and that the domestic trade cost factor  $t_{ii} = 1$ . This is to ensure that commodities are more expensive abroad than on the domestic market. Modeling trade costs in this way leads to three properties. First, since the exports including transport costs (gross exports) are  $X_{ij} = t_{ij} \cdot p_i \cdot c_{ij}$  with quantity  $c_{ij}$  sent from  $i$  to  $j$ , the exports can be deconstructed into total trade costs  $(t_{ij} - 1) \cdot p_i \cdot c_{ij}$  plus (net) exports exclusive transport cost  $p_i \cdot c_{ij}$ .<sup>7</sup> Second, it can be shown that a fraction  $(t_{ij} - 1)/t_{ij}$  of the amount of goods shipped from  $i$  to  $j$  is lost in transport.<sup>8</sup> Finally, iceberg-costs are a measure of average trade costs and not total trade costs. This property is important for the message of this chapter. Obviously, the iceberg-factor can be denoted as gross exports divided by net exports:

$$t_{ij} = \frac{p_{ij} \cdot c_{ij}}{p_i \cdot c_{ij}}.$$

This implies that  $t_{ij}$  is nothing more than the tariff-equivalent factor for bringing \$1.00 of country  $i$ 's composite export good to country  $j$ . Therefore, iceberg-cost-factor  $t_{ij}$  is nothing more than an *average cost* of trade.

Keeping these properties of iceberg-costs in mind, we can start to build the trade model. The procedure follows the work of Anderson (1979), whose development of a very simple form without trade costs was illustrated in the previous chapter, see equation (2.3). Consider an importing country  $j$ . Recall that consumers around the world are assumed to have identical preferences, so that preferences of the consumers in country  $j$  can be

---

<sup>7</sup> To bring this mathematically into one line:  $X_{ij} = p_{ij} \cdot c_{ij} = t_{ij} \cdot p_i \cdot c_{ij} = (t_{ij} - 1) \cdot p_i \cdot c_{ij} + p_i \cdot c_{ij}$ . The last expression shows that  $X_{ij}$  equals total trade costs (first summand) plus the net exports (second summand).

<sup>8</sup> Assume for simplicity that  $p_i = 1$  and  $c_{ij} = 1$  and e.g.  $t_{ij} = 1.25$ . This means, country  $i$  must send 1.25 units to  $j$  so that one unit arrives. In this case a fraction  $0.25/1.25 = 0.2$  or 20% of the exports sent by country  $i$  would be lost.

represented by the CES utility function<sup>9</sup>

$$U_j = \left( \sum_i \varphi_i \cdot c_{ij}^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}}. \quad (3.1)$$

Here,  $c_{ij}$  is the quantity of  $i$ 's commodity imported by  $j$  (including country  $j$ 's domestic consumption  $c_{jj}$ ),  $\varphi_i$  is a distribution parameter to weight the preference of the representative consumer for country  $i$ 's composite good and  $\sigma$  is the elasticity of substitution between all goods in the world. This elasticity of substitution is assumed to be  $\sigma > 1$ , meaning that there is a substitutive relationship between individual commodities by the different countries.<sup>10</sup> The budget constraint of country  $j$  postulates that its GDP  $Y_j$  must equal the expenditure of country  $j$  on all goods of all countries  $i$  (inclusive of the good of country  $j$  itself,  $T_{jj} = p_{jj} \cdot c_{jj}$ ):

$$Y_j = \sum_i p_{ij} \cdot c_{ij} = \sum_i t_{ij} \cdot p_i \cdot c_{ij}, \quad (3.2)$$

with  $p_{ij}$  as the price of  $i$ 's commodity in country  $j$ . The factory price of  $i$ 's commodity, i.e. the price net of all trade costs, will be denoted with  $p_i$ .

Maximizing country  $j$ 's utility function subject to its budget constraint yields the demand function and multiplying both sides of this demand function by  $p_{ij}$  yields the import function<sup>11</sup>

$$X_{ij} = \varphi_i^\sigma \cdot \left( \frac{t_{ij} \cdot p_i}{P_j} \right)^{1-\sigma} \cdot Y_j, \quad (3.3)$$

---

<sup>9</sup> Recent work by Melitz and Ottaviani (2008) or Behrens, Mion, Murata, and Suedekum (2008) criticizes the usage of CES utility functions in the theoretical gravity equations and uses a more complex specification for the utility function where demand elasticity becomes endogenous.

<sup>10</sup> In a review of empirical literature, Anderson and van Wincoop (2004) point out that this value of the elasticity of substitution  $\sigma$  lies between 5 and 10.

<sup>11</sup> The individual mathematical steps to achieve this and the following results are documented in appendix A.



with  $X_{ij} = t_{ij} \cdot p_i \cdot c_{ij}$  being the gross value of imports of  $j$  from  $i$  and

$$P_j = \left( \sum_i \varphi_i^\sigma t_{ij}^{1-\sigma} p_i^{1-\sigma} \right)^{1/(1-\sigma)} \quad (3.4)$$

being a CES-price-index of country  $j$ .

Now, consider an exporting country  $i$ . In a general equilibrium with cleared markets, the GDP of country  $i$  must equal the sum of all exports (including the export into  $i$  itself –  $i$ 's intra-national trade  $T_{ii}$ ). Combining this general equilibrium condition with equation (3.3) yields:

$$\begin{aligned} Y_i &= \sum_j X_{ij} & (3.5) \\ &= \sum_j \varphi_i^\sigma \cdot \left( \frac{t_{ij} \cdot p_i}{P_j} \right)^{1-\sigma} \cdot Y_j \\ &= \varphi_i^\sigma p_i^{1-\sigma} \cdot \sum_j \left( \frac{t_{ij}}{P_j} \right)^{1-\sigma} \cdot Y_j \\ &= \varphi_i^\sigma p_i^{1-\sigma} \cdot Y_w \cdot \sum_j \left( \frac{t_{ij}}{P_j} \right)^{1-\sigma} \cdot s_j \\ &= \varphi_i^\sigma p_i^{1-\sigma} \cdot Y_w \cdot \Pi_i^{1-\sigma}, \end{aligned}$$

with  $Y_w = \sum_j Y_j$  being the world's GDP,  $s_j = Y_j/Y_w$  being country  $j$ 's share of world GDP and

$$\Pi_i \equiv \left( \sum_j \left( \frac{t_{ij}}{P_j} \right)^{1-\sigma} \cdot s_j \right)^{1/(1-\sigma)} \quad (3.6)$$

being a measure for country  $i$ 's *multilateral resistance*. This is an index for mean trade costs of country  $i$  with all countries (summed over  $j$ ), weighted by country size and elasticity of substitution.

Solving equation (3.5) for the scaled prices ( $\varphi_i^\sigma p_i^{1-\sigma}$ ) and using this for the CES-index

(3.4) yields the multilateral resistance term for country  $j$ :

$$P_j = \left( \sum_i \left( \frac{t_{ij}}{\Pi_i} \right)^{1-\sigma} \cdot s_i \right)^{1/(1-\sigma)}. \quad (3.7)$$

Substituting the solution of equation (3.5) for the scaled prices ( $\varphi_i^\sigma p_i^{1-\sigma}$ ) into the import volume function (3.3) finally gives the theory-based gravity equation

$$X_{ij} = \frac{Y_i \cdot Y_j}{Y_w} \cdot \left( \frac{t_{ij}}{\Pi_i \cdot P_j} \right)^{1-\sigma}. \quad (3.8)$$

Note, that (3.8) includes trade costs on both sides. It will be useful to consider trade flows without trade costs. The corresponding gravity equation for *net* exports follows from dividing (3.8) by  $t_{ij}$ :

$$X_{ij}^0 = \frac{Y_i \cdot Y_j}{Y_w} \cdot t_{ij}^{-\sigma} \cdot (\Pi_i \cdot P_j)^{\sigma-1}, \quad (3.9)$$

where  $X_{ij}^0$  denotes trade cost adjusted trade flows ( $X_{ij}/t_{ij}$ ) or net exports while  $X_{ij}$  denotes *gross* exports.<sup>12</sup>

As long as the elasticity of substitution between the countries' goods,  $\sigma$ , is larger than 1, higher bilateral iceberg-trade-costs lower the bilateral exports. Since factor  $t_{ij}$  can be interpreted as the cost of bringing a value of \$1.00 from country  $i$  to  $j$ , a kind of average trade cost, it follows from gravity equation (3.8) and (3.9): the higher the average trade costs between two countries, the lower the exports. Considering factor  $t_{ij}$  not as some undefined measure of trade costs but explicitly as the average trade cost value will be a central message of this chapter. A higher value for the elasticity of substitution,  $\sigma$ , increases the impact of trade costs on exports because foreign goods can be substituted more easily by the composite domestic good.

An important outcome of the Anderson/van-Wincoop-model is that these average trade

---

<sup>12</sup>If trade costs were only costs of insurance and freight,  $X_{ij}$  would be the CIF-trade-volume and  $X_{ij}^0$  the FOB-trade-volume, but in the context of this model, trade costs can play a much broader role.

costs do not simply enter the gravity equation (like in older versions), but they must be seen as relative to the product of the multilateral resistances of the trading partners: It is not enough to consider average trade costs between two countries, bilateral average trade costs *relative* to all other trading partners must enter the model. Several studies show that controlling for these multilateral resistances lowers implausibly high border effects (see Hummels, 1999; Rose and van Wincoop, 2001; Anderson and van Wincoop, 2003, and others).

### 3.2 A Micro-founded Form of Trade Costs

In the setup with iceberg-costs of the previous section,  $t_{ij}$  is a constant factor that represents average costs of trade. This factor is not directly measurable and is usually proxied by distance and several control variables (e.g. dummies for common border and common language). However, since  $t_{ij}$  denotes average costs of trade, to my knowledge it has never been modeled as a micro-founded average cost function. From microeconomic theory, it is well known that an average cost function not only depends on cost factors like factor prices but also on the quantity produced. Therefore, I argue that average trade costs are dependent on export values.

Assume that between each pair of countries there is a trade sector or a representative firm that carries out all services to bring goods from the factories in country  $i$  to the consumers in country  $j$ . To keep the model general, we will not assume scale effects or fixed costs on these services. If the trade sector is able to make profit, it does not account for any country's GDP: In the gravity model, GDP is implicitly defined as a country's output of (composite) tradable goods while shipping is assumed to be a non-tradable service. If bilateral net exports are the output of this trade sector's production function, we can denote it as a function of input vector  $x_{ij}^k = (x_{ij}^1, \dots, x_{ij}^K)$ :

$$X_{ij}^0 = X_{ij}^0(x_{ij}^k). \tag{3.10}$$

An input  $x_{ij}^k$  in this context can mean, for example, shipping one good via ocean or air, paying for tariffs, translating contracts, and so on. Now, let  $w_{ij}^k = (w_{ij}^1, \dots, w_{ij}^K)$  be the vector of input prices. Minimizing trade costs  $\sum_k w_{ij}^k \cdot x_{ij}^k$  subject to given net exports, provided that second order conditions hold, yields the trade cost function

$$TC_{ij} = TC_{ij}(w_{ij}^k, X_{ij}^0). \quad (3.11)$$

Dividing both sides by  $X_{ij}^0$  yields average trade costs,

$$\frac{TC_{ij}}{X_{ij}^0} = \overline{TC}_{ij}(w_{ij}^k, X_{ij}^0) = t_{ij}(w_{ij}^k, X_{ij}^0) - 1 = \tau_{ij}(w_{ij}^k, X_{ij}^0). \quad (3.12)$$

These average trade costs  $\overline{TC}_{ij}$  describe the costs of bringing a value of \$1.00 from country  $i$ 's composite exports to country  $j$ . Keeping the properties of iceberg-trade-costs in mind, this is equal to the interpretation of the trade cost markup  $\tau_{ij}$ . Thus, trade costs or the iceberg-factor, respectively, become endogenous. As long as there are economies of scale in the trading sector, e.g. due to the presence of fixed costs of infrastructure, the average cost function (3.12) will decline with rising bilateral exports: The more two countries trade with each other, the lower the average bilateral trade costs are. The result is the assumption that  $\partial\tau_{ij}/\partial X_{ij}^0 < 0$ , or  $\partial t_{ij}/\partial X_{ij}^0 < 0$ .

### 3.3 Economies of Scale in International Commodity Trade

Why should there be economies of scale in the trade sector? In the context of this study, trade costs are all costs for providing a foreign market with the products from the domestic market. This is the international economics interpretation of trade costs.<sup>13</sup>

---

<sup>13</sup>The spatial economics interpretation concentrates more on the geographical distance and less on the incidence of a border. Thus it also applies for domestic trade. In international economics trade costs are more readily outlined as discrete values varying over country pairs, in spatial economics they are alternately modeled continually.

Following Anderson and van Wincoop (2004) they can be subdivided into three different kinds of trade costs: (a) transport costs, (b) border-related trade barriers and (c) costs for retail and wholesale on the foreign market.

Transport costs are the costs of shipping goods. They can be separated into direct transport costs, the so-called costs of insurance and freight (CIF), and indirect transport costs, which include holding costs for goods in transit, inventory costs due to buffering the variability of delivery dates, preparation costs associated with shipment size and other costs. Hummels (2007) argues that the most important technologies for transporting goods between countries are ocean and air shipping. As one reason for this, he points out that only one quarter of the world's exports takes place between countries that share a common border. There are several approaches for capturing trade costs with empirical data (see Hummels, 1999; Limao and Venables, 2001; Redding and Venables, 2002; Hummels, 2007, for example), although indirect transport costs are particularly difficult to observe. In gravity equations, transport costs are usually proxied by the distance between the capitals or economic centers of two trading countries.

Border-related trade barriers are trade impediments which occur between countries due to political, currency, language, cultural and other reasons. The problem with these barriers is that there are many unobservable and probably even unknown effects. Some barriers are observable, e.g. tariffs, currency volatilities and so on. However, there are data limitations to the political barriers, as Anderson and van Wincoop (2004, section 2.1.1) criticize. Notably, it is very service-intensive to overcome barriers like language, mentality, bureaucracy and so on. In gravity equations, border-related trade effects are usually controlled by a set of dummy variables for common properties of the countries.

Costs for wholesale and retail have to be borne by suppliers both foreign and domestic. Since these costs differ between countries, they are likely to enter the exporter's decisions. Wholesale and retail costs are captured in gravity equations via price indices (following Baier and Bergstrand, 2001, and the earlier work by Bergstrand) or, more commonly, multilateral resistance terms which are usually controlled by country fixed-effects (Anderson

and van Wincoop, 2003).

In summary, per-unit costs of bringing goods from one country,  $t_{ij}$ , into another country should depend on (a) transport and (b) border effect cost, while (c) costs for wholesale and retail should be captured by individual country effects ( $\Pi_i, P_j$ ).

The transport sector typically uses fixed cost intensive infrastructures: harbors, airports, rail networks, road systems. Limao and Venables (2001) find that infrastructure plays an important role for the determination of transport costs, especially for landlocked countries. As market power indicates a presence of fixed costs and economies of scale. For example, Hummels, Lugovskyy, and Skiba (2009) find evidence for market power and price discrimination in the ocean shipping industry. Furthermore, work by Hummels (and co-authors) indicates that the usage of fixed iceberg-melting-costs is an inappropriate measure for transport costs. Hummels and Skiba (2002, pp. 2–6) give a detailed discussion of the sources of scale effects in the transport sector. As an introductory example, they argue that shipping goods from Ivory Coast to the U.S. East Coast is twice as costly as shipping goods from Japan to the U.S. West Coast, although distance is the same in both cases.

Costs for border-related effects are likely to have economies of scale as well. As noted above, overcoming border-related effects can be closely related to services. Here, social networks, communication networks and many more factors play an important role (see Jones and Kierzkowski, 1990, for a discussion of the particular case in which traded goods are produced in a fragmented industry) and there should be a relationship between costs for these service networks and the underlying exports, similar to a technology with fixed costs. If there is more trade, per-unit costs for translations, filling-out forms, overcoming bureaucracy, and others should be lower.

These arguments lead to the proposition that average trade costs should depend on bilateral exports and that the relationship between them is inverse.

### 3.4 Interaction between the Gravity Equation and Trade Cost Function

If bilateral trade costs depend on the underlying exports, an endogeneity problem may bias estimations from gravity equations. After inserting the endogenous average trade costs into the gravity equation, we can extract a functional term that reveals the bias and discuss it.

The first step is to bring endogenous trade costs into a functional form that is suitable to empirical studies using the gravity equation. In a critique of modeling trade costs as iceberg costs, Grossman (1998) suggests a log-linear form to concretize the trade cost function. Following this suggestion, a logarithmic form is applied to the average trade cost function (3.12) and, according to the outlined hypotheses, augmented by the exports  $X_{ij}^0$ :

$$t_{ij}(w_{ij}^k, X_{ij}^0) = e^{\beta_0} (w_{ij}^k)^{\beta_k} (X_{ij}^0)^{\beta_X}. \quad (3.13)$$

If there are economies of scale in the trade sector, the elasticity of average trade costs with respect to net exports,  $\beta_X$ , is expected to be lower than 0. The empirical question of these scale effects will be checked in chapter 4 of this study. Logarithmizing equations (3.9) and (3.13) yields:

$$\ln X_{ij}^0 = K + FE_{ij} + \ln Y_i + \ln Y_j - \sigma \ln t_{ij}, \quad (3.14)$$

$$\ln t_{ij} = \beta_0 + \beta_k \ln w_{ij}^k + \beta_X \ln X_{ij}^0, \quad (3.15)$$

with a constant  $K = \left[ \ln \frac{1}{Y_w} \right]$  and the fixed-effects  $FE_{ij} = (\sigma - 1) \ln (\Pi_i \cdot P_j)$ .

The second step is to insert the trade cost function into the gravity equation. Substituting equation (3.15) into equation (3.14) yields:

$$\ln X_{ij}^0 = K + FE_{ij} + \ln Y_i + \ln Y_j - \sigma (\beta_0 + \beta_k \ln w_{ij}^k + \beta_X \ln X_{ij}^0). \quad (3.16)$$

In equation (3.16), net exports  $X_{ij}^0$  appear on both sides and since it has got an impact

on trade costs ( $\beta_X \neq 0$ ) it should cause a bias.

The third step is to extract the bias term. Equation (3.16) can easily be solved for  $\ln X_{ij}^0$ :

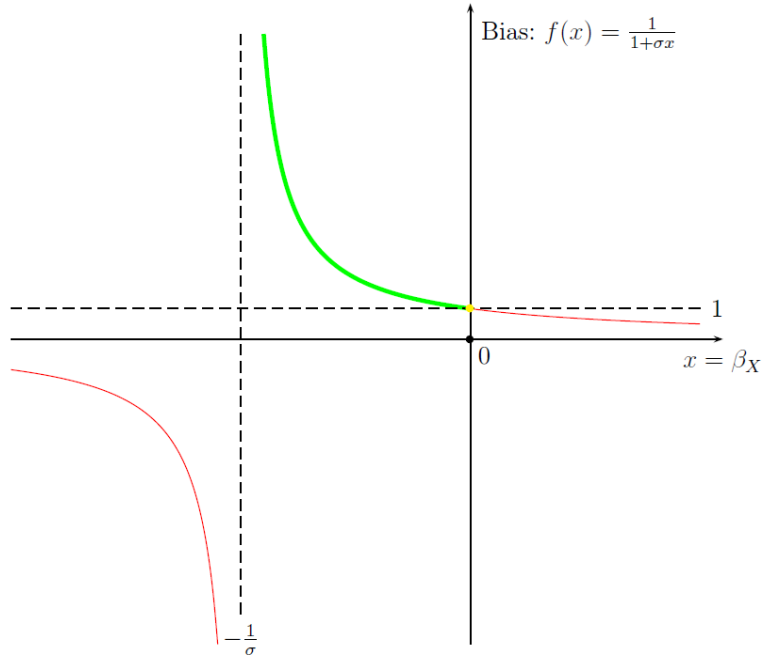
$$\ln X_{ij}^0 = \frac{1}{1 + \sigma\beta_X} (K' + FE_{ij} + \ln Y_i + \ln Y_j - \sigma\beta_k \ln w_{ij}^k), \quad (3.17)$$

with  $K' = K - \sigma\beta_0$ . The bias of ignoring endogeneity of trade costs is thus given by the fraction  $1/(1 + \sigma\beta_X)$ .

As noted before, the elasticity of substitution  $\sigma$  is assumed to be larger than 1, based on empirical evidence (Anderson and van Wincoop, 2004). If there are economies of scale in the trade sector and per-unit trade costs decrease for exports,  $\beta_X$  should be negative. Thus, the product  $\sigma\beta_X$  is expected to be negative. If  $\sigma\beta_X$  lies between 0 and  $-1$ , the bias is positive and larger than one. This would imply that coefficients are overestimated as long as trade costs are not considered to be endogenous. With  $\sigma\beta_X$  converging against  $-1$ , the bias grows exponentially toward infinity. At  $\sigma\beta_X = -1$  there is no solution for the bias. If  $\sigma\beta_X$  is smaller than  $-1$ , the fraction becomes negative. This would imply that the signs of the estimated effects are changed – which would lead to implausible estimates and that would be contradictory to the success of the gravity equation.

According to Anderson and van Wincoop (2004), most studies of the substitutability of internationally traded goods estimate substitution elasticities  $\sigma$  between 5 and 10. If  $\beta_X$  is exactly  $-1/5$  or  $-1/10$ , respectively, the bias would be indefinite. As long as the value of  $\beta_X$  is smaller than these values, the bias would reverse the parameters' signs and the gravity equation would not be as famous as it is. Given that  $\beta_X$  is higher than these values, but smaller than 0, estimated parameters are biased upwards. Insofar as  $\beta_X = 0$ , which has been implicitly assumed in gravity works until now, the fraction would be one, implying that there is no bias. Provided that  $\beta_X > 0$ , exports would have an additive effect on average trade costs which is hard to explain in a sector that is likely to deal with fixed costs. In this case, standard estimations with the gravity equation would be underestimated and the discussion about the border puzzle would go the wrong direction. The bias term as a function of  $\beta_X$  with a fixed  $\sigma$  is plotted in figure 3.1.





**Figure 3.1:** Plot of the Bias Term with a Given Value for the Elasticity of Substitution  $\sigma$ .

If exports have an impact on per-unit trade costs and if the discussion about the puzzle of the implausibly high estimates for trade barriers in gravity equations is on the right track,  $\beta_X$  must lie between 0 and the inverse value of  $-\sigma$ . This indicates that average trade costs' elasticity with respect to exports should be low, but not zero.

It is important to recognize that this bias only appears if the effects of the right-hand side variables are interpreted as *direct* effects on trade. A change in trade cost factors, e.g. a tariff reduction between two countries, lowers trade costs which leads to more trade between the two respective countries. This is the direct effect of the cost reduction on trade. But theoretically, this direct effect would lead to a domino effect: The increased export value lowers trade costs due to scale effects, the lower trade costs again increase trade and so on. If we would take the bias-affected gravity function (3.17) and insert it back into the trade cost function, the newly achieved trade cost function back into the gravity function, and so on, we would converge to the original gravity function (3.14). Thus, at the end, the results from traditional estimations must be interpreted as *overall*

effects, assuming that the domino effect of trade costs and exports works completely free of obstructions or time delays.

### 3.5 Conclusion

The theoretical literature about the gravity equation takes trade costs between two countries as exogenously given: trade costs affect the volume of bilateral trade. In all these models, trade costs enter in terms of iceberg costs which can be interpreted as trade costs per unit of exports. In this chapter, an alternative form of bringing trade costs into a theory-based gravity equation is presented. The main argument is that, from a microeconomic point of view, trade costs should not be independent of the underlying exports. Further, if we presume economies of scale in the trade sector, average trade costs should decline with the underlying exports. If this relationship is not controlled in empirical studies using the gravity equation, the estimated effects must be interpreted as overall effects and might be biased if they are interpreted as direct effects on exports. The analysis of the bias suggests that the impact of exports on average trade costs should be very inelastic, otherwise the results of gravity studies would be hard to explain. Yet, if this impact is significant, the bias might explain overestimations, which could contribute to the discussion of the “border-puzzle”.

Another interesting outcome of this model is that input prices of trade factors (denoted as  $w_{ij}^k$ ) play an important role in a micro-founded trade cost function. Usually, trade barriers are proxied by distance and some dummy-variables. However, micro-founded cost theory postulates that proxies for input prices should enter the trade cost terms to reflect the aggregated technology of the trade sector. Notably, Brun, Carrère, Guillaumont, and de Melo (2005) achieve a higher explanatory power with additional price variables like an oil price index which controls for such input prices.

Of course, it remains an empirical question whether the propositions of these theoretical considerations hold. Recent work by Novy (2007) makes it possible to compute a theory-

based index of bilateral trade costs which is equivalent to the geometric mean of two countries' iceberg-factors,  $t_{ij}$  and  $t_{ji}$ . Jacks, Meissner, and Novy (2008) use this approach to regress (average) trade costs on the usual variables, but they still do not control for exports. In the following chapter, I use this index for bilateral trade costs to test if there is a simultaneous relationship between the gravity equation and the trade cost function.

# Chapter 4

## Estimation

A basic assumption of the gravity equation for international trade is that increasing trade costs lower exports. Yet, intuition and the theoretical considerations presented in the previous chapter imply that a high export volume lowers bilateral trade costs as well: A fixed cost intensive trade sector probably bears lower average costs with more trade. In this case, standard gravity estimation might be biased due to simultaneity if they are interpreted as direct effects. The empirical analysis pursued in this chapter finds an empirical interdependency between exports and trade costs. Using a simultaneous equation model to address this problem improves the estimates compared to the standard gravity specification.

The theory introduced in the previous chapter argues that trade costs do not only determine trade flows, but trade flows also determine trade costs if there are economies of scale (falling transport costs per dollar of export) in the trade sector, since the trade sector is likely to be fixed cost intensive. This was shown by combining the gravity derivation from a general equilibrium model (Anderson and van Wincoop, 2003) with a simple model of a bilateral trade cost sector that minimizes trade costs with a given volume of export to derive an endogenous tariff equivalent for trade cost. The consequence is a mutual causality between trade flows and trade costs which might bias gravity results and thus has to be checked. This chapter suggests a way to do so. Two empirical questions will be proposed:

1. Are there economies of scale in the trade sector? This should be the case if trade costs decline with increasing exports between two countries.
2. How are estimates biased if economies of scale in the trade sector persist, but go unaddressed? This question will be analyzed by comparing the results of models with and without the mutual interaction between exports and trade costs.

The chapter is structured as follows. Section 4.1 introduces the econometric estimation strategy: the estimation of a simultaneous equation model consisting of a gravity equation and a trade cost equation, where exports and a recently developed index for trade costs (Novy, 2007) are the endogenous variables. Section 4.2 describes the data used in the

estimation, section 4.3 presents the results of estimating the simultaneous equation model and provides a comparison to the standard estimation strategies. Section 4.4 concludes.

## 4.1 Econometric Model

The traditional strategy to estimate a gravity equation is: take trade as the endogenous variable and regress it on country sizes and a set of trade cost proxies. Yet our theoretical considerations suggest, firstly, that these trade cost proxies affect trade costs rather than exports, and, secondly, that trade costs could be affected by the exports inversely, due to economies of scale in international trade. If this inverse causality between exports and trade costs exists, estimating a simultaneous equation model (SEM) should be the appropriate strategy.<sup>1</sup>

Consider equations (3.15) and (3.14). Replace the theoretical coefficients by empirical parameters, and augment the equations by the residual terms  $u_{ij}$  and  $v_{ij}$  to get the structural equations:<sup>2</sup>

$$\ln X_{ij}^0 = \alpha_0 + \alpha_{Y_i} \ln Y_i + \alpha_{Y_j} \ln Y_j + \alpha_t \ln t_{ij} + u_{ij}, \quad (4.1)$$

$$\ln t_{ij} = \beta_0 + \beta_k \ln w_{ij}^k + \beta_X \ln X_{ij}^0 + v_{ij}, \quad (4.2)$$

Since equation (4.1) depends on  $t_{ij}$  and equation (4.2) depends on  $X_{ij}^0$ , the gravity equation becomes a system of interdependent or simultaneous equations. The adequate estimator is the two-stage-least-squares (2SLS) or three-stage-least-squares (3SLS) estimator. 2SLS means that, in the first step, all endogenous variables of the equation system ( $t_{ij}$  and  $X_{ij}$ )

---

<sup>1</sup> For an conceptual overview see e.g. Greene (2000, chapter 16).

<sup>2</sup> The subscripts for the time dimension are omitted here for simplicity.

are regressed on all exogenous variables of the equation system ( $Y_i$ ,  $Y_j$  and  $w_{ij}^k$ ):

$$\ln X_{ij}^0 = \pi_1 + \pi_2 \ln Y_i + \pi_3 \ln Y_j + \pi_4^k w_{ij}^k + \eta_{ij}, \quad (4.3)$$

$$\ln t_{ij} = \pi_5 + \pi_6 \ln Y_i + \pi_7 \ln Y_j + \pi_8^k w_{ij}^k + \epsilon_{ij}, \quad (4.4)$$

with parameters  $\pi_{(\cdot)}$  and residual terms  $\eta_{ij}$  and  $\epsilon_{ij}$ . These two equations are called *reduced form equations*. Note that the reduced form equation (4.3) is identical with the traditional specification of the gravity equation. In the second step, the estimated values for the endogenous variables ( $\hat{t}_{ij}$  and  $\hat{X}_{ij}$ ) are used as instruments to estimate the initial structural equations (4.1) and (4.2).

This procedure is necessary because theory implies that both structural equations, (4.1) and (4.2), contain endogenous variables. From an econometric point of view, endogeneity of variables signifies that these variables are correlated with the error terms  $u_{ij}$  and  $v_{ij}$ . The consequence is inconsistent estimates of the parameters. Since there is a correlation between the error terms  $u_{ij}$  and  $v_{ij}$ , using a “feasible generalized least-squares” (FGLS) estimator (where the estimators are weighted by the variance-covariance-matrix) helps to improve the results. This procedure is known as 3SLS.

Since a panel data set will be used, certain techniques must be used to achieve consistent results.<sup>3</sup> As a baseline case, a pooled regression model is estimated where the panel structure of the data is not considered. Anderson and van Wincoop (2003) suggest to control for the countries’ multilateral resistances. To do so, I shall estimate a least-square dummy variable (LSDV) model with dummy variables controlling for exporting and importing countries as well as for the respective year (country-year or two-way fixed effects). This strategy was first suggested by Mátyás (1997). Most recent studies use the three-way fixed effects model (or country-pair-year fixed effects model), which controls for country-pairs and time. This specification appears in most studies as the most appropriate one.

---

<sup>3</sup> See the explanations of section 2.2 in this study. See also e.g. Cheng and Wall (1999) or Baltagi, Egger, and Pfaffermayr (2003) for an overview of panel data estimation strategies for gravity equations.

There are two central questions from an analytical point of view:

**Question 1** *Are there economies of scale in the trade sector?*

This should be the case if, in a SEM specification, the estimated value of  $\beta_X$  in equation (4.2) is significantly lower than 0.

**Question 2** *How are estimates biased if economies of scale in the trade sector persist, but go unaddressed?*

To analyze this question, I will estimate a restricted version of the equation system (4.1) and (4.2) using an instrumental variable (IV) regression where  $\alpha_t = 0$  for both the gravity and the trade cost equation. If the theoretical suggestions regarding the bias term are correct, the estimates for the parameters of the exogenous trade variables  $w_{ij}^k$  should be systematically higher than in the SEM specification.

## 4.2 Data

The data set comprises annual data for all 30 OECD countries<sup>4</sup> for the years 1995 to 2006. These countries are the largest economies in the world. They account for roughly 80% of global GDP. Many studies use broader databases like the IMF's directions of trade statistics, which provide over 50 years of data for more than 100 countries. In these data sets, zero trade flows sometimes occur, meaning that very small and remote countries might have no trade relations with any other (see Helpman, Melitz, and Rubinstein, 2008).<sup>5</sup> The presence of zero trade flows requires alternative non-linear estimation strategies. An advantage of the OECD data set, apart from the sheer density of data, is that the proposed estimation of a linear simultaneous equation model remains consistent because there are no cases of zero trade flows.

The data for bilateral exports is taken from the OECD Database for Structural Analysis

---

<sup>4</sup> Chile as the 31st member is not considered since it joins the OECD not until 2010.

<sup>5</sup> This is also discussed in section 2.2 in this study.



2008 (OECD STAN) and converted into logs ( $\text{exp}_{ij}$ ). GDP data is taken from the OECD national account statistics and converted into logs ( $\text{gdp}_i$  and  $\text{gdp}_j$ ). The  $w_{ij}^k$ -variables are performed as follows. Distance is calculated using the great-circle formula between the capitals or economic centers<sup>6</sup> of two countries in kilometers, converted into logs. As it is a tradition in gravity equations, a set of variables is used to control for country characteristics. These country characteristic variables are distinguished by exporting countries (suffix  $i$ ) and importing countries (suffix  $j$ ). For the measure of common language ( $\text{lang}$ ), common border ( $\text{bor}$ ), commonwealth of nations ( $\text{cwn}$ ), former east block ( $\text{ebl}$ ), island ( $\text{isl}$ ), landlocked countries ( $\text{landl}$ ) and EU-membership ( $\text{eu}$ ), dummy variables are used. These dummy variables take the value 1 if the condition that is controlled by the dummy applies, and 0 if not. Note that all of these dummy variables, as well as the distance variable, are constant over time except for the variable for EU-membership (since several countries became EU-members during the period). The variable  $\text{trf}$  records the log of the “Freedom of Trade Index” published by the Heritage Foundation. Exchange rate volatility is calculated as the monthly standard deviation from the annual mean relative to the annual mean for each bilateral exchange rate and converted into logs. The monthly data for the US Dollar exchange rates of the respective countries were taken from the OECD Financial Indicators database and recalculated into bilateral exchange rates.<sup>7</sup>

So far, the data set for a standard gravity framework is explained:  $\text{exp}_{ij}$ ,  $\text{gdp}_i$ ,  $\text{gdp}_j$  and a set of trade cost proxies. Since these trade cost proxies influence exports indirectly via trade costs, a measure of overall average trade costs is needed. Novy (2007) derives an index for the geometric mean of the overall trade costs (measured in iceberg costs) between two countries from the theoretical gravity equation derived by Anderson and van Wincoop (2003), which is also a starting point for the theoretical considerations of this

---

<sup>6</sup> In the case of: Canada (Toronto), Germany (Frankfurt), Turkey (Istanbul) and United States (Chicago).

<sup>7</sup> Because the Euro-countries are taken as one in this database, exchange rate changes between Euro-countries before the introduction of the Euro were calculated from historical data taken from EUROSTAT. Because this EUROSTAT database does not cover Greek Drachma and the Slovak Korun (since these countries introduced the Euro later) monthly data for the Greek Drachma exchange rate was taken from the US Federal Bank, for the Slovak Korun from the Slovak National Bank.

study.<sup>8</sup> This trade cost index can be computed by the formula:

$$t_{ij} = \left( \frac{X_{ii}X_{jj}}{X_{ij}X_{ji}} \right)^{\frac{1}{2(\sigma-1)}}. \quad (4.5)$$

The higher the exports *inside* the respective countries relative to the exports *between* the two countries, the higher the bilateral trade costs will be, and vice versa. The exports within a country (intra-national trade), can be interpreted as the country's production minus the sum of the exports into all countries. Since export data are measured in gross shipments while GDP data are based on value added (and services that are not considered in the export data) GDP is not suitable to calculate this index. Instead, following Wei (1996) and Novy (2007), production data for goods extracted from the OECD STAN Database are used and converted into US Dollars using the OECD Financial Indicators annual exchange rates.

Unfortunately, production data are not available for 6 of the 30 countries.<sup>9</sup> Therefore, missing values for production were constructed using the following three steps.

In the first step, I assume that in countries with higher productivity (measured by per-capita-income, source: World Development Indicators, WDI, 2008) the ratio between value added and production is higher. Thus, I calculate the elasticity of the value added/production-ratio with respect to per-capita-income using ordinary least-squares. In the second step, I compute the missing data points from the estimated values of this regression (if there are no data for production, but data for value added in the OECD data).

There are still some missing data points because there are no value added data for Mexico, Turkey, UK or USA available in the OECD database. Hence, in the third step, I take the value added data from the WDI database and, using an adjusted regression (intercept =

---

<sup>8</sup> The derivation of the index is displayed in appendix B. Novy (2007) shows in the latest version of his paper (November 2009) that this index can also be derived from a number of other theory-based gravity equations. Also note that already Head and Ries (2001) and Head and Mayer (2004) have derived versions of this index using a Dixit-Stiglitz-Krugman model of international trade. This model corresponds to the gravity equation of Anderson and van Wincoop (2003), used by Novy (2007), as it was shown in chapter 2.

<sup>9</sup> These countries are: Australia, Ireland, Mexico, Turkey, UK and US.

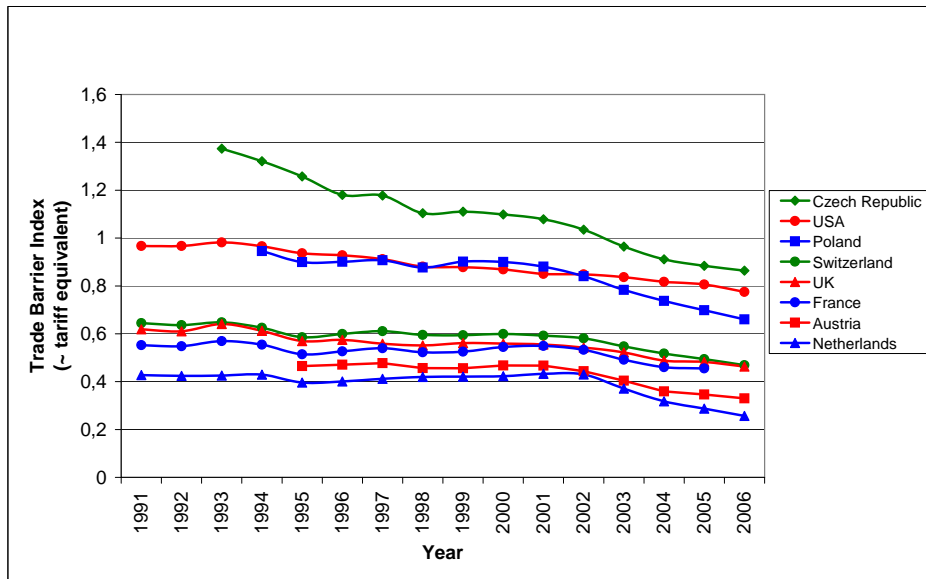
0) between OECD and WDI data, I find that OECD data systematically is 95% of the WDI values. Consequently, I multiply WDI data for the value added by factor 0.95 and pursue the same procedure as in the first and the second step to compute missing production estimates for cases there are no value added data available in the OECD STAN database, but in the WDI database. Missing values for the countries' total exports are also supplemented from WDI data, where 0.95 turns out to be the adequate adjustment factor as well. The differences between the resulting total production values and the total export values reflect the remaining goods produced (and therefore traded) inside the respective country.

Another crucial issue is the elasticity of substitution between the countries' composite goods,  $\sigma$ . In a survey of the empirical literature, Anderson and van Wincoop (2004) find that this elasticity takes values between 5 and 10. Thus, following Novy (2007), the elasticity of substitution is set  $\sigma = 8$ .<sup>10</sup> Figure 4.1 illustrates the trade costs between Germany and 8 of her trading partners. The index is measured as a tariff equivalent,  $\tau_{ij} = t_{ij} - 1$ . For example, trade costs between Germany and the USA declined from 100% in the early 1990s to 80% in 2006. 100% tariff equivalent means in this context that the cost for transportation, overcoming national borders and retail/wholesale in the target market equals 100% of the value of the exported goods. Netherland and Austria have, compared with other EU-members, relatively low trade costs with Germany. Switzerland as a non-EU-member, which has a similar economic size and geographic and cultural distance (or connectivity) to Germany like Austria, has comparatively high trade costs with Germany. The trade costs with the eastern European countries Czech Republic and Poland are nearly as high as those with the USA, although they declined faster over the period. Overall, Germany's trade costs became lower over the considered period.

Table 4.1 summarizes the data used for the estimation. Notice, that the log of the trade cost index is negative between Belgium and the Netherlands for the three years from 2004 to 2006 (six observations). This is a counterintuitive number because a negative value

---

<sup>10</sup> A sensitivity analysis using  $\sigma = 5$  and  $\sigma = 10$  leads to exactly the same results for the estimated parameters of the exogenous variables.



**Figure 4.1:** Trade cost index for Germany and 8 certain countries measured as tariff equivalent.

implies that the iceberg factor  $t_{ij}$  would be lower than 1 and thus must be interpreted as a negative trade cost markup on the export value. The reason for this phenomenon is probably that the markets of Belgium and the Netherlands are integrated on an extremely high level. Furthermore, many goods from overseas arrive at the Dutch harbor of Rotterdam – the most important harbor of this region – and are then sent to Belgium which systematically increases the exports compared to the respective national production values and thus affects the trade cost index. The six observations with negative values for the log of the trade cost index are excluded from the estimation.

### 4.3 Results

Table 4.2 shows the estimates of the traditional gravity estimation strategy, where all determinants of trade costs appear directly in the estimation equation. In column 1, the results of the pooled regression approach are shown. Only two regressors, namely

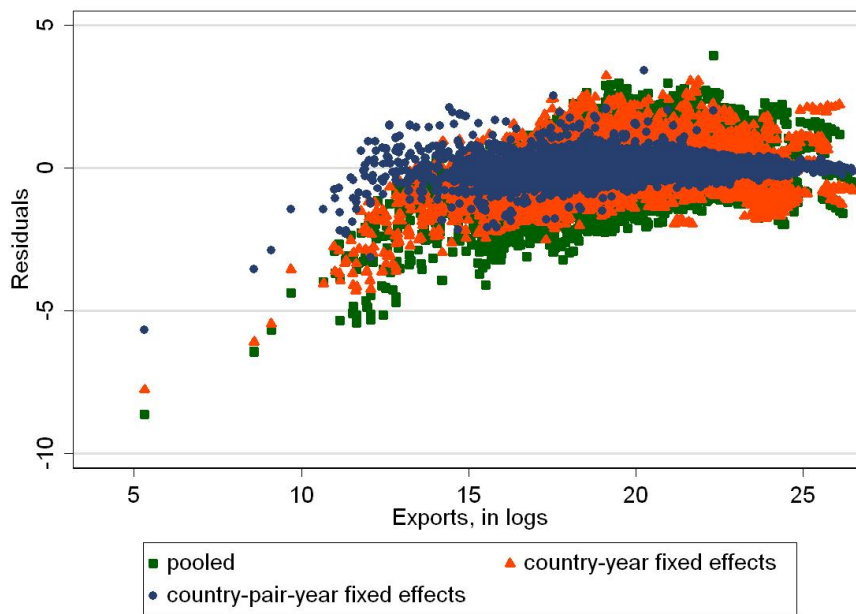
**Table 4.1:** Summary statistics of the OECD data set, over 12 years.

Variable	Mean	Std. Dev.	Min.	Max.	Obs.
expij	20.147	2.344	5.323	26.481	9878
tij	0.810	0.256	-0.156	1.893	9418
gdpi/gdpj	26.419	1.532	22.672	30.204	10226
trfi/trfj	4.351	0.069	3.904	4.443	9972
dist	7.962	1.189	4.043	9.895	10226
exvol	0.938	1.144	-12.281	3.537	9428
lang	0.067	0.25	0	1	10226
bor	0.076	0.265	0	1	10226
cwni/cwnj	0.133	0.34	0	1	10226
ebli/eblj	0.133	0.34	0	1	10226
isli/islj	0.2	0.4	0	1	10226
landli/landlj	0.2	0.4	0	1	10226
eui/euj	0.544	0.498	0	1	10226

those of the island location of the importing country (islj) and the EU-membership of the importing country (euj), do not have a significant impact on bilateral exports. Note that the freedom of trade index for both, the exporting and the importing countries (trfi and trfj), has a significantly negative impact on bilateral trade. That means that more liberal importers have lower imports. This result is counter-intuitive.

Column 2 represents the results of estimating the LSDV model with country-year fixed effects. In this model, the two freedom of trade variables (trfi and trfj,) the importing country's commonwealth-of-nations-membership (cwnj,) and the importing country's landlocked location (landlj), are not significant. Note that exchange rate volatility seems to have a significantly positive effect on trade in this specification: More uncertainty about exchange rates enhances trade. Also note that the signs of some dummies change compared to the pooled regression specification.

Column 3 of table 4.2 shows the results for the three-way fixed effects estimator with country-pair-year fixed effects. Here, all time-invariant variables are dropped due to collinearity. All estimated parameters have got the expected signs: higher trade-freedom (at least in the exporting country) and membership in the European Union have positive impacts on trade. Note that the coefficients of the importing country's freedom of trade and of the exchange rate volatility are not significant at the 10%-level. A neutral exchange



**Figure 4.2:** Residual-Analysis of the Standard Gravity Specification: Pooled Regression, Country-Year Fixed Effects and Country-pair-Year Fixed Effects.

rate risk is not surprising. First, firms have possibilities to hedge this risk on financial markets. Second, exchange rate risks are likely to play a higher role in the trade with less-developed countries rather than those of the OECD.<sup>11</sup>

To compare the three kinds of specifications, the residuals are plotted in figure 4.2. While the residuals of the pooled and country-year fixed effects specifications increase with the logarithmized export volume, in the country-pair-year fixed effect model they are distributed around zero independently from the export volume. This observation indicates that the country-pair-year fixed effects model should be preferred. Furthermore the standard errors of the estimated parameters are lower and the results are more intuitive overall. Therefore, the three-way fixed effects estimator should be preferred over the other two estimators.<sup>12</sup>

<sup>11</sup>In an earlier version of this study, where the dummy variables were distinguished by “one country or both countries” instead of “exporting country or importing country”, the coefficients of all variables were significant and had the expected sign.

<sup>12</sup>It was also tested whether a random effects model is adequate. The Hausman-test rejects the null-hypothesis that there are no systematic differences in the parameters of three-way fixed effects and random effects, which implies that the three-way fixed effects estimator has to be preferred. In the remaining analysis, the random effects estimator is not further discussed.

## 4. Estimation

**Table 4.2:** Basic Case: Results of the Standard Gravity Specification.

	Pooled Regression (1)	Country-year FE (2)	Country-pair-year FE (3)
gdpi	0.942 (0.009)***	0.308 (0.089)***	0.408 (0.049)***
gdpij	0.858 (0.009)***	0.737 (0.085)***	0.808 (0.040)***
trfi	-0.396 (0.197)**	0.268 (0.199)	0.252 (0.113)**
trfj	-0.315 (0.178)*	-0.023 (0.180)	-0.063 (0.086)
dist	-1.034 (0.014)***	-1.200 (0.018)***	—
lang	0.655 (0.045)***	0.604 (0.048)***	—
bor	0.579 (0.044)***	0.417 (0.046)***	—
cwni	-0.098 (0.036)***	2.045 (0.180)***	—
cwnj	0.370 (0.037)***	-0.154 (0.130)	—
ebli	-0.113 (0.038)***	0.683 (0.074)***	—
eblj	-0.075 (0.043)*	-0.867 (0.247)***	—
isli	0.312 (0.031)***	3.693 (0.316)***	—
islj	0.025 (0.034)	-0.175 (0.096)*	—
landli	0.057 (0.030)*	-0.946 (0.187)***	—
landlj	-0.425 (0.034)***	-0.138 (0.103)	—
eui	0.221 (0.024)***	0.639 (0.052)***	0.603 (0.027)***
euj	0.012 (0.025)	0.182 (0.056)***	0.181 (0.028)***
exvol	-0.042 (0.010)***	0.029 (0.009)***	-0.005 (0.005)
Constant	-16.768 (1.091)***	-0.734 (3.365)	-13.382 (1.697)***
Observations	8492	8492	8492
Adjusted $R^2$	0.83	0.90	0.98

Note: Robust standard errors in parentheses.

\* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%.

## 4. Estimation

**Table 4.3:** Theory-based estimates without economies of scale in trade (IV) and with economies of scale in trade (SEM).

	Instrumental Variable Estimator (IV)			Simultaneous Equation Model (SEM)		
	(1)	(2)	(3)	(1)	(2)	(3)
	3SLS	3SLS	2SLS	3SLS	3SLS	2SLS
gdpi	0.540 (0.005)***	0.344 (0.043)***	0.378 (0.030)***	0.537 (0.005)***	0.346 (0.042)***	0.378 (0.030)***
gdpj	0.513 (0.005)***	0.527 (0.043)***	0.516 (0.030)***	0.516 (0.005)***	0.525 (0.043)***	0.516 (0.030)***
tij	-6.251 (0.037)***	-7.000 (0.039)***	-6.857 (0.227)***	-6.252 (0.037)***	-7.000 (0.039)***	-6.857 (0.227)***
Constant	-2.613 (0.211)***	2.356 (1.421)*	2.211 (1.371)	-2.619 (0.210)***	3.282 (1.843)*	2.180 (1.371)
expij	–	–	–	-0.065 (0.001)***	-0.031 (0.010)***	-0.036 (0.004)***
trfi	-0.146 (0.029)***	-0.035 (0.021)*	-0.028 (0.009)***	-0.019 (0.013)	-0.023 (0.017)	-0.014 (0.008)*
trfj	-0.156 (0.029)***	-0.036 (0.021)*	-0.028 (0.009)***	-0.024 (0.014)*	-0.021 (0.017)	-0.019 (0.007)**
dist	0.141 (0.002)***	0.172 (0.002)***	–	0.097 (0.001)***	0.134 (0.012)***	–
lang	-0.149 (0.009)***	-0.086 (0.005)***	–	-0.066 (0.004)***	-0.067 (0.007)***	–
bor	-0.103 (0.008)***	-0.055 (0.005)***	–	-0.027 (0.004)***	-0.043 (0.006)***	–
cwni	-0.032 (0.006)***	-0.119 (0.014)***	–	-0.009 (0.003)***	0.055 (0.023)**	–
cwnj	-0.032 (0.006)***	-0.243 (0.010)***	–	-0.010 (0.003)***	-0.122 (0.042)***	–
ebli	0.080 (0.007)***	0.057 (0.011)***	–	0.019 (0.003)***	0.176 (0.041)***	–
eblj	0.080 (0.007)***	0.050 (0.011)***	–	0.021 (0.003)***	-0.004 (0.026)	–
isli	0.035 (0.005)***	0.029 (0.010)***	–	-0.013 (0.003)***	0.035 (0.008)***	–
islj	0.036 (0.005)***	0.146 (0.014)***	–	-0.011 (0.003)***	0.026 (0.008)***	–
landli	0.081 (0.006)***	0.136 (0.008)***	–	0.011 (0.003)***	0.187 (0.064)***	–
landlj	0.079 (0.006)***	0.134 (0.008)***	–	0.012 (0.003)***	-0.012 (0.015)	–
eui	-0.044 (0.004)***	-0.067 (0.006)***	-0.061 (0.003)***	-0.017 (0.002)***	-0.051 (0.009)***	-0.037 (0.004)***
euj	-0.041 (0.004)***	-0.067 (0.006)***	-0.061 (0.003)***	-0.016 (0.002)***	-0.048 (0.006)***	-0.049 (0.003)***
exvol	0.002 (0.002)	-0.004 (0.001)***	0.000 (0.001)	0.004 (0.001)***	-0.003 (0.001)***	0.000 (0.001)
Constant	1.004 (0.172)***	-0.041 (0.129)	1.188 (0.056)***	1.559 (0.081)***	0.570 (0.285)**	1.722 (0.086)***

Note: Robust standard errors in parentheses.

\* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%.



Table 4.3 shows the results after using the instrumental variable and simultaneous equation model techniques. The left three columns show the results under the restriction that there are no economies of scale in the trade sector using the IV estimator. The right three columns comprise the unrestricted case of estimating a SEM specification. Three important outcomes can be seen from this table. First, the impact of exports on trade costs is significantly negative in the three SEM specifications. Second, the signs of the trade cost variables ( $w_{ij}^k$ ) have the expected signs at least in the country-pair-year fixed effects specification. Third, the estimates for trade cost proxies are lower in the SEM specification compared to the IV specification. That means that controlling for economies of scales in the trade sector lowers the estimated direct impacts of the trade cost proxies by 20 to 50% in the three-way specification.

To examine which kind of specification of the SEM estimation has the best fit, the residuals of the pooled, country-year fixed effects and country-pair-year fixed effects model are plotted in figures 4.3 (for the gravity equation) and 4.4 (for the trade cost equation). Just like in figure 4.2, the residuals of the three-way model are distributed around zero, which indicates that the country-pair-year fixed effects specification makes the best fit. Note that in the SLS estimation, the deviation of the residuals from zero is lower than in the OLS estimation.

The Durbin-Wu-Hausman-test is performed to check whether a simultaneity problem exists. The test consists of two steps. In the first step,  $t_{ij}$  is regressed on all exogenous variables in the model to calculate the estimated residual vector  $\hat{\epsilon}_{ij}$  – see reduced form equation (4.4). In the second step, this residual vector is plugged into the structural equation of interest (4.1) as an additional regressor. The null hypothesis is that the coefficient of the residuals is 0. If the Wald-test suggests that the coefficient of  $\hat{\epsilon}_{ij}$  is significantly unequal to 0, an interdependent relationship between equations (4.2) and (4.1) is likely. In the present case, such a relation could be found in all three panel specifications. Estimating the structural form directly via OLS yields biased and inconsistent results while the results of the simultaneous equation model are at least consistent. In all three panel specifications, the null hypothesis is rejected and consequently the application

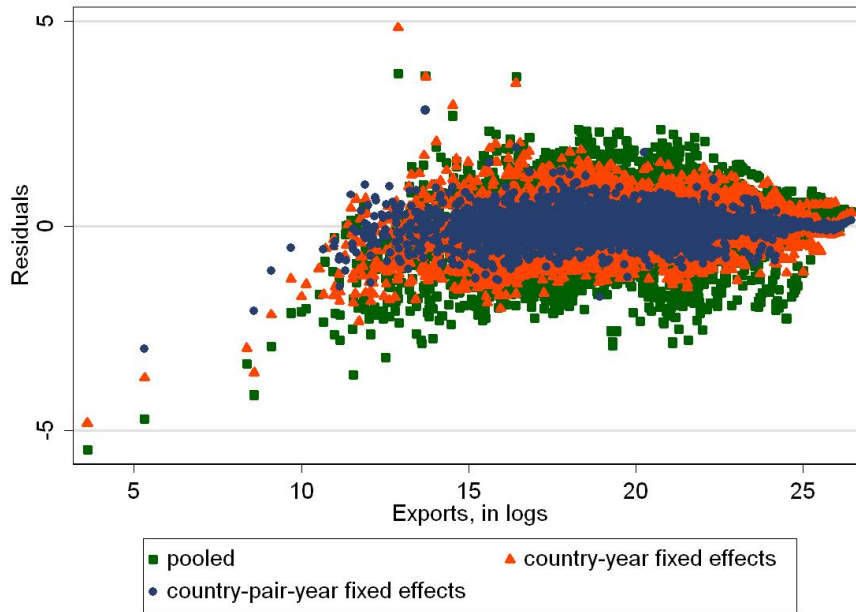
of the SEM-strategy should be preferred.

However, the Sargan-test rejects the hypothesis that the instrument variables are chosen adequately. In this test,  $\text{exp}_{ij}$  is regressed on  $\text{gdp}_i$ ,  $\text{gdp}_j$  and  $\text{tij}$  instrumented by the trade cost proxies (and fixed-effects dummies, respectively) to achieve the residual vectors  $\hat{u}_{ij}^{IV}$ . This is the residual vector from the IV-regression reported in the upper-left part of table 4.3. Then, the obtained residual vector  $\hat{u}_{ij}^{IV}$  is regressed on all exogenous variables in the model:  $\text{gdp}_i$ ,  $\text{gdp}_j$  and all trade cost proxies. From this regression, the Sargan-test-statistic can be computed as the product of the number of observations and the coefficient of determination (R-square). R-square close to 0 implies that there is less correlation between the instruments and the error term, and therefore the instruments tend to be exogenous. This is the null hypothesis. Under the alternative hypothesis, the instruments are correlated with the error term and are therefore endogenous. The test statistic follows a  $\chi_{k-r}^2$  distribution, where  $k - r$  is the difference between the number of instruments (or trade cost proxies, respectively) minus the number of endogenous variables on the right-hand side (which is one:  $\text{tij}$ ). The null hypothesis of valid instruments must be rejected in all three panel specifications, meaning that the results of the instrumental variable and simultaneous approach should be biased. Note that the instruments used in these approaches are also used in the traditional gravity equations – like equation (4.3) – to proxy trade costs. If they are endogenous, they also bias the standard (reduced form) specification. Consequently, not only is the IV/SEM approach biased, but the traditional approach to estimate the gravity equation as well, if these variables are not chosen adequately. It will be a task for future research to find more adequate variables that help to fit the trade cost function. Especially, those variables which reflect more direct cost sources, might be a key as it was discussed in the conclusions of the previous chapter, though there are hardly any data available.

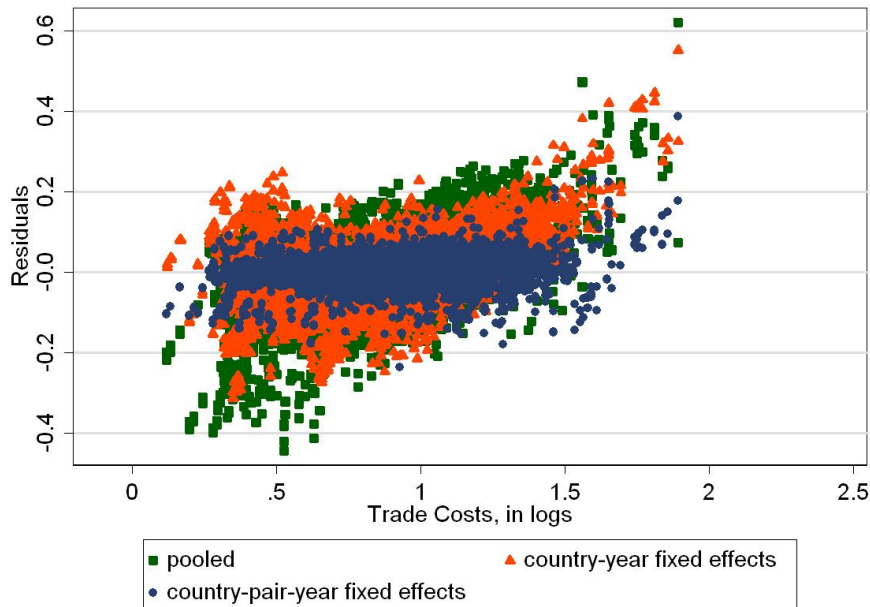
Two central research questions were proposed in section 4.1. In summary, the results suggest the following answers:

**Answer 1** *Are there economies of scale in the trade sector?*

Evidently yes. The results reported by table 4.3 yield the following conclusions: First, the



**Figure 4.3:** Residual-Analysis of the Theory-Based Gravity Equation: Pooled Regression, Country-Year Fixed Effects and Country-pair-Year Fixed Effects.



**Figure 4.4:** Residual-Analysis of the Theory-Based Trade Cost Equation: Pooled Regression, Country-Year Fixed Effects and Country-pair-Year Fixed Effects.

effect of exports on trade costs is significantly negative in the SEM-specification. Second, the Hausman-test indicates the presence of simultaneity.

**Answer 2** *How are estimates biased if economies of scale in the trade sector persist, but go unaddressed?*

Ignoring the endogeneity of trade costs tends to overestimate the effects of trade cost proxies. This can be seen after comparing the left part of table 4.3 with the right part. Nearly all of the  $w_{ij}^k$ -variables are considerably lower in the SEM estimation. However, the results of all estimated models must be interpreted carefully since the Sargan-test indicates that the  $w_{ij}^k$ -variables are not chosen adequately – a problem that cannot easily be solved. This is due to the lack of data for variables that better reflect the components of trade costs rather than of country characteristics.

## 4.4 Conclusion

Studies that apply the gravity equation take trade costs as exogenously given. However, theoretical considerations and intuition from chapter 3 suggest that exports between two countries depend on bilateral trade costs and bilateral trade costs depend on the exports between the two countries (if there are economies of scale in the trade sector). If this interdependence between exports and trade costs exists and is not addressed, estimates might be biased.

The empirical results of this chapter give evidence that economies of scale in international trade do exist. Using a 3SLS/2SLS regression yields the result that higher trade between two countries implies lower bilateral trade costs. Comparing the results of this regression with an assimilable IV approach, where the impact of exports on trade costs is neglected by assumption, shows that ignoring the interaction of exports and trade costs tends to result in higher coefficients. This result might be a contribution to the broad discussion of presumably, overly high coefficient estimates in studies using the gravity equation. It also shows that purely regressing the trade cost index on its determinants and ignoring

the effect of exports on the trade cost index in the structural form (4.2) – or similarly the GDPs in the reduced form trade cost equation (4.4) – should lead to biased results.<sup>13</sup>

However, it will be a task for future research to find more adequate exogenous variables for the trade cost function. The Sargan-test shows that these variables are not completely exogenous. This also implies that they are not appropriate as proxies for trade costs in the traditional estimation strategy. As a consequence, this leads to biased results for both the traditional approach and the IV/SEM approach pursued in this study. Using variables that more exactly reflect trade cost components instead of country characteristics might help to solve this problem, but such variables are hardly available.

The estimation of a gravity model as a simultaneous equation model with a gravity equation and a trade cost equation becomes feasible with a comprehensive index for the tariff equivalents of bilateral trade costs, as it has been proposed by Novy (2007). The presence of a measure for comprehensive trade costs between country pairs also makes it possible to compute empirical values for the multilateral resistance terms (that Anderson and van Wincoop (2003) introduced into the gravity literature). With concrete data for trade costs and multilateral resistances, we can quantify each variable on the right-hand side of the gravity equation by Anderson and van Wincoop (2003). This will be the task of the next two chapters. The following chapter demonstrates a way to compute the concrete values of multilateral resistances. Chapter 6 presents the results of estimating the theory-based gravity equation with respect to quantified multilateral resistances.

---

<sup>13</sup>Examples of the application of a trade cost function that ignores exports or country size are: Jacks, Meissner, and Novy (2008), and Olper and Raimondi (2009).

## Chapter 5

# Computing Multilateral Resistances

This chapter builds on a theoretical derivation of the gravity equation provided by Anderson and van Wincoop (2003). They conclude that exports depend not only on bilateral trade costs, but also on bilateral trade costs relative to a measure of both countries' trade costs with all other countries (i.e. the so-called *multilateral resistances*).

The aim in this chapter is to find a direct computational solution for multilateral resistances. On its right-hand side, the theory-based gravity equation (3.8) has a directly measurable part, containing the GDPs, and an indirectly measurable part, containing trade costs and multilateral resistances. The indirectly measurable part is usually obtained by replacing  $t_{ij}$  via proxy variables (like distance, exchange rate volatilities, membership in a certain country group and many more) and controlling multilateral resistances by fixed-effect dummies (country or country-pair dummies). In the previous chapter, I replaced the indirect method of considering bilateral trade costs by a novel index (Novy, 2007) that makes it possible to yield direct data for bilateral trade costs. In this chapter, I use this index to compute numerical values for the multilateral resistances of the trading countries. Consequently, I achieve data for all the right-hand side variables of Anderson and van Wincoop's theory-based gravity equation.

Recent work by Baier and Bergstrand (2009) pursues a similar aim. They use a Taylor-series expansion to solve for multilateral resistances. However, this approach requires a normalization of the multilateral resistances to a reference country, so that each computed multilateral resistance must be interpreted relative to a certain country that has to be chosen in advance. In contrast, my approach is able to compute direct absolute values for the multilateral resistances. A normalization to a certain country is not necessary.

This chapter is organized as follows. Section 5.1 explains how multilateral resistances work in the theory-based gravity equation. The theory-based index for trade costs by Novy (2007) was already introduced in the previous chapter. The calculation of this index is briefly repeated in section 5.2. The presence of direct data for bilateral trade costs makes it possible to solve the multilateral resistance terms. A procedure to do so is demonstrated in section 5.3. Section 5.4 concludes.

## 5.1 Background

In a general equilibrium framework, with many countries trading composite goods that are differentiated by country of origin, Anderson and van Wincoop (2003) derive the following gravity equation:

$$X_{ij} = \frac{Y_i \cdot Y_j}{Y_w} \cdot \left( \frac{t_{ij}}{P_i \cdot P_j} \right)^{1-\sigma}. \quad (5.1)$$

Here,  $Y_i$  and  $Y_j$  are the exogenously given GDPs of the countries.  $Y_w$  symbolizes the GDP of the whole world. Each country's GDP is assumed to present a country characteristic composite good, and  $\sigma$  is the elasticity of substitution between these goods. Moreover, it is assumed that  $\sigma > 1$ , which is supported by empirical evidence (see Anderson and van Wincoop, 2004). Trade costs in terms of iceberg trade costs are indicated by  $t_{ij} > 1$ . These iceberg costs can be interpreted as a tariff equivalent: selling a good from country  $i$  in country  $j$  raises the price on country  $j$ 's market by  $(t_{ij} - 1)\%$ . Keep in mind that this modeling of trade costs reflects per-unit trade costs rather than total trade costs. This means that  $t_{ij}$  describes the average markup of trade costs on each dollar of transport value. Furthermore, it is assumed that domestic trade costs are benchmarked to 1,  $t_{ii} = 1$ , and that transport costs between two countries are symmetric,  $t_{ij} = t_{ji}$ .<sup>1</sup>  $P_i$  and  $P_j$  denote the exogenously given multilateral resistances of the exporting or importing country, respectively. They are derived from a Dixit-Stiglitz price index and can be

---

<sup>1</sup> This assumption of symmetry could be relaxed, but because the empirical trade cost index introduced in section 5.2 is a geometric mean of trade costs and thus a symmetric measure of trade costs, this assumption helps to simplify.



written as:<sup>2</sup>

$$P_i = \left( \sum_j \left( \frac{t_{ij}}{P_j} \right)^{1-\sigma} \cdot s_j \right)^{1/(1-\sigma)}, \quad (5.2)$$

$$P_j = \left( \sum_i \left( \frac{t_{ij}}{P_i} \right)^{1-\sigma} \cdot s_i \right)^{1/(1-\sigma)}. \quad (5.3)$$

Multilateral resistances can be interpreted as an index for the overall accessibility to trade of a country. In the second multiplier of gravity equation (5.1), bilateral per-unit trade costs  $t_{ij}$  appear in relation to the respective countries' multilateral resistances.<sup>3</sup> For illustration, imagine two countries lying isolated from the rest of the world on one island in the ocean, far away from the next continent. Bilateral average trade costs measured by iceberg-factor  $t_{ij}$  might be low and this should guarantee a higher trade volume between both countries. Yet, the relatively high trade costs with the rest of the world have an additional, positive effect on the bilateral trade volume. If the same two island-countries were two small countries in the middle of a huge continent, surrounded by many large countries, multilateral resistances would probably be much lower and thus the trade volume between the two countries would be lower, even if for the GDPs and  $t_{ij}$  the same levels are chosen.

---

<sup>2</sup> Note that Anderson and van Wincoop (2003) distinguish more precisely between multilateral resistances of exporting countries on the one hand and multilateral resistances of importing countries on the other hand, denoted as  $\Pi_i$  and  $P_j$  in the previous chapters, see equation (3.8). If trade costs are assumed to be symmetrical between all countries ( $t_{ij} = t_{ji}$ ), which is also a relevant assumption for this study, it can be shown that the export multilateral resistance of a country equals the import multilateral resistance, so that  $\Pi_i = P_i$  (Anderson and van Wincoop, 2003, p. 175).

<sup>3</sup> Notably, the effect of multilateral price indices was already stated in the first theoretical derivations of the gravity equation (see Anderson, 1979; Bergstrand, 1985). But Anderson and van Wincoop (2003) concentrated this issue on the elegant formulation of equation (5.1) and were able to conclude that ignoring multilateral resistances leads to biased results.

## 5.2 Computing Bilateral Trade Costs

Before we start to solve the multilateral resistance equations (5.2) and (5.3), we need data for the bilateral trade costs  $t_{ij}$ . Building on the theoretical framework of the gravity equation introduced by Anderson and van Wincoop (2003), Novy (2007) derives an index for the geometric mean of the bilateral trade costs between two countries:

$$t_{ij} = \left( \frac{X_{ii}X_{jj}}{X_{ij}X_{ji}} \right)^{\frac{1}{2(\sigma-1)}}. \quad (3.15)$$

In this index, trade barriers between two countries are a function of the ratio between intra-national trade ( $X_{ii}, X_{jj}$ ) and international trade ( $X_{ij}, X_{ji}$ ). The higher the trade inside a country relative to its exports to the other country, the higher the bilateral trade costs will be, since  $\sigma$  is assumed to be larger than 1. Note that this index is a comprehensive measure of trade costs. These comprehensive trade costs can be deconstructed into measurable components and not measurable components.<sup>4</sup>

As the necessary data are available from several sources, equation (3.15) makes it feasible to compute a theory-based index for the overall trade costs between two countries. In this chapter, I use a set of 23 OECD countries for the years 1995 to 2005.<sup>5</sup> The data source for bilateral exports is the bilateral trade statistics of OECD's Structural Analysis (OECD STAN). Following Novy (2007), intra-national trade flows are computed as a country's total production minus total exports. If it is available, the production data is taken from the OECD STAN data (converted into US Dollars using the OECD Financial Indicators annual exchange rates). Since there are many missing observations (e.g. Turkey is altogether unreported in this data set), I compensate for the missing data by using data

---

<sup>4</sup> See Anderson and van Wincoop (2004) for a comprehensive discussion of trade costs. They decompose overall trade costs into three classes: transport costs, border-related costs and retail/wholesale costs. See also section 3.3 of this study.

<sup>5</sup> The countries are selected so that the full data for all variables which are necessary to compute multilateral resistances are available for the considered period. This is necessary to make the resulting values for multilateral resistances comparable over time. The countries covered by the data set are: Australia, Austria, Canada, Denmark, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Japan, Korea, Netherlands, Norway, Poland, Portugal, Spain, Sweden, Switzerland, Turkey, United Kingdom and the United States.





that the start value is independent of the respective country. Then we use the resulting  $\mathcal{P}_{i(1)}$ -vector from this calculation to compute  $\mathcal{P}_{j(2)} = \left( \sum_i \left( \frac{1}{\mathcal{P}_{i(1)}} \right) \cdot \vartheta_{ij} s_i \right)$  in a second round. This procedure is repeated until each  $\mathcal{P}_i$  converges, meaning that there are no (or negligibly few) changes after several repeated recalculation rounds. In this case the value of each  $\mathcal{P}$  on the right-hand side equals the value of the corresponding  $\mathcal{P}$  on the left-hand side: we yield a certain value for each  $\mathcal{P}_i$  on the left-hand side that is equal to each  $\mathcal{P}_i$ s plugged in on the right-hand sides of the equations. This must be one solution of the equation system.

How do we choose the right value,  $\mathcal{P}_{(0)}^*$ ? Running the recalculation of the data sample with overly small values of  $\mathcal{P}_{(0)}$  leads to an alternating sequence: the results of the odd rounds of recalculation are too low, the results of the even rounds are too high and so on. Running the recalculation of the data sample with overly high values for  $\mathcal{P}_{(0)}$  leads to an adverse alternating sequence, where the odd recalculation rounds are too high and the even recalculations too low. The closer  $\mathcal{P}_{(0)}$  is to the optimal starting value  $\mathcal{P}_{(0)}^*$ , the smaller is the amplitude of the recalculated values.

### 5.3.2 An Illustrative Example

An example shall help to understand this procedure. With a few tricks, a set of three polynomial equations can easily be transformed into a square linear equation system solvable with Cramer's Rule. This direct solution is a reliable benchmark for the results from the numerical procedure. Although the tricks to achieve the square linear system do not exactly meet the assumptions of the original model, this example might help one to understand the mechanics of solving the equations.

Starting from equation system (5.5), define the unknown  $z_{ij} = z_i \cdot z_j$  and multiply each summand in the parentheses on the right-hand side of (5.5) with  $z_i$  for the case  $C = 3$ ,

to get:

$$\begin{aligned}
 1 &= z_{11}b_{11} + z_{12}b_{12} + z_{13}b_{13}, \\
 1 &= z_{21}b_{21} + z_{22}b_{22} + z_{23}b_{23}, \\
 1 &= z_{31}b_{31} + z_{32}b_{32} + z_{33}b_{33}.
 \end{aligned} \tag{5.6}$$

Assume that  $b_{ii} = 0$ . Note that this assumption does not adequately reflect the definition of multilateral resistances. But it is necessary to get a symmetric linear equation system. Following the assumptions of the economic model, each  $b_{ii}$  is strictly greater than 0 because  $t_{ii} = 1$ ,  $\sigma > 1$  and  $0 < s_i < 1$ . More precisely, this changed assumption ignores country  $i$  itself in the summation of all countries to compute multilateral resistances as given by equation (5.2).<sup>7</sup> Since  $z_{ij} = z_{ji}$ , because  $1/(\mathcal{P}_i\mathcal{P}_j) = 1/(\mathcal{P}_j\mathcal{P}_i)$ , it becomes possible to rewrite equation system (5.6) into a square linear equation system:

$$\begin{aligned}
 1 &= z_{12}b_{12} + z_{13}b_{13} + 0, \\
 1 &= z_{12}b_{21} + 0 + z_{23}b_{23}, \\
 1 &= 0 + z_{13}b_{31} + z_{23}b_{32},
 \end{aligned} \tag{5.7}$$

or in matrix form:

$$\underbrace{\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}}_{\mathbf{1}} = \underbrace{\begin{pmatrix} b_{12} & b_{13} & 0 \\ b_{21} & 0 & b_{23} \\ 0 & b_{31} & b_{32} \end{pmatrix}}_{\mathbf{B}} \cdot \underbrace{\begin{pmatrix} z_{12} \\ z_{13} \\ z_{23} \end{pmatrix}}_{\mathbf{z}}. \tag{5.8}$$

Using Cramer's Rule, this square linear equation system can easily be solved for the three unknowns ( $z_{12}^*$ ,  $z_{13}^*$ ,  $z_{23}^*$ ):

$$z_{12}^* = \frac{1}{\mathcal{P}_1\mathcal{P}_2} = \frac{-b_{13}b_{23} + b_{23}b_{31} + b_{13}b_{32}}{b_{12}b_{23}b_{31} + b_{13}b_{21}b_{32}}, \tag{5.9}$$

$$z_{13}^* = \frac{1}{\mathcal{P}_1\mathcal{P}_3} = \frac{b_{12}b_{23} - b_{12}b_{32} + b_{21}b_{32}}{b_{12}b_{23}b_{31} + b_{13}b_{21}b_{32}}, \tag{5.10}$$

$$z_{23}^* = \frac{1}{\mathcal{P}_2\mathcal{P}_3} = \frac{b_{13}b_{21} + b_{12}b_{31} - b_{21}b_{31}}{b_{12}b_{23}b_{31} + b_{13}b_{21}b_{32}}. \tag{5.11}$$

---

<sup>7</sup> Equation (5.2) becomes:

$$P_i = \left( \sum_{j \neq i} \left( \frac{t_{ij}}{\mathcal{P}_j} \right)^{1-\sigma} \cdot s_j \right)^{1/(1-\sigma)} \quad \text{instead of} \quad \left( \sum_j \left( \frac{t_{ij}}{\mathcal{P}_j} \right)^{1-\sigma} \cdot s_j \right)^{1/(1-\sigma)}.$$

The general theory-based case of transforming the polynomial equation system into a linear equation system is presented in appendix C.

**Table 5.1:** Assumed Data for the Numerical Example.

Country	1	2	3
GDP-share $s_i$	40%	35%	25%
	Trade Costs $t_{ij}$		
1	0	1.2	1.3
2	1.2	0	1.4
3	1.3	1.4	0

From these solutions it is possible to compute the values for the desired multilateral resistances. Solving the system  $z_{ij} = 1/(\mathcal{P}_i \mathcal{P}_j)$  for each  $\mathcal{P}_i$  with  $i, j \in \{1, 2, 3\}$  yields

$$P_1 = \mathcal{P}_1^{1/(1-\sigma)} = \left( \sqrt{\frac{z_{23}^*}{z_{12}^* z_{13}^*}} \right)^{1/(1-\sigma)}, \quad (5.12)$$

$$P_2 = \mathcal{P}_2^{1/(1-\sigma)} = \left( \sqrt{\frac{z_{13}^*}{z_{12}^* z_{23}^*}} \right)^{1/(1-\sigma)}, \quad (5.13)$$

$$P_3 = \mathcal{P}_3^{1/(1-\sigma)} = \left( \sqrt{\frac{z_{12}^*}{z_{13}^* z_{23}^*}} \right)^{1/(1-\sigma)}. \quad (5.14)$$

Note that a real solution for the multilateral resistances can only be obtained, if there is no negative  $z_{ij}^*$ . This condition depends on the values of the  $b_{ij}$ -coefficients:  $t_{ij}$ ,  $s_j$  and  $\sigma$ .<sup>8</sup>

Now, we consider a numerical example for the three country model. First we solve the equations using Cramer's Rule to get a benchmark and then we apply the numerical algorithm. We use the data for countries 1, 2 and 3 of table 5.1 as given with an elasticity of substitution  $\sigma = 8$  and remember that the impact of a country on its own multilateral resistances is ignored by assumption ( $t_{ii} = 0$ ) to provide a special case, where a simple solution of the equation system is possible. Country 1 is the biggest country and has the lowest trade costs compared to the other two countries. Thus, we expect a low multilateral resistance. For country 3, the opposite should be the case.

We directly solve the multilateral resistances to obtain a reliable reference. The values for

---

<sup>8</sup> All four determinants used for Cramer's Rule, the main determinant and the three column-replaced determinants, must be greater than 0, or all of them must be smaller than 0.

the coefficients  $b_{ij} = t_{ij}^{-7} s_j$  can be directly computed from table 5.1. Applying equations (5.9) to (5.11) yields the solution to the equation system (5.8):

$$z_{12}^* = 6.399, \tag{5.15}$$

$$z_{13}^* = 9.412, \tag{5.16}$$

$$z_{23}^* = 12.047. \tag{5.17}$$

Because no  $z_{ij}^*$  is negative, it is possible to find real positive solutions for the multilateral resistances by applying equations (5.12) to (5.14):

$$P_1 = 1.122 \tag{5.18}$$

$$P_2 = 1.162 \tag{5.19}$$

$$P_3 = 1.228 \tag{5.20}$$

Does the algorithm yield the same numbers for multilateral resistances? Under the assumptions and with the numbers of table 5.1, we can rewrite equation system (5.4) thus:

$$\begin{aligned} \mathcal{P}_1 &= 0 + \frac{1}{\mathcal{P}_2} \cdot 0.098 + \frac{1}{\mathcal{P}_3} \cdot 0.040, \\ \mathcal{P}_2 &= \frac{1}{\mathcal{P}_1} \cdot 0.112 + 0 + \frac{1}{\mathcal{P}_3} \cdot 0.024, \\ \mathcal{P}_3 &= \frac{1}{\mathcal{P}_1} \cdot 0.064 + \frac{1}{\mathcal{P}_2} \cdot 0.033 + 0, \end{aligned} \tag{5.21}$$

Now we choose a common value for all multilateral resistances,  $P_{(0)} \equiv \mathcal{P}_0^{-7} = \mathcal{P}_{1(0)}^{-7} = \mathcal{P}_{2(0)}^{-7} = \mathcal{P}_{3(0)}^{-7}$ , and insert this value into all the  $\mathcal{P}$ 's on the right-hand side of equation system (5.21) to obtain the values for the  $\mathcal{P}$ 's on the left-hand side,  $\mathcal{P}_{1(1)}$ ,  $\mathcal{P}_{2(1)}$  and  $\mathcal{P}_{3(1)}$ . Next, these values are used on the right-hand side to compute  $\mathcal{P}_{1(2)}$ ,  $\mathcal{P}_{2(2)}$ ,  $\mathcal{P}_{3(2)}$  and so on.

The results for different start values  $P_{(0)}$  are reported in table 5.2. Four important outcomes can immediately be seen from the numbers of this table. First, where the start value  $P_{(0)}$  is 1.15 (third line), differences between the individual rounds of recalculation are relatively small, compared to the other results. These differences grow both if the value for  $P_{(0)}$  shrinks to 1.00 or rises to 1.50. Second, if starting values are lower than

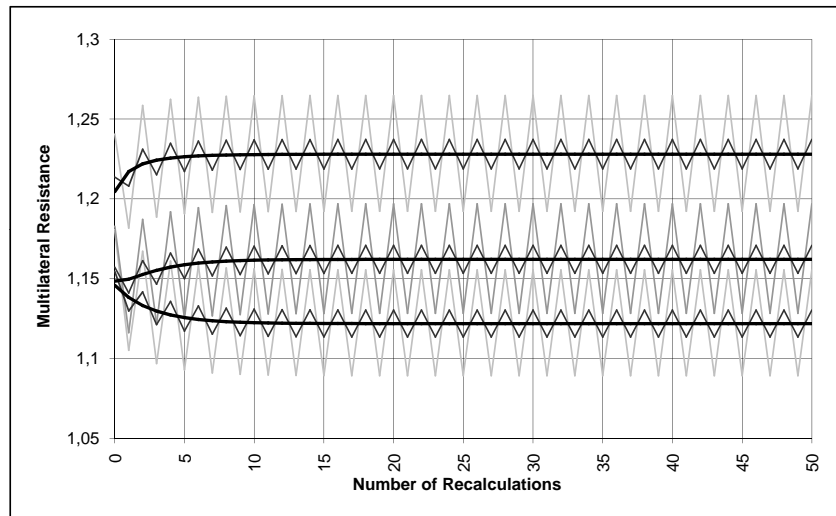


1.15 (line 1 and 2), results of an odd recalculation round are lower than those of an even one. If a starting value larger than 1.15 is used (line 4 and 5), the opposite is the case: the values from the 99th recalculation, for example, are always higher than the values from the 100th recalculations. Third, after many recalculations the results alternate around singular values of  $P_1$ ,  $P_2$  and  $P_3$ . The value after 97 recalculations is the same as after 99 recalculations, the value after 98 is the same as after 100 recalculations. And fourth, after a sufficiently high number of recalculations, the desired values given by the direct solution of the equation (5.18) to (5.20) always lie between the maximum and minimum values of the alternating sequences.

Is it possible to find a start value  $P_{(0)}^*$ , where no differences between the single recalculation rounds appear anymore? To face this problem, a search algorithm is used. Start for example with  $P_{(0)} = 1.00$ . The values after 100 recalculations are larger than those after 99 recalculations. Repeat the procedure with  $P_{(0)} = 1.10$  and then with  $P_{(0)} = 1.20$ . In the latter case, the alternating sequence changes: values after 100 recalculations are lower than after 99. Now go down in steps of 0.01 until the structure changes at 1.15, go up in steps of 0.001 and stop when the changes between recalculation 99 and 100 are small enough, i.e. measured by the sum of differences between the values in the last two recalculation rounds undertaken. This procedure leads to an optimal start value  $P_{(0)}^* \approx 1.158672$ . The more exact this start value is chosen, the closer the results of the numerical simulation fit the results of equations (5.18) to (5.20).

**Table 5.2:** Simulation of the Numerical Example.

$P^{(0)}$	0	1	2	3	4	5	6	...	97	98	99	100
$P_A$	1.328	0.982	1.313	0.975	1.306	0.971	1.302		0.968	1.300	0.968	1.300
1.00 $P_B$	1.331	0.992	1.335	0.997	1.341	1.000	1.344	...	1.003	1.346	1.003	1.346
$P_C$	1.396	1.050	1.415	1.057	1.420	1.058	1.422		1.060	1.423	1.060	1.423
$P_A$	1.207	1.081	1.194	1.072	1.187	1.069	1.184		1.065	1.182	1.065	1.182
1.10 $P_B$	1.210	1.091	1.214	1.097	1.219	1.100	1.222	...	1.103	1.224	1.103	1.224
$P_C$	1.269	1.155	1.287	1.162	1.291	1.164	1.292		1.166	1.293	1.166	1.293
$P_A$	1.154	1.130	1.142	1.121	1.136	1.117	1.133		1.113	1.130	1.113	1.130
1.15 $P_B$	1.157	1.141	1.161	1.147	1.166	1.150	1.168	...	1.153	1.171	1.153	1.171
$P_C$	1.217	1.208	1.231	1.215	1.235	1.217	1.236		1.219	1.237	1.219	1.237
$P_A$	1.106	1.179	1.094	1.170	1.088	1.166	1.086		1.162	1.083	1.162	1.083
1.20 $P_B$	1.109	1.191	1.113	1.196	1.117	1.200	1.120	...	1.204	1.122	1.204	1.122
$P_C$	1.163	1.260	1.180	1.268	1.183	1.270	1.185		1.272	1.186	1.272	1.186
$P_A$	0.885	1.474	0.875	1.462	0.871	1.457	0.869		1.452	0.867	1.452	0.867
1.50 $P_B$	0.887	1.488	0.890	1.495	0.894	1.500	0.896	...	1.504	0.898	1.504	0.898
$P_C$	0.930	1.576	0.944	1.585	0.947	1.588	0.948		1.590	0.949	1.590	0.949



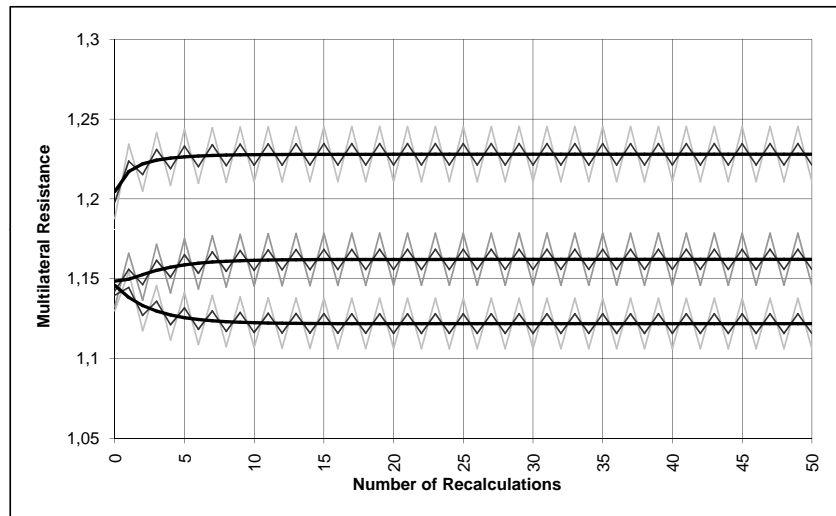
**Figure 5.1:** Simulation Results of the Three Country Example (Upward Approximation):  
 $P_{(0)} = 1.125$  (light gray, large amplitude),  $P_{(0)} = 1.15$  (gray, small amplitude),  
 $P_{(0)} = P_{(0)}^* = 1.158672$  (black, steady course).

Figures 5.1 and 5.2 finally illustrate the procedure graphically. Figure 5.1 starts with values  $P_{(0)} < P_{(0)}^*$ . The amplitude first goes down, then up. Figure 5.2 starts with values  $P_{(0)} > P_{(0)}^*$ . The amplitude first goes up, then down. In both cases, more distance from  $P_{(0)}^*$  increases the amplitude between the recalculation results (light gray and gray alternating sequences). Choosing the optimal value  $P_{(0)}^*$  leads to a steady course (black bold sequence). Notice that the sequences with the optimal values converge toward the solutions of the equation system which were detected with Cramer's Rule.

### 5.3.3 Multilateral Resistances of the 23 OECD Countries

The numerical procedure is applied to the data set with 23 OECD Countries.<sup>9</sup> Because this panel data set comprises data of 11 years, it is necessary to compute the multilateral resistances separately for each year. That means, we must separate the data by year and

<sup>9</sup> To implement the computation, the programming environment of the software package STATA was applied to write a custom program. This program is available upon request.



**Figure 5.2:** Simulation Results of the Three Country Example (Downward Approximation):  
 $P_{(0)} = 1.175$  (light gray, large amplitude),  $P_{(0)} = 1.165$  (gray, small amplitude),  
 $P_{(0)} = P_{(0)}^* = 1.158672$  (black, steady course).

find 11 different start values. The algorithm to find these start values follows the same idea as described in the example: start with 1, go up in steps of 1 until the amplitude changes, then go down in steps of 0.1 until the amplitude changes, then go up in steps of 0.01 and so on.<sup>10</sup> The *start value* of the algorithm is 1. The first *step size* is 1. After each change of the amplitude, the step size is set to one tenth of the step size before. The *number of recalculations* with a given start value is 100.<sup>11</sup>

It is necessary to choose a condition when the convergence is sufficient and the program stops searching for the optimal start value. As a measure of sufficiency, I choose the sum of all differences between the last and the penultimate recalculation round. We thus take the differences of all observations between the 100th and 99th recalculation round and sum it up. If this sum is lower than  $\pm 10^{-6}$  the accuracy of the simulated values is

<sup>10</sup>This algorithm is surely not the most efficient one. For larger data sets more advanced programming efforts should be applied to minimize the runtime of the program.

<sup>11</sup>To check the robustness of the simulation, the number of recalculations was extended up to 150. The results remain exactly the same. The results also remain robust if other start values or step sizes are used. The results remained also the same after taking other start values than 1 (e.g. 0 and 10).

**Table 5.3:** Amplitude and Convergence of the Simulation.

Difference	Mean	Std. Dev.	Min.	Max.
$\Delta_1 = P_{i(100)} - P_{i(99)}$	$-4.19 \cdot 10^{-8}$	$7.33 \cdot 10^{-8}$	$-2.38 \cdot 10^{-7}$	$1.19 \cdot 10^{-7}$
$\Delta_9 = P_{i(100)} - P_{i(91)}$	$-4.19 \cdot 10^{-8}$	$7.33 \cdot 10^{-8}$	$-2.38 \cdot 10^{-7}$	$1.19 \cdot 10^{-7}$
$\Delta_2 = P_{i(100)} - P_{i(98)}$	0	0	0	0
$\Delta_{10} = P_{i(100)} - P_{i(90)}$	0	0	0	0

**Table 5.4:** Descriptive statistics of  $\Delta_1 = P_{i(100)} - P_{i(99)}$ .

Value	Frequency	Percentage	Cumulation
$-2.38 \cdot 10^{-7}$	242	4.35	4.35
$-1.19 \cdot 10^{-7}$	1,650	29.64	33.99
0	3,498	62.85	96.84
$1.19 \cdot 10^{-7}$	176	3.16	100.00
Total	5,566	100.00	

considered to be sufficiently high. For some years a marginal amplitude remains. In these cases, the algorithm is stopped when the step size obtains a value of  $10^{-9}$ .

Are these conditions for breaking the algorithm adequate? There are two important requirements on the simulation: first, the amplitude (meaning the difference between odd and even recalculation rounds) should be zero, and second, the simulation must converge after less than 100 recalculations. Table 5.3 reports some descriptive statistics over all 5,566 observations<sup>12</sup> of the differences between the last (100th) and selected previous recalculation rounds. The approximately optimal start value  $P_{(0)}^*$  results from the search algorithm described above. The summary statistics of these differences between each single realization of the 100th the 99th recalculation round ( $\Delta_1 \equiv P_{i(100)} - P_{i(99)}$ ) are reported in the first line. If these differences deviate from zero (even if only a very small number), there is still an amplitude and the start value obtained by the algorithm is not yet adequate. As can be seen, the mean and standard deviation values are not zero. However, the values are very low: they do not become relevant before the seventh decimal place. So if these deviations from zero are due to a remaining amplitude, they are so small that they can be neglected. It is worthwhile to take a closer look at this amplitude.

<sup>12</sup> 23 countries  $\times$  22 trade partners  $\times$  11 years = 5,566 observations.

Table 5.4 shows that there are only four values taken by  $\Delta_1$  over the whole 5,566 observations:  $-2.38 \cdot 10^{-7}$ ,  $-1.19 \cdot 10^{-7}$ , 0 and  $+1.19 \cdot 10^{-7}$ . 63% of the observations take exactly the value 0, the relatively “large” amplitude of  $-2.38 \cdot 10^{-7}$  affects only 4% of the observations. Because  $-2.38 \cdot 10^{-7}$  is two times  $-1.19 \cdot 10^{-7}$  which is the negative value of  $+1.19 \cdot 10^{-7}$ , it is not unlikely that this bias results from a computational problem like a systematic rounding error caused by the software used.

The second line of table 5.3 shows that over the last 9 recalculation rounds these differences are the same. This means that the amplitude is constant at least over the last 9 recalculation rounds. Analyzing the amplitude yields the same results as described in table 5.4. The two last lines of table 5.3 show that there is no change reported for the 5,566 realizations of the multilateral resistances over the last 2 and the last 10 recalculation rounds:  $\Delta_2$  and  $\Delta_{10}$  are exactly reported to be zero. This implies that the values have a sufficiently low amplitude and that the values have attained their full convergence after less than 90 rounds of recalculation.

Table 5.5 describes the derived multilateral resistance data for the 23 countries of the data set, taken from the eleven years of observation. The country with the lowest multilateral resistances is Canada. This result is unsurprising, because trade costs between Canada and the United States are very low. The United States is the biggest economy in the set which guarantees for a high weight ( $s_j$  around 40%) in the summation over all countries. It has relatively high multilateral resistances. From the definition of multilateral resistances given by equation (5.2) or (5.3), it becomes obvious that the multilateral resistances of a country must be low if this country has extremely low trade costs with an extremely large country that has high multilateral resistances. Additional countries with low multilateral resistances are the Netherlands, Germany, Ireland and the United Kingdom. Countries with high multilateral resistances are the mediterranean countries Greece, Portugal and Turkey, as well as the former Eastern Bloc countries Poland and Hungary. Note that Poland, Turkey and especially Hungary were able to decrease their multilateral resistances over the period from 1995 to 2005, while the level of Greece did not change over this time and Portugal even raised its multilateral resistances. Another country with increasing

**Table 5.5:** Summary statistics of the Multilateral Resistances by Country, over 11 years.

Country	Min.	Mean	Max.	Std. Dev.
Australia	1.444	1.483	1.519	0.022
Austria	1.371	1.399	1.424	0.017
Canada	1.002	1.041	1.123	0.035
Denmark	1.402	1.424	1.437	0.012
Finland	1.487	1.495	1.516	0.008
France	1.287	1.308	1.342	0.017
Germany	1.209	1.244	1.292	0.027
Greece	1.660	1.679	1.703	0.013
Hungary	1.451	1.510	1.632	0.061
Ireland	1.222	1.276	1.377	0.051
Italy	1.358	1.370	1.386	0.008
Japan	1.322	1.361	1.442	0.033
Korea	1.314	1.345	1.398	0.026
Netherlands	1.151	1.206	1.244	0.028
Norway	1.434	1.448	1.469	0.011
Poland	1.500	1.579	1.640	0.043
Portugal	1.623	1.637	1.651	0.008
Spain	1.418	1.442	1.479	0.018
Sweden	1.337	1.356	1.386	0.016
Switzerland	1.332	1.356	1.389	0.016
Turkey	1.551	1.613	1.703	0.047
United Kingdom	1.241	1.269	1.313	0.020
United States	1.442	1.506	1.542	0.028
Total	1.002	1.406	1.703	0.150

multilateral resistances is the United States which might be caused by the terrorist attacks of September 11th 2001 for the years following this date. Figures 5.3 to 5.5 show the development of the multilateral resistances of some OECD countries over the time period from 1995 to 2005.

It must be stated that these results are a rough measure, because the data do not cover the whole world but only 23 OECD countries. This leads to a bias of the results. Because the equation system is a summation, the results must be systematically lower than the true values as the missing countries do not appear in this summation. Furthermore, the missing sum of the absent countries differs for each multilateral resistance solution since the respective bilateral trade costs differ. Therefore, the parameters of the missing multilateral resistances ( $b_{ij}$ ) are affected accordingly. Yet, since the total GDP share of

## 5. Computing Multilateral Resistances

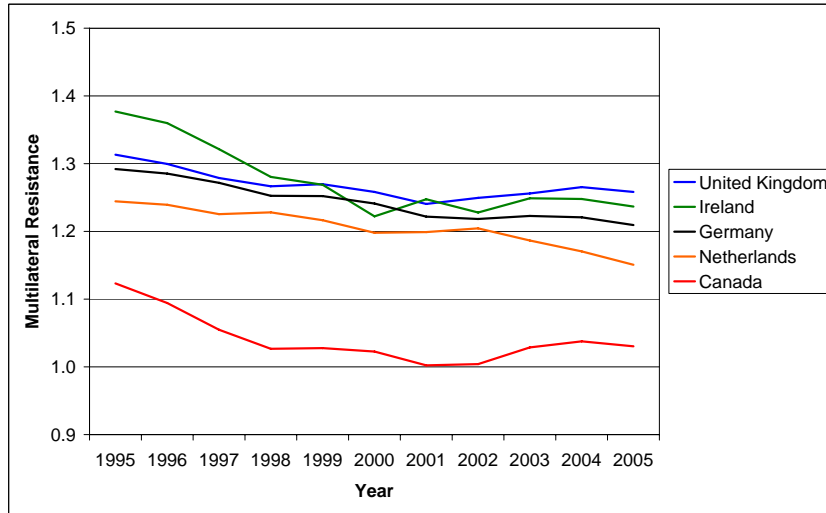


Figure 5.3: Countries with Low Multilateral Resistances 1995 to 2005.

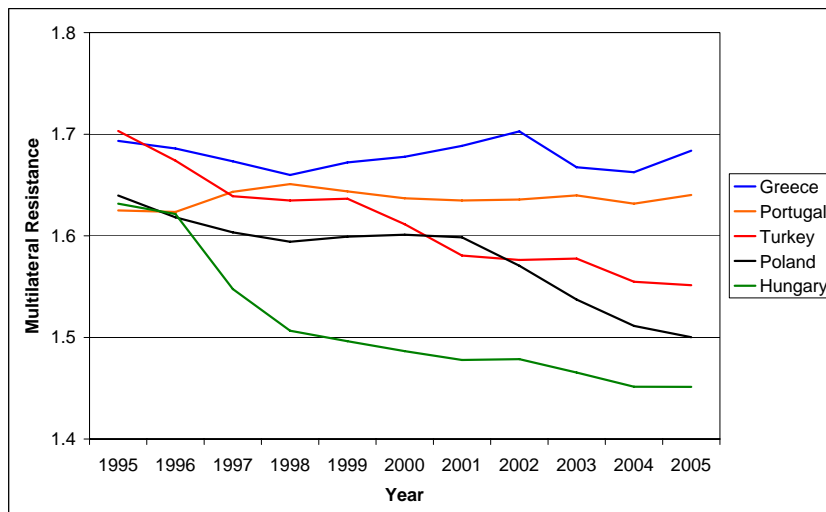
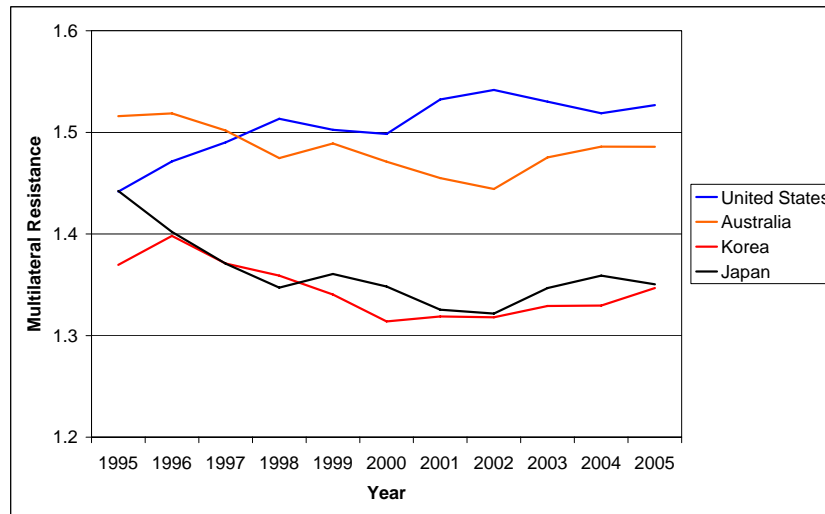


Figure 5.4: Countries with High Multilateral Resistances 1995 to 2005.





**Figure 5.5:** Multilateral Resistances of Australia, Japan, Korea, and the U.S., 1995 to 2005.

the countries is more than 75% of the world GDP (which is the weight in the multilateral resistance formula), the results should not differ too much from the real values. However, for forthcoming studies, there might be one solution to this problem: One could take values for GDP, production and exports for the whole world and subtract the values of the countries directly used for the computation. These residual numbers make it possible to compute the parameters for the rest of the world, taken as one composite, additional country. Adding this to the equations should yield unbiased results.

## 5.4 Conclusion

In their theoretical foundation of the gravity equation, Anderson and van Wincoop (2003) found that trade costs in gravity equations must be seen relatively to the trading countries' multilateral resistances (that reflect the countries' trade barriers to all other countries in the model). Neglecting this issue normally leads to (upward) biased estimates. When applying the gravity equation, it became commonplace in the empirical literature to control

for multilateral resistances by country or country-pair dummies.

This chapter has shown a way to quantify multilateral resistances. Using an index for comprehensive bilateral trade costs as proposed by Novy (2007), it becomes possible to solve the equation system that defines multilateral resistances. Since a direct solution is neither possible nor feasible, a numerical procedure has been developed to compute multilateral resistances. The idea of this procedure is to find an optimal common start value for all countries' multilateral resistances, so that the equation system converges after repeated recalculations. This procedure works with OECD data. For all 11 calculations (for the 11 years) the equation systems converge. Since only 23 countries and not the whole world are considered, the results are biased downwards. But the fact that these considered countries contain the strongest economies of the world should keep the bias small. However, the calculated values of the multilateral resistances are plausible nonetheless.

## Chapter 6

# Estimation with Multilateral Resistances

With the methodology to yield measures for trade costs ( $t_{ij}$ ) and multilateral resistances ( $P_i$  and  $P_j$ ) it becomes possible to estimate the gravity equation (5.1) directly. As it was shown in chapters 3 and 4, direct estimates of equation (5.1) could be biased if there is evidence that per-dollar trade costs are endogenously affected by policy variables and natural trade cost barriers; and further, if they are inversely connected to bilateral export volumes due to economies of scale in the trade sector. In this chapter, I use both the index for trade costs and the computed index for multilateral resistances to estimate the theoretically-founded gravity equation and show how the values of the estimated coefficients shrink.

This chapter is structured as follows. Section 6.1 drafts the estimation strategy and the data to be used. In section 6.2 the results of the estimation are presented. Section 6.3 concludes.

## 6.1 Econometric Model and Data

The standard approach of estimating the gravity equation is:

$$\text{exp}_{ij} = \pi_1 + \pi_2 \text{gdpi} + \pi_3 \text{gdpj} + \sum_{k=4}^{19} \pi_k w_{ijt}^{k-3} + \pi_{pi} p_i + \pi_{pj} p_j + \epsilon_{ijt}, \quad (6.1)$$

where  $\text{exp}_{ij}$  is the log of bilateral exports,  $\text{gdpi}$  and  $\text{gdpj}$  are the logs of the exporting and importing country's GDP, respectively, as it was described in the previous chapter. The exporting country is always denoted by  $i$ , the importing country by  $j$ . Data source for annual data of exports and GDPs is the OECD Structural Analysis Data Base (OECD STAN). The logs of the exporting or importing country's multilateral resistances, as they are computed in section 5.3, are denoted by  $p_i$  and  $p_j$ , respectively. These values result from the calculations presented in the previous chapter. The vector  $w_{ij}^k$  concludes the following trade cost proxies: the freedom of trade index by the Heritage Foundation ( $\text{trfi}$  and  $\text{trfj}$ ), geographic distance between the trading countries in logs ( $\text{dist}$ ), exchange rate volatility in logs ( $\text{exvol}$ ), dummies for common language ( $\text{lang}$ ), common border ( $\text{bor}$ ),

**Table 6.1:** Summary statistics of the OECD data set, over 11 years.

Variable	Mean	Std. Dev.	Min.	Max.	Obs.
expij	20.974	1.77	14.9	26.434	5566
tij	0.753	0.195	0.236	1.34	5566
gdpi/gdpj	26.836	1.257	24.548	30.149	5566
trfi/trfj	4.348	0.069	3.904	4.443	5456
dist	7.907	1.074	5.451	9.803	5566
exvol	0.906	1.259	-12.281	3.463	4982
lang	0.057	0.232	0	1	5566
bor	0.079	0.27	0	1	5566
cwni/cwnj	0.13	0.337	0	1	5566
ebli/eblj	0.087	0.282	0	1	5566
isli/islj	0.174	0.379	0	1	5566
landli/landlj	0.13	0.337	0	1	5566
eui/euj	0.589	0.492	0	1	5566

EU membership of the exporting or importing country (eui and euj), landlocked location (landli and landlj), location on an island (isli and islj), membership in the commonwealth of nations (cwni and cwnj) and former eastern bloc (ebli and eblj). The data set includes 23 OECD countries for the period from 1995 to 2005, it is summarized in table 6.1.

Estimating this panel data set requires certain techniques to control for the effects of the countries and the years. Therefore, three specifications will be reported:

1. a pooled regression, where the panel data properties are ignored,
2. a least-squares dummy variable (LSDV) model with 23 dummies for the exporting countries, 23 dummies for the importing countries, and 11 dummies for the years (following e.g. Mátyás, 1997; Anderson and van Wincoop, 2003), and
3. a LSDV model with  $23 \times 22 = 506$  country pair dummies plus 11 year dummies (see Cheng and Wall, 1999; Baltagi, Egger, and Pfaffermayr, 2003, for a discussion of the adequate panel specification of gravity equations as well as the explanations in chapter 2).

Since we have constructed data for bilateral trade costs and multilateral resistances, the estimation of the standard gravity specification is not adequately based on the theory of

Anderson and van Wincoop (2003). One problem of estimating the theory-based gravity equation (5.1) directly is that trade costs are exogenous. Changes in policy variables like freedom of trade or membership in a group of countries like the EU do not affect export levels between two countries directly. However, they do affect trade costs between the two countries directly and changes in those bilateral trade costs affect bilateral trade volumes. Ignoring this endogeneity of trade costs may lead to biased estimates. Following the logic introduced in chapters 3 and 4, I also estimate the simultaneous equation model:

$$\text{exp}_{ij} = \alpha_0 + \alpha_{Y_i} \text{gdpi} + \alpha_{Y_j} \text{gdpj} + \alpha_{t_{ij}} \text{tij} + \alpha_{P_i} \text{pi} + \alpha_{P_j} \text{pj} + u_{ijt}, \quad (6.2)$$

$$\text{tij} = \beta_0 + \beta_X \text{exp}_{ij} + \sum_{k=1}^{16} \beta_k w_{ij}^k + v_{ijt}, \quad (6.3)$$

with the trade cost index  $\text{tij}$  introduced in 5.2.

As a reference, I first estimate the standard gravity equation (6.1) as a pooled regression, a country-year fixed effects and a country-pair-year fixed effects model. To study the impact of introducing the multilateral resistances into the equation system (6.2) and (6.3), I first estimate both equations simultaneously without multilateral resistances (or multilateral resistances assumed to be captured by the fixed-effects dummies); then, the second equation (6.2) with multilateral resistances; and finally, the third equation (6.2) with adjusted multilateral resistances using a common coefficient for  $\alpha_{P_i}$  and  $\alpha_{P_j}$ .

## 6.2 Empirical Results

Table 6.2 presents the results of estimating the standard gravity equation (6.1). The first two columns show the pooled regression that ignores the presence of panel data. Columns 3 and 4 show the country-year fixed effects model. Columns 5 and 6 show the country-pair-year fixed effects model. Columns 1, 3 and 5 display the reference case, where the effect of the multilateral resistances is assumed to be zero (or assumed to be completely captured by the fixed-effects, respectively). In the results of columns 2, 4 and 6 the effects

of the computed multilateral resistances are contained.

An analysis of the residuals shows that in the case of the country-pair-year fixed effects model, the residuals are closer to zero and that they are distributed rather independently from the endogenous variable  $\text{exp}_{ij}$  in comparison to the other specifications (figures 6.1 and 6.2). This indicates that the model with the country-pair-year fixed effects has the best properties to fit the model. The estimated coefficients of the multilateral resistances affect the exports negatively. The coefficient of the exporting country's multilateral resistances has an especially strong effect on trade flows. A negative value indicates that higher multilateral resistances tend to lead to lower trade activity with another, particular country. Note that this result is not directly in line with the theory by Anderson and van Wincoop (2003), described in section 5.1, which states that higher multilateral resistances of two trading countries enhance their bilateral trade flows because their bilateral trade costs are then relatively low (*ceteris paribus*). But the relation between multilateral and overall bilateral trade costs is not reflected in this reduced form specification at all. Controlling for the multilateral resistance index lowers the estimated effects of the other exogenous variables, as can be seen immediately from the comparison of the results in columns 5 and 6.

Again, note that the standard gravity model does not exactly reflect the theory presented above, as long as constructed data for bilateral trade costs and multilateral resistances are available. Table 6.3 shows the results of the theory-based simultaneous equation model with country-pair-year fixed effects using a 2SLS estimator.<sup>1</sup> As a reference case, column 1 displays the results without multilateral resistances. Column 2 displays the case that multilateral resistances enter the gravity equation (6.2) unrestrictedly. Note that the multilateral resistances of the importing country ( $p_j$ ) foster trade while the multilateral resistances of the exporting country ( $p_i$ ) lowers trade in this specification (upper part of table 6.3). Since the theory of Anderson and van Wincoop (2003) predicts that the effect

---

<sup>1</sup> Using a 3SLS estimator for the country-pair-year fixed effects specification with its 517 dummy variables was not feasible. Table 6.3 only presents the results of a country-pair-year fixed effects estimation because this specification has the best fit of the model compared to pooled regression and country-year fixed effects. The results of the other specifications are available on request.

6. Estimation with Multilateral Resistances

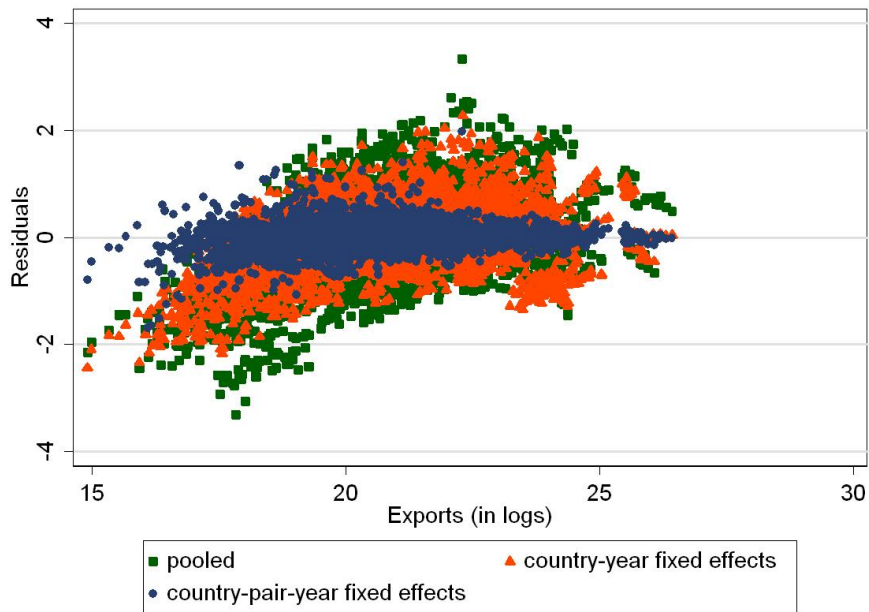
**Table 6.2:** Basic Case: Results of the Standard Gravity Specification.

	(1)	(2)	(3)	(4)	(5)	(6)
gdpi	0.864 (0.010)***	0.789 (0.010)***	0.368 (0.092)***	0.288 (0.094)***	0.386 (0.040)***	0.324 (0.039)***
gdpj	0.818 (0.010)***	0.817 (0.010)***	0.840 (0.106)***	0.808 (0.107)***	0.775 (0.040)***	0.763 (0.039)***
trfi	-0.734 (0.223)***	-0.816 (0.211)***	0.765 (0.218)***	0.798 (0.216)***	0.710 (0.079)***	0.732 (0.077)***
trfj	-0.378 (0.211)*	-0.511 (0.196)***	-0.020 (0.197)	-0.004 (0.198)	0.009 (0.079)	0.016 (0.077)
dist	-0.909 (0.015)***	-0.875 (0.015)***	-1.100 (0.024)***	-1.101 (0.024)***	—	—
lang	0.500 (0.045)***	0.452 (0.050)***	0.419 (0.041)***	0.414 (0.041)***	—	—
bor	0.322 (0.047)***	0.341 (0.046)***	0.115 (0.042)***	0.116 (0.042)***	—	—
cwni	-0.389 (0.036)***	-0.820 (0.044)***	-1.464 (0.146)***	-1.698 (0.150)***	—	—
cwnj	0.282 (0.043)***	0.281 (0.045)***	-0.024 (0.289)	0.038 (0.290)	—	—
ebli	-0.503 (0.055)***	-0.327 (0.054)***	-2.701 (0.284)***	-1.045 (0.350)***	—	—
eblj	-0.152 (0.054)***	-0.163 (0.051)***	0.426 (0.156)***	0.451 (0.154)***	—	—
isli	0.405 (0.031)***	0.514 (0.033)***	-0.013 (0.109)	1.147 (0.194)***	—	—
islj	0.018 (0.039)	0.006 (0.038)	-0.082 (0.117)	0.257 (0.193)	—	—
landli	0.109 (0.034)***	-0.031 (0.033)	-2.511 (0.299)***	-1.282 (0.333)***	—	—
landlj	-0.194 (0.036)***	-0.169 (0.034)***	-0.712 (0.190)***	-0.229 (0.291)	—	—
eui	0.088 (0.028)***	0.022 (0.027)	0.569 (0.079)***	0.397 (0.081)***	0.556 (0.027)***	0.399 (0.028)***
euj	0.066 (0.028)**	0.113 (0.028)***	0.084 (0.075)	0.032 (0.077)	0.087 (0.027)***	0.063 (0.028)**
exvol	-0.037 (0.010)***	-0.025 (0.008)***	-0.014 (0.007)**	-0.014 (0.007)**	-0.009 (0.004)**	-0.009 (0.004)**
pi	—	-2.853 (0.132)***	—	-4.668 (0.627)***	—	-4.251 (0.248)***
pj	—	0.014 (0.123)	—	-1.358 (0.595)**	—	-0.564 (0.255)**
Constant	-12.241 (1.213)***	-8.569 (1.159)***	-4.562 (3.803)	-1.325 (3.830)	-13.767 (1.547)***	-10.237 (1.525)***
Observations	4782	4782	4782	4782	4782	4782
$R^2$	0.81	0.83	0.89	0.90	0.98	0.98

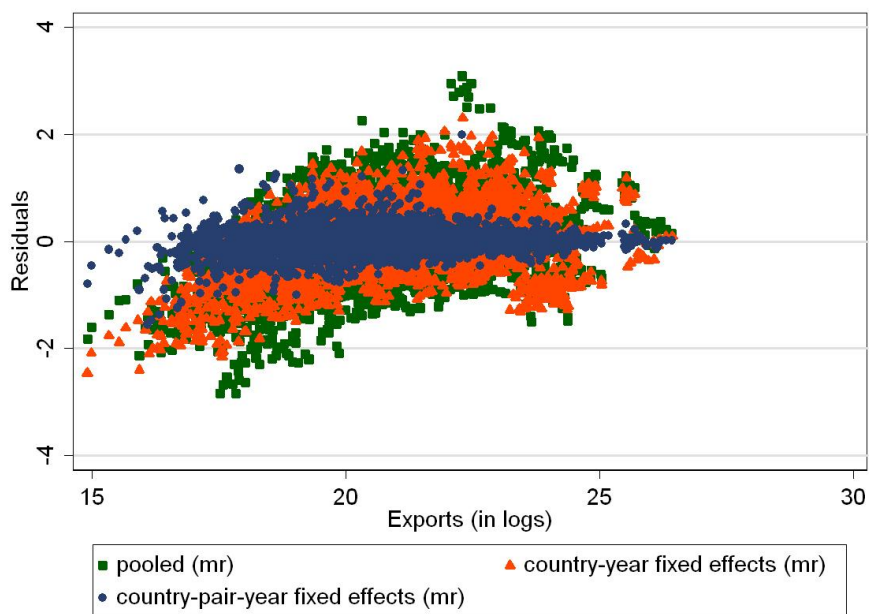
Robust standard errors in parentheses

\* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%





**Figure 6.1:** Residual-Analysis of the Standard Gravity Specification: Pooled Regression, Country-Year Fixed Effects and Country-pair-Year Fixed Effects (without Multilateral Resistances).



**Figure 6.2:** Residual-Analysis of the Standard Gravity Specification: Pooled Regression, Country-Year Fixed Effects and Country-pair-Year Fixed Effects (with Multilateral Resistances).

**Table 6.3:** Results of the Simultaneous Equation Model with Country-pair-Year Fixed Effects.

	(1)	(2)	(3)
gdpi	0.350 (0.030) <sup>***</sup>	0.281 (0.029) <sup>***</sup>	0.346 (0.030) <sup>***</sup>
gdpij	0.471 (0.030) <sup>***</sup>	0.560 (0.029) <sup>***</sup>	0.474 (0.030) <sup>***</sup>
tij	-6.761 (0.266) <sup>***</sup>	-7.367 (0.371) <sup>***</sup>	-7.379 (0.397) <sup>***</sup>
pi	—	-1.433 (0.262) <sup>***</sup>	—
pj	—	3.481 (0.271) <sup>***</sup>	—
pp	—	—	0.962 (0.256) <sup>***</sup>
Constant	4.023 (1.405) <sup>***</sup>	3.223 (1.317) <sup>**</sup>	3.895 (1.395) <sup>***</sup>
expij	-0.039 (0.004) <sup>***</sup>	-0.068 (0.003) <sup>***</sup>	-0.081 (0.003) <sup>***</sup>
trfi	-0.027 (0.008) <sup>***</sup>	-0.002 (0.007)	0.010 (0.007)
trfj	-0.048 (0.007) <sup>***</sup>	-0.037 (0.006) <sup>***</sup>	-0.032 (0.006) <sup>***</sup>
exvol	0.001 (0.000)	0.000 (0.000)	0.000 (0.000)
eui	-0.028 (0.004) <sup>***</sup>	-0.010 (0.003) <sup>***</sup>	-0.002 (0.003)
euj	-0.044 (0.003) <sup>***</sup>	-0.038 (0.002) <sup>***</sup>	-0.035 (0.002) <sup>***</sup>
Constant	1.929 (0.077) <sup>***</sup>	2.371 (0.059) <sup>***</sup>	2.563 (0.061) <sup>***</sup>
Obs.	4782	4782	4782

Standard errors in parentheses

\* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%

of trade costs  $t_{ij}$  on exports must be seen in relation to multilateral resistances  $p_i$  and  $p_j$ , we would expect a positive sign for both coefficients and not only for the import country coefficient.

To adjust the empirical model to the theory of equation (5.1) where multilateral resistances have the same coefficient, I comprise  $p_i$  and  $p_j$  to  $p_{ij} = p_i \cdot p_j$ . The results of this restricted model are shown in columns (3). Here, the coefficient of the multilateral resistances' product has a highly significant impact on the exports. If the product of multilateral resistances (the trade barriers of two certain countries to all countries) increases and everything else (especially the trade costs between the two countries) is kept constant, exports between these two certain countries increase because it becomes relatively more expensive for both countries to trade with the rest of the world than with each other. This is exactly the logic of the multilateral resistances introduced by Anderson and van Wincoop (2003). What happens to the coefficients of the remaining  $k$  exogenous variables which directly affect trade costs? Controlling directly for unrestricted or restricted multilateral resistances in columns 2 and 3 clearly reduces the estimated effects of these variables compared to the estimation without any multilateral resistances in column 1.<sup>2</sup>

### 6.3 Conclusion

Constructed data for trade costs and multilateral resistances make it possible to estimate the theory-based gravity equation by Anderson and van Wincoop (2003) directly. Since trade costs should be endogenous and also depend on the bilateral exports they explain, the estimation should be worked out using a simultaneous equation model. The results of this estimation show that the computed multilateral resistances have a significant influence on bilateral exports. It also appears that multilateral resistances clearly reduce the estimated effects of the remaining exogenous variables in the trade cost equation.

---

<sup>2</sup> As already noted in chapter 4, it is unsurprising that the coefficient of exchange rate volatility is not significant. Firms can hedge the exchange rate risk, and overall the exchange rate risk should not play an overly important role between the large OECD economies.

Again, it is important to note that the computed multilateral resistances are systematically biased downwards since not all countries of the world are included (see chapter 5). In a robustness check, 0.2 was added to the values of the multilateral resistances in order to increase these downward biased values. This transformation hardly affects the parameters of the model variables (except the parameters of the multilateral resistances themselves).

# Chapter 7

## Summary

Gravity Equations are the most used tool to explain the effects of trade costs on export flows between countries. The estimated coefficients of proxy variables for trade costs are frequently criticized as being too high. This is very crucial as long as these coefficients are broadly used to consult policymakers. Recent approaches have helped to improve the results by considering both theoretical and econometrical aspects. The main objective of this study was to make a contribution to this discussion by studying possible interactions between trade costs and exports.

**The First Part:** The introduction of a theory of endogenous trade costs and an empirical application.

The theoretical argument introduced in chapter 3 is that there is probably not an unidirectional effect of trade costs on exports, as it is described by the original gravity equation (1.1). It is argued that there is rather an interdependency between exports and trade costs since trade costs presumably bear fixed costs: The more actively two countries trade with each other, the lower the trade costs will be (per dollar of trade volume). Intuitive examples for these fixed costs which lead to declining average costs of trade can be found in transport and social infrastructures. To study the consequences of this dualism between exports and trade costs, iceberg (or average) trade costs in a theory-based gravity equation (Anderson and van Wincoop, 2003) are endogenized. Using a simple microeconomic model of the trade sector, it is shown that bilateral iceberg trade costs depend on cost components and – if there are economies of scale – on the bilateral export volume as well.

The consequence is a system of two equations: a gravity equation and a trade cost function. It is shown that the presence of economies of scale results in an upward bias of the estimated coefficients, i.e. using the traditional specification of the gravity equation, if these coefficients are interpreted as direct effects of trade costs on exports. This regards the immediate impact of a *ceteris paribus* change in a trade cost component like an import tariff on exports. But the duality of exports and trade costs implies that the affected exports change the trade costs again and the affected trade costs the exports,

and so on, like a domino effect. The resulting overall effect is implicitly measured by the traditional specification of the gravity equation. The direct effect and the overall effect are only equal, if the interaction between trade costs and exports works immediately and frictionlessly. It is necessary to re-interpret the outcomes of traditional gravity approaches as such overall effects.

A theory becomes especially useful if there is empirical evidence for it. This is proposed in chapter 4. Because the theoretical considerations suggest a dual system of a gravity equation and a trade cost equation, I estimate a simultaneous equation system using a three/two stage least-squares (3SLS/2SLS) estimator. Data are basically taken from OECD data bases covering the 30 OECD member countries (without respect to Chile, which joins the OECD not before January 2010). Since trade costs are not directly measurable, I use a micro-founded index of comprehensive trade costs (Novy, 2007). The theoretical implication is confirmed. The estimation strategy pursued provides a significantly negative effect of exports on trade costs (implying the presence of economies of scale) and clearly lower estimates for the parameters of interest.

**The Second Part:** A numerical solution for multilateral resistances and their empirical application with respect to trade cost endogeneity.

The presence of a measure for comprehensive bilateral trade costs (Novy, 2007) enables the computation of multilateral trade costs. The theory emphasized by Anderson and van Wincoop (2003) shows that it is not pure bilateral trade costs but rather bilateral trade costs relative to multilateral trade costs that matter in gravity equations. Theoretically, these multilateral trade costs are not pure (weighted) averages. They are a complex system of equations called multilateral resistances. Chapter 5 offers a convenient methodology to solve this equation system. The equation system displays that the unknown multilateral resistances (right-hand side) depend on all countries' multilateral resistances being multiplied by known country-specific parameters (left-hand side). I choose a common start value for the multilateral resistances on the right-hand side to compute values for the left-hand side. Then I use these computed values from the left-hand side on the

right-hand side to compute new values for the left-hand side. I repeat this until the left- and right-hand side vectors of the unknowns are equal. The adequate start value can be found by a search algorithm.

It is important to note that the data set does not contain the whole world which should lead to downward biases. However, the countries involved have such a high share of the global GDP that this bias should be in a tolerable scale. The resulting measures for the multilateral resistances appear to be plausible.

Chapter 6 extends the estimation procedure of chapter 4 to the computed multilateral resistance data. Multilateral resistances have a significant effect on both exports and trade costs. They again tend to lessen the estimated coefficients.

### **The Contribution of this Study**

What is the contribution of this study? First, a new theory of endogenous trade costs was provided which shows that iceberg trade costs are likely to depend on exports. An interaction between exports and trade costs (or the gravity function and a trade cost function) leads to a simultaneity problem. Second, this theory could be confirmed after estimating the gravity equation with a new strategy: a simultaneous equation system using a theory-based index to compensate for the directly immeasurable trade costs. Third, a methodology was developed to make the heretofore unknown index of multilateral resistances (Anderson and van Wincoop, 2003) visible. The consequence of the theoretical considerations and the use of constructed data for bilateral and multilateral trade costs is that the estimated direct effects of variables influencing trade volumes decrease. This could achieve more plausible and more reliable results from the gravity equation as the “workhorse for empirical studies” (Eichengreen and Irwin, 1998) of international trade.

### **The Limitations of this Study**

What are the limitations of this study? One source for criticism might be the usage of the index for overall trade costs by Novy (2007). This index is derived from the



gravity equation itself and, by definition, it depends on the trade flows. Of course, from a theoretical point of view that might be a problem. Yet, from an application-oriented point of view it is a useful measure for bilateral tariff equivalent overall trade costs and there is probably no other available data source for this variable.

Another limitation of the study is that the exogenous variables chosen to explain trade costs turned out to be inadequate (Sargan-test). Consequently, the estimates are biased. But this bias does not only affect the estimation strategy with the simultaneous equations pursued in this study. If the variables are not chosen adequately, this should also affect the traditional strategy to estimate gravity equations. The theory outlined in chapter 3 of this study suggests that appropriate variables for this purpose should reflect the cost components of trade costs rather than of country characteristics. It will be a task for future research to find adequate variables, but the requisite data are scarcely available.

A third limitation is related to the solution for the multilateral resistances. Computing them requires a very high density of data. Countries with insufficient data must be excluded from the computation and each excluded country biases the results downwards. For this reason, the computed values for the multilateral resistances must be interpreted carefully. However, a robustness check has shown that using higher values for multilateral resistances (with 0.2 added) in the different regression models does not noticeably change the estimates of the model variables. The problem could be solved in future research by taking the rest of the world as an additional, composite country into account (computed as the difference between whole global – or “whole world” – data and the data of the used countries).

### **Outlook**

This study has provided a new strategy to estimate gravity equations in a simultaneous model of two equations: a gravity equation and a trade cost function. The model was estimated using linear instrumental variable approaches. The data set consists of OECD countries, which are the largest economies of the world and where all countries trade with each other. Extending the strategy to a data set with more dissimilar economies and zero

trade flows requires non-linear IV estimators (see the discussion in chapter 2). Future endeavors could try to find and apply such estimators to make the strategy applicable to data sets with more heterogeneous countries and zero trade flows.

A major point of interest in this study was the phenomenon of trade costs. Despite the omnipresence of trade costs in real economic life, there are still many open questions and much to explore in future research. One aspect is to find explanatory variables and adequate functional forms to explain trade costs. This study has introduced one simple approach in a very general form, where trade costs depend on the underlying export volume and a theoretically unspecified set of cost factors. A more detailed theory-based insight into the functional form and the theoretical determinants of trade costs would be desirable.

Finally, I want to draw on Anderson and van Wincoop (2004), who criticize the shortage of data which reflect the cost factors of trade costs – especially data for political trade barriers. The availability of such data would help to gain more knowledge about the determinants of trade costs. This knowledge would improve the validity of the gravity equation's results and its implications for policymakers.

# Appendix A

## Derivation of the Gravity Equation by Anderson and van Wincoop (2003): Calculus

Consider the CES utility function of country  $j$

$$U_j = \left( \sum_i \varphi_i c_{ij}^{(\sigma-1)/\sigma} \right)^{\sigma/(\sigma-1)} \quad (3.1)$$

and the budget restriction of country  $j$

$$Y_j = \sum_i t_{ij} \cdot p_i \cdot c_{ij}. \quad (3.2)$$

Solving the Lagrange Function

$$\max_{c_{ij}} \mathcal{L} = \left( \sum_i \varphi_i c_{ij}^{(\sigma-1)/\sigma} \right)^{\sigma/(\sigma-1)} + \lambda \left( Y_j - \sum_i t_{ij} p_i c_{ij} \right) \quad (A.1)$$

yields the first order conditions

$$\frac{\partial \mathcal{L}}{\partial c_{ij}} = 0 = \varphi_i c_{ij}^{-1/\sigma} \cdot \left( \sum_i \varphi_i c_{ij}^{(\sigma-1)/\sigma} \right)^{1/(\sigma-1)} - \lambda t_{ij} p_i, \quad (\text{A.2})$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = 0 = Y_j - \sum_i t_{ij} p_i c_{ij}. \quad (\text{A.3})$$

Solving the first order condition (A.2) for  $c_{ij}$  yields

$$\begin{aligned} c_{ij} &= \left[ \frac{\lambda t_{ij} p_i}{\varphi_i \left( \sum_i \varphi_i c_{ij}^{(\sigma-1)/\sigma} \right)^{1/(\sigma-1)}} \right]^{-\sigma} \\ &= \underbrace{(\lambda)^{-\sigma}}_{\equiv \Lambda} \cdot (t_{ij} p_i)^{-\sigma} \varphi_i^\sigma \underbrace{\left( \sum_i \varphi_i c_{ij}^{(\sigma-1)/\sigma} \right)^{\sigma/(\sigma-1)}}_{=U_j} \\ &= \Lambda \cdot (t_{ij} p_i)^{-\sigma} \varphi_i^\sigma \cdot U_j. \end{aligned} \quad (\text{A.4})$$

Multiplying both sides of equation (A.4) with  $t_{ij} p_i$  and summing up over all countries  $i \in \{1, \dots, C\}$  yields country  $j$ 's expenditure function

$$\underbrace{\sum_i t_{ij} p_i c_{ij}}_{Y_j} = \Lambda \cdot U_j \cdot \sum_i (t_{ij} p_i)^{1-\sigma} \varphi_i^\sigma, \quad (\text{A.5})$$

$$\Rightarrow \Lambda = \frac{Y_j}{U_j \sum_i (t_{ij} p_i)^{1-\sigma} \varphi_i^\sigma}.$$

Inserting this solution for  $\Lambda$  back into equation (A.4) yields

$$\begin{aligned}
 c_{ij} &= \frac{Y_j (t_{ij} p_i)^{-\sigma} \varphi_i^\sigma}{\sum_i (t_{ij} p_i)^{1-\sigma} \varphi_i^\sigma} \\
 &= \frac{1}{t_{ij} p_i} Y_j \varphi_i^\sigma \cdot \frac{(t_{ij} p_i)^{1-\sigma}}{\sum_i (t_{ij} p_i)^{1-\sigma} \varphi_i^\sigma} \\
 &= \frac{1}{t_{ij} p_i} Y_j \varphi_i^\sigma \left[ \frac{t_{ij} p_i}{\left( \sum_i (t_{ij} p_i)^{1-\sigma} \varphi_i^\sigma \right)^{1/(1-\sigma)}} \right]^{1-\sigma} \\
 &= \frac{1}{t_{ij} p_i} Y_j \varphi_i^\sigma \left[ \frac{t_{ij} p_i}{P_j} \right]^{1-\sigma}. \tag{A.6}
 \end{aligned}$$

with the CES price index

$$P_j = \left( \sum_i (t_{ij} p_i)^{1-\sigma} \varphi_i^\sigma \right)^{1/(1-\sigma)}. \tag{A.7}$$

To achieve the gross import function, we multiply both sides of equation (A.6) with  $t_{ij} p_i$ :

$$X_{ij} = t_{ij} p_i c_{ij} = Y_j \varphi_i^\sigma \left( \frac{t_{ij} p_i}{P_j} \right)^{1-\sigma}. \tag{3.3}$$

Using equation (3.3) we can extend the budget restriction (3.2):

$$\begin{aligned}
 Y_i &= \sum_j X_{ij} \\
 &= \sum_j \varphi_i^\sigma \cdot \left( \frac{t_{ij} \cdot p_i}{P_j} \right)^{1-\sigma} \cdot Y_j \\
 &= \varphi_i^\sigma p_i^{1-\sigma} \cdot \sum_j \left( \frac{t_{ij}}{P_j} \right)^{1-\sigma} \cdot \underbrace{Y_j}_{=s_j \cdot Y_w} \\
 &= \varphi_i^\sigma p_i^{1-\sigma} \cdot Y_w \cdot \underbrace{\sum_j \left( \frac{t_{ij}}{P_j} \right)^{1-\sigma}}_{=\Pi_i^{1-\sigma}} \cdot s_j \\
 &= \varphi_i^\sigma p_i^{1-\sigma} \cdot Y_w \cdot \Pi_i^{1-\sigma}, \tag{3.5}
 \end{aligned}$$

with the multilateral resistance of country  $i$

$$\Pi_i \equiv \left( \sum_j \left( \frac{t_{ij}}{P_j} \right)^{1-\sigma} \cdot s_j \right)^{1/(1-\sigma)}. \quad (3.6)$$

Equation (3.5) can be solved for scaled prices:  $\varphi_i^\sigma p_i^{1-\sigma} = Y_i / (Y_w \cdot \Pi_i^{1-\sigma})$ . Inserting this solution in the CES price index (A.7) gives the multilateral resistance of country  $j$ :

$$P_j = \left( \sum_i \left( \frac{t_{ij}}{\Pi_i} \right)^{1-\sigma} \cdot s_i \right)^{1/(1-\sigma)}. \quad (3.7)$$

Inserting the scaled prices  $\varphi_i^\sigma p_i^{1-\sigma}$  into the gross import function (3.3) yields the gravity function with gross exports

$$X_{ij} = \frac{Y_i \cdot Y_j}{Y_w} \cdot \left( \frac{t_{ij}}{\Pi_i \cdot P_j} \right)^{1-\sigma}, \quad (3.8)$$

or, after dividing bot sides by  $t_{ij}$ , the net gravity function

$$X_{ij}^0 = \frac{Y_i \cdot Y_j}{Y_w} \cdot t^{-\sigma} \cdot (\Pi_i \cdot P_j)^{\sigma-1}. \quad (3.9)$$

# Appendix B

## Derivation of the Index for Bilateral Trade Costs

Starting point to derive the trade cost index suggested by Novy (2007) is the Anderson and van Wincoop (2003) gravity equation:

$$X_{ij} = \frac{Y_i \cdot Y_j}{Y_w} \cdot \left( \frac{t_{ij}}{\Pi_i \cdot P_j} \right)^{1-\sigma}. \quad (3.8)$$

Applying this equations for the intra-national trade of country  $i$ , meaning the exports from country  $i$  to country  $i$  itself, yields:

$$X_{ii} = \frac{Y_i \cdot Y_i}{Y_w} \cdot \left( \frac{t_{ii}}{\Pi_i \cdot P_i} \right)^{1-\sigma}. \quad (B.1)$$

Solving equation (B.1) for the product of the multilateral resistances yields:

$$\Pi_i P_i = \left( \frac{X_{ii} Y_w}{Y_i Y_i} \right)^{1/(\sigma-1)} \cdot t_{ii}. \quad (B.2)$$

Now, we define a bidirectional gravity equation that is defined as the product of the two bilateral export flows between countries  $i$  and  $j$ :

$$X_{ij}X_{ji} = \left(\frac{Y_i \cdot Y_j}{Y_w}\right)^2 \cdot \left(\frac{t_{ij}t_{ji}}{\Pi_i P_i \cdot \Pi_j P_j}\right)^{1-\sigma}. \quad (\text{B.3})$$

Rearranging the last multiplier of equation (B.3) and using the solution for the product of multilateral resistances of country  $i$  (B.2) and its respective formulation for  $\Pi_j P_j$  in equation (B.3) gives:

$$X_{ij}X_{ji} = X_{ii}X_{jj} \cdot \left(\frac{t_{ii}t_{jj}}{t_{ij}t_{ji}}\right)^{\sigma-1}. \quad (\text{B.4})$$

The next step is disentangling trade costs and export values from equation (B.4):

$$\frac{t_{ij}t_{ji}}{t_{ii}t_{jj}} = \left(\frac{X_{ii}X_{jj}}{X_{ij}X_{ji}}\right)^{1/(\sigma-1)}. \quad (\text{B.5})$$

Taking the square root of the left-hand side of equation (B.5) is the geometric mean of the bilateral trade:

$$\bar{t}_{ij} = \sqrt{\frac{t_{ij}t_{ji}}{t_{ii}t_{jj}}} = \left(\frac{X_{ii}X_{jj}}{X_{ij}X_{ji}}\right)^{\frac{1}{2(\sigma-1)}}. \quad (\text{B.6})$$

It measures the bilateral trade costs relative to the intra-country trade costs which are thus implicitly set 1. Thus, it must be interpreted as the international component of trade costs because trade costs inside a country are faded out. As a geometric mean it is a symmetric measure of bilateral trade costs between country  $i$  and  $j$ . It can also be denoted as a tariff equivalent ( $\tau_{ij}$ ) after subtracting 1 from both sides of equation (B.6).



# Appendix C

## Multilateral Resistances as a Linear Equation System

In a set of many countries  $(1, \dots, i, j, \dots, C)$ , the linearized equation system in analogy to (5.5) gets the following structure:

$$\mathbf{1} = \mathbf{B} \cdot \mathbf{z}$$

with the left-hand side vector of dimension  $C \times 1$

$$\mathbf{1}^\top = (1^1, \dots, 1^C),$$

the vector of the unknowns  $z_{ij} = 1/(\mathcal{P}_i \mathcal{P}_j)$  of dimension  $\frac{C \cdot (C+1)}{2} \times 1$

$$\mathbf{z}^\top = (z_{11}, z_{12}, \dots, z_{1C}, z_{22}, \dots, z_{2C}, z_{33}, \dots, z_{CC}),$$

and the coefficient matrix of the dimension  $C \times \frac{C \cdot (C+1)}{2}$ :

$$\mathbf{B} = \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1C} & & & & & \\ & b_{21} & & & b_{22} & \cdots & b_{2C} & & & \\ & & \ddots & & & \ddots & & b_{33} & \cdots & \\ & & & b_{C1} & & & b_{C2} & & \ddots & \cdots & b_{CC} \end{pmatrix}.$$

This linear equation system consists of more unknowns than equations since  $\frac{C \cdot (C+1)}{2} > C$ . Thus, this equation system is underdetermined and an underdetermined linear equation system has usually infinitely many solutions. Therefore, it is not tractable to pursue the linearization of the polynomial equation system (5.5).

# Bibliography

- AITKEN, N. D. (1973): “The Effect of EEC and EFTA on European Trade: A Temporal Cross-Section Analysis,” *American Economic Review*, 63(5), 881–891.
- AITKEN, N. D., AND R. S. OBUTELEWICZ (1976): “A Cross-Sectional Study of EEC Trade with the Association of African Countries,” *Review of Economics and Statistics*, 58(4), 425–433.
- ANDERSON, J. E. (1979): “A Theoretical Foundation for the Gravity Equation,” *American Economic Review*, 69(1), 106–116.
- ANDERSON, J. E., AND E. VAN WINCOOP (2003): “Gravity with Gravitas: A solution to the Border Puzzle,” *American Economic Review*, 93(1), 170–192.
- (2004): “Trade Costs,” *Journal of Economic Literature*, 42(3), 691–751.
- ARMINGTON, P. S. (1969): “A Theory of Demand Distinguished by Place of Production,” *IMF Staff Papers*, 16(1), 159–178.
- BABETSKAIA-KUKHARCHUK, O., AND M. MAUREL (2004): “Russia’s Accession to the WTO: The Potential for Trade Increase,” *Journal of Comparative Economics*, 32(4), 680–699.
- BAIER, S. L., AND J. H. BERGSTRAND (2001): “The Growth of World Trade: Tariffs, Transport Costs, and Income Similarity,” *Journal of International Economics*, 53(1), 1–27.
- (2007): “Do Free Trade Agreements Actually Increase Members’ International Trade?,” *Journal of International Economics*, 71(1), 72–95.

## BIBLIOGRAPHY

---

- (2009): “Bonus Vetus OLS: A Simple Method for Approximating International Trade-Cost Effects Using the Gravity Equation,” *Journal of International Economics*, 77(1), 77–85.
- BALDWIN, R., AND D. TAGLIONI (2006): “Gravity for Dummies and Dummies for Gravity Equations,” NBER Working Papers 12516, National Bureau of Economic Research, Inc.
- BALDWIN, R. E. (2006): “The Euro’s Trade Effect,” Working Paper Series 594, European Central Bank, Frankfurt.
- BALTAGI, B. H. (2005): *Econometric Analysis of Panel Data*. Wiley, New York, 3rd edn.
- BALTAGI, B. H., P. EGGER, AND M. PFAFFERMAYR (2003): “A Generalized Design for Bilateral Trade Flow Models,” *Economics Letters*, 80(3), 391–397.
- BEHRENS, K., G. MION, Y. MURATA, AND J. SUEDEKUM (2008): “Trade, Wages, and Productivity,” IZA Discussion Papers 3682, Institute for the Study of Labor (IZA).
- BERGSTRAND, J. H. (1985): “The Gravity Equation in International Trade: Some Microeconomic Foundations and Empirical Evidence,” *Review of Economics and Statistics*, 67(3), 474–481.
- (1989): “The Generalised Gravity Equation, Monopolistic Competition, and the Factor-Proportions Theory in International Trade,” *Review of Economics and Statistics*, 71(1), 143–153.
- (1990): “The Heckscher-Ohlin-Samuelson Model, the Linder Hypothesis and the Determinants of Bilateral Intra-Industry Trade,” *Economic Journal*, 100(403), 1216–1229.
- BRUN, J.-F., C. CARRÈRE, P. GUILLAUMONT, AND J. DE MELO (2005): “Has Distance Died? Evidence from a Panel Gravity Model,” *World Bank Economic Review*, 19(1), 99–120.

## BIBLIOGRAPHY

---

- BUCH, C. M., J. KLEINERT, AND F. TOUBAL (2004): “The Distance Puzzle: On the Interpretation of the Distance Coefficient in Gravity Equations,” *Economics Letters*, 83(3), 293–298.
- CHANEY, T. (2008): “Distorted Gravity: The Intensive and Extensive Margins of International Trade,” *American Economic Review*, 98(4), 1707–1721.
- CHENG, I.-H., AND H. J. WALL (1999): “Controlling for Heterogeneity in Gravity Models of Trade and Integration,” Working Paper 1999-010E, Federal Reserve Bank of St. Louis.
- CHRISTENSEN, R. (1987): *Plain Answers to Complex Questions. The Theory of Linear Models*, Springer Texts in Statistics. Springer, New York.
- DEARDORFF, A. V. (1984): “Testing Trade Theories and Predicting Trade Flows,” in *Handbook of International Economics*, ed. by R. W. Jones, and P. B. Kenen, vol. 1, pp. 467–517. North-Holland Publishing Company, New York.
- (1998): “Determinants of Bilateral Trade: Does Gravity Work in a Neoclassical Work?,” in *The Regionalization of the World Economy*, ed. by J. A. Frankel, pp. 7–22. The University of Chicago Press, Chicago.
- DEBAERE, P. (2005): “Monopolistic Competition and Trade, Revisted: Testing the Model without Testing for Gravity,” *Journal of International Economics*, 66(1), 249–266.
- DISDIER, A.-C., AND K. HEAD (2008): “The Puzzling Persistence of the Distance Effect on Bilateral Trade,” *Review of Economics and Statistics*, 90(1), 37–48.
- EATON, J., AND S. KORTUM (2002): “Technology, Geography, and Trade,” *Econometrica*, 70(5), 1741–1779.
- EGGER, P. (2000): “A Note on the Proper Econometric Specification of the Gravity Equation,” *Economics Letters*, 66(1), 25–31.
- (2005): “Alternative Techniques for Estimation of Cross-Section Gravity Models,” *Review of International Economics*, 13(5), 881–891.

## BIBLIOGRAPHY

---

- EGGER, P., AND M. PFAFFERMAYR (2003): “The Proper Panel Econometric Specification of the Gravity Equation: A Three-Way Model with Bilateral Interaction Effects,” *Empirical Economics*, 28(3), 571–580.
- EICHENGREEN, B. J., AND D. A. IRWIN (1998): “The Role of History in Bilateral Trade Flows,” in *The Regionalization of the World Economy*, ed. by J. A. Frankel, pp. 33–62. The University of Chicago Press, Chicago.
- EVENETT, S. J., AND W. KELLER (2002): “On Theories Explaining the Success of the Gravity Equation,” *Journal of Political Economics*, 110(2), 281–316.
- FEENSTRA, R. C. (2004): *Advanced International Trade*. Princeton University Press, Princeton.
- FELBERMAYR, G. J., AND W. KOHLER (2006): “Exploring the Intensive and Extensive Margins of World Trade,” *Review of World Economics*, 142(4), 642–674.
- FRANKEL, J. A., AND D. ROMER (1999): “Does Trade Cause Growth?,” *American Economic Review*, 89(3), 379–399.
- FUCHS, M., AND K. WOHLRABE (2008): “Institutions, Trade and Intergration: What Can be Expected within the Enlarged EU?,” Dresden Discussion Papers in Economics 16/08, Technische Universität Dresden.
- GREENE, W. H. (2000): *Econometric Analysis*. Prentice Hall International, 4th edn.
- GROSSMAN, G. M. (1998): “Comment on A. V. Deardorff, Determinants of International Trade: Does Gravity Work in a Neoclassical World?,” in *The Regionalization of the World Economy*, ed. by J. A. Frankel, pp. 29–31. The University of Chicago Press, Chicago.
- HEAD, K., AND T. MAYER (2004): “The Empirics of Agglomeration and Trade,” in *Handbook of Regional and Urban Economics*, ed. by J. V. Henderson, and J.-F. Thisse, vol. 4, chap. 59, pp. 2609–2669. Elsevier.

## BIBLIOGRAPHY

---

- HEAD, K., AND J. RIES (2001): “Increasing Returns Versus National Product Differentiation as an Explanation for the Pattern of U.S.-Canada Trade,” *American Economic Review*, 91(4), 858–876.
- HELPMAN, E. (1987): “Imperfect Competition and International Trade: Evidence from Fourteen Industrial Countries,” *Journal of the Japanese and International Economies*, 1, 62–81.
- HELPMAN, E., AND P. R. KRUGMAN (1985): *Market Structure and International Trade: Increasing Returns, Imperfect Competition and the International Economy*. MIT Press, Cambridge.
- HELPMAN, E., M. J. MELITZ, AND Y. RUBINSTEIN (2008): “Estimating Trade Flows: Trading Partners and Trading Volumes,” *Quarterly Journal of Economics*, 123(2), 441–487.
- HERZ, B., AND M. WAGNER (2007): “Do the World Trade Organization and the Generalized System of Preferences Foster Bilateral Trade?,” Discussion Paper 01-07, Universität Bayreuth.
- HUMMELS, D. L. (1999): “Towards a Geography of Trade Costs,” Working Paper, Purdue University.
- (2007): “Transportation Costs and International Trade in the Second Era of Globalization,” *Journal of Economic Perspectives*, 21(3), 131–154.
- HUMMELS, D. L., AND J. LEVINSOHN (1995): “Monopolistic Competition and International Trade: Reinterpreting the Evidence,” *Quarterly Journal of Economics*, 110(3), 799–836.
- HUMMELS, D. L., V. LUGOVSKYY, AND A. SKIBA (2009): “The Trade Reducing Effects of Market Power in International Shipping,” *Journal of Development Economics*, 89(1), 84–97.
- HUMMELS, D. L., AND A. SKIBA (2002): “A Virtuous Circle? Regional Tariff Liberalization and Scale Economies in Transport,” Working Paper, Purdue University.

## BIBLIOGRAPHY

---

- JACKS, D. S., C. M. MEISSNER, AND D. NOVY (2008): “Trade Costs, 1870–2000,” *American Economic Review*, 98(2), 529–534.
- JONES, R. W., AND H. KIERZKOWSKI (1990): “The Role of Services in Production and International Trade: A Theoretical Framework,” in *The Political Economy of International Trade, Essays in Honor of Robert E. Baldwin*, ed. by R. W. Jones, and A. O. Krueger, pp. 31–48. Blackwell, Oxford.
- KRUGMAN, P. R. (1979): “Increasing Returns, Monopolistic Competition, and International Trade,” *Journal of International Economics*, 9(4), 469–479.
- KRUGMAN, P. R., AND M. OBSTFELD (2009): *International Economics*, The Addison-Wesley Series in Economics. Pearson/Addison-Wesley, Boston, 8th edn.
- LIMAO, N., AND A. J. VENABLES (2001): “Infrastructure, Geographical Disadvantage, Transport Costs, and Trade,” *World Bank Economic Review*, 15(3), 451–479.
- LINNEMANN, H. (1966): *An Econometric Study of International Trade Flows*, Contributions to Economic Analysis 42. North-Holland Publishing Company, Amsterdam.
- MCCALLUM, J. (1995): “National Borders Matter: Canada-U.S. Regional Trade Patterns,” *American Economic Review*, 85(3), 615–623.
- MELITZ, M. J. (2003): “The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity,” *Econometrica*, 71(6), 1695–1725.
- MELITZ, M. J., AND G. I. P. OTTAVIANI (2008): “Market Size, Trade, and Productivity,” *Review of Economic Studies*, 75(1), 295–316.
- MÁTYÁS, L. (1997): “Proper Econometric Specification of the Gravity Model,” *The World Economy*, 20(3), 363–368.
- (1998): “The Gravity Model: Some Econometric Considerations,” *The World Economy*, 21(3), 397–401.
- NITSCH, V. (2002): “Honey, I Shrunk the Currency Union Effect on Trade,” *The World Economy*, 25(4), 457–474.



## BIBLIOGRAPHY

---

- (2007a): “Does G7/G8 Promote Trade?,” *Economics Letters*, 94, 136–140.
- (2007b): “State Visits and International Trade,” *The World Economy*, 30(12), 1797–1816.
- NITSCH, V., AND D. SCHUMACHER (2004): “Terrorism and International Trade: An Empirical Investigation,” *European Journal of Political Economy*, 20(2), 423–433.
- NOVY, D. (2007): “Gravity Redux: Measuring International Trade Costs with Panel Data,” Warwick Economic Research Papers No. 861.
- OBSTFELD, M., AND K. S. ROGOFF (2001): “The Six Major Puzzles in International Macroeconomics. Is There a Common Cause?,” in *NBER Macroeconomics Annual 2000*, ed. by B. S. Bernanke, and K. Rogoff, pp. 339–390.
- OLPER, A., AND V. RAIMONDI (2009): “Patterns and Determinants of International Trade Costs in the Food Industry,” *Journal of Agricultural Economics*, 60(2), 273–297.
- PÖYHÖNEN, P. (1963): “A Tentative Model for the Volume of Trade between Countries,” *Weltwirtschaftliches Archiv*, 90(1), 93–99.
- REDDING, S. J., AND A. J. VENABLES (2002): “The Economics of Isolation and Distance,” *Nordic Journal of Political Economy*, 28, 93–108.
- ROSE, A. K. (2000): “One Money, One Market: The Effect of Common Currencies on Trade,” *Economic Policy*, 15(30), 9–45.
- (2004): “Do We Really Know that the WTO Increases Trade?,” *American Economic Review*, 94(1), 98–114.
- ROSE, A. K., AND M. M. SPIEGEL (2009): “The Olympic Effect,” NBER Working Papers 14854, National Bureau of Economic Research, Inc.
- ROSE, A. K., AND E. VAN WINCOOP (2001): “National Money as a Barrier to Trade: The Real Case for Currency Union,” *American Economic Review*, 91(2), 386–390.
- SAMUELSON, P. A. (1954): “The Transfer Problem and Transport Costs: Analysis of Effects of Trade Impediments,” *Economic Journal*, 64(254), 264–289.

## BIBLIOGRAPHY

---

- SILVA, J. M. C. S., AND S. TENREYRO (2006): “The Log of Gravity,” *Review of Economics and Statistics*, 88(4), 641–658.
- SUBRAMANIAN, A., AND S.-J. WEI (2007): “The WTO Promotes Trade, Strongly But Unevenly,” *Journal of International Economics*, 72(1), 151–175.
- TINBERGEN, J. (1962): *Shaping the World Economy*. Twentieth Century Fund, New York.
- TOMZ, M., J. L. GOLDSTEIN, AND D. RIVERS (2007): “Do We Really Know That the WTO Increases Trade? Comment,” *American Economic Review*, 97(5), 2005–2018.
- WEI, S.-J. (1996): “Intra-National versus International Trade: How Stubborn are Nations in Global Integration?,” NBER Working Paper 5531.