Copenhagen Business School
Department of Economics

Master Thesis
MSc in Applied Economics and Finance

The Theoretical Foundation of Gravity Modeling:

What are the developments that have brought gravity modeling into mainstream economics?

Sarah Cathérine Starck

August 2012

Number of Pages: 71 Number of characters: 182.299

Master Thesis Supervisor
Professor Pascalis Raimondos-Møller
Department of Economics
Copenhagen Business School
Porcelænshaven 16A, 1. fl., 1.78
2000 Frederiksberg C
Denmark

Executive Summary

Gravity modeling is one of the most applied empirical methods to model and explain international trade flows. Besides the success gravity models experience in empirical research, a profound theoretical rationale of the model did not exist a decade ago. In its traditional form, the gravity equation puts bilateral trade flows in a positive relationship to the countries' gross domestic products and links it negatively to the distance between them. Alluding to a concept in physics, it intends to explain trade flows in relation to the entities' economic productivity or mass and the geographic distance parting them. Originally, the distance variable absorbed many different varieties of trade barriers. However, with the evolution of trade theory, research focused on a manifold of trade hindering and enhancing factors such as transport costs, tariffs, trade agreements, currency unions or cultural similarities.

This work gives an overview of the theoretical development in gravity modeling. It aims to offer insights on the theoretical foundation within gravity modeling. The research question this work is based on can be summarized as following: *The theoretical foundation of gravity modeling: What are the developments that have brought gravity modeling into mainstream economics?*

My findings indicate that during the last years a new research approach has been developed which concentrates on gravity model theory. The development of multilateral resistance terms by Anderson and van Wincoop (2003, 2004) allow a new interpretation of gravity modeling within international economics. Multilateral resistance represents the importing and exporting country's average trade resistance between itself and every possible trade partner that could potentially be traded with. Thus, multilateral resistance does not solely take bilateral trade barriers into account as before, but sets country-pair trade into a multilateral context. The incorporation of inward and outward resistance terms results in an in depth understanding of empirical phenomena such as the border puzzle which was introduced by McCallum (1995).

This suggests the structural gravity equation is a step towards a more comprehensive, holistic and context-sensitive analysis of international trade flows. Ensuring a general equilibrium setting which derives from the choice and comparison of opportunity costs, it features one of the main characteristics in international trade: a balanced state of markets around the world. This enriches a broad field within international economics, such as the modeling of migration, foreign direct investment or international capital movement. Due to its complexity in application, multilateral resistance is sometimes neglected in the empirical estimation. Reasons for this are often incomplete data because of missing observations or nonexistent industry specific information. Further research could focus on diverse assumptions and strive for a general theoretical justification which is more independent from the assumptions chosen.

Table of Contents

List of Tables and Figures III

List of Abbreviations IV

1 Introduction 1

2 The gravity model in international economics 3

2.1 Origins of the gravity model 3

2.2 Standard specification of the gravity equation 4

2.3 Research question 5

3 Theoretical foundations of the gravity model 6

3.1 Product differentiation by place of origin 7

3.2 Trade and monopolistic competition 12

3.3 Trade and differing factor endowments 15

3.4 Trade and differences in production technologies 20

3.5 Summary 23

4 Recent developments in the theoretical foundation 24

4.1 Motivation: The border puzzle 24

4.2 The structural gravity equation 26

4.2.1 Assumptions of multilateral resistance 27

4.2.2 Derivation of the structural gravity equation 30

4.2.3 Interpretation of multilateral resistance 34

4.3 Solving the border puzzle: Empirical results 38

4.4 Gravity modeling and international trade 42

4.4.1 General equilibrium and opportunity costs 43

4.4.2 Estimation of the incidence of trade costs 45

4.4.3 Discussion 46

4.5 Other prospects of gravity models 49

4.5.1 The intensive and extensive margin of trade 49

4.5.2 Demand functions other than CES preferences 50

5 Other applications in economics 52

5.1 Migration flows 52

5.2 Foreign direct investment 55

5.3 Portfolio investment 58

5.4 Currency unions 61

5.5 Trade agreements 62

6 Conclusion 65

7 Bibliography 68

List of Tables and Figures

Table 1: Estimation of the gravity equation by McCallum (1995) ………………………….. 25

Table 2: Estimation of McCallum's gravity equation by Anderson and van Wincoop (2003)……………………………………………………………….... 27

Table 3: Estimation of multilateral resistance under border barriers and borderless trade in the multi-country model by Anderson and van Wincoop (2003)……………………………………………………………….... 39

Tabel 4: Impact of border barriers on bilateral trade in the multi-country model by Anderson and van Wincoop (2003)………………………………………. 40

Figure 1: Overview of theoretical gravity approaches before 2003 ……………………...…..... 23

List of Abbreviations

BB	Border Barriers
NB	Borderless Trade
CES	Constant Elasticity of Substitution
c.i.f.	cost, insurance, freight
CHB	Constructed Home Bias
EU	European Union
EMU	European Monetary Union
FDI	Foreign Direct Investment
f.o.b	free on board
GDP	Gross Domestic Product
GSP	Generalized System of Preferences
NAFTA	North American Free Trade Agreement
OECD	Organization for Economic Co-operation and Development
OLS	Ordinary Least Squares
OVB	Omitted Variable Bias
ROW	Rest Of World
UNCTAD	United Nations Conference on Trade and Development
U.S.	United States of America
WTO	World Trade Organization

1 Introduction

During the last decades world trade has increased significantly. This rise in trade has been explained to a large extent by factors of globalization such as international policy changes or new technologies in transportation facilities. Changes in international policy refer to a decline in cross-border taxes and tariff quotas driven by regional and free trade agreements. Often, supranational institutions seem to gain influence. Countries try to keep trade barriers as low as possible or even abolish them by negotiating trade agreements and joining international organizations such as the world trade organization (WTO). Furthermore, an extensive advancement in the provision of infrastructure and enhanced transportation technologies allowed for falling transportation costs (Anderson and van Wincoop, 2004; Krugman and Obstfeld, 2006). As the World Development Report 2009 suggests, economic development is characterized by a threefold: *higher density, shorter distance and fewer division* (The International Bank for Reconstruction and Development, 2009). This refers to the concentration of economic production hubs, the increasing mobility aiming at minimizing distance to participate in economic opportunity and less economic division because of liberalizing trade policies.

With markets throughout the world moving closer together and being more dependent on each other, the modeling of trade flows between countries within the research field of international economics and trade has yield an empirically common used method of modeling and estimating trade flows. The gravity equation of trade sets the volume of trade between two countries proportionally in relation to their gross domestic products (GDP) and inversely refers to the distance between them. The reasoning behind this equation is straight forward: the bigger the two economies (measured by the size of their GDP), the more trade is expected to take place between them. Additionally, the further apart the two countries are, meaning the greater the distance between them, the smaller is the predicted trade (Head, 2003). Originally, distance was synonymous with the geographical distance between two countries, but later included many forms of possible trade barriers such as transportation costs, tariffs and taxes, contract, information or distribution costs.

Alongside the success of gravity modeling in empirical application, the theoretical foundation of gravity modeling was rather insufficient or even missing completely for a long time. During the last decade however, research has focused on a theoretical rationale. Different theoretical approaches have tried to solve the *model identification problem* (Evenett and Keller, 2002). This tries to identify the underlying theories and involved assumptions which account for the success of the gravity equation. There are a number of approaches from which gravity equations have been derived in the past: Ricardian, Heckscher-Ohlin and Helpman and Krugman models as well as other determinants that add to a possible explanation of bilateral trade flows, such as transportation or trade costs. They are based accordingly on differing opportunity costs, e.g. technological differences (Ricardian model), factor endowments (Heckscher-Ohlin model) or emphasize the importance of monopolistic competition and increasing returns to scale (Helpman and Krugman model).

A major finding of this work suggests that a milestone in the search for a theoretical foundation of gravity modeling was achieved by Anderson and van Wincoop. Their work published in 2003 and 2004 contributes first and foremost by including a *multilateral resistance term* in the gravity equation (Anderson and van Wincoop, 2003; 2004). The *multilateral resistance term* represents a reflection of the average trade resistance between a country and *all* of its other possible trading partners. The *bilateral* relation between the two trading countries no longer determines trade flows, but bilateral trade is dependent on *multilateral* resistance, meaning the dependence on all other trading partners of the exporting and importing country. Hence, multilateral resistance sets trade into a contextual framework, consequently founding a new theoretical underpinning. It reflects the essence of what international trade is about: a choice made upon different opportunity costs which then results in general equilibrium.

Anderson and van Wincoop's approach is state of the art from a theoretical perspective nowadays as many gravity models incorporate multilateral resistance terms. However, capturing multilateral resistance empirically remains a challenge. The quality of data is essential for the incorporation and eventually the measurability of multilateral resistance.

The aim of this work is to describe the theoretical development of gravity modeling from traditional to structural gravity models. It offers insight on the theoretical advancement of gravity modeling throughout the last decade. It is based on the following research question:

The theoretical foundation of gravity modeling: What are the developments that have brought gravity modeling into mainstream economics?

The missing rationale of the gravity equation is theoretically justified and well-grounded nowadays. Since international trade is primarily explained by three classic approaches - the Ricardian model, the Heckscher-Ohlin model and the Helpman and Krugman model - I will demonstrate the theoretical underpinning of traditional gravity modeling on the basis of these models. Each approach features different assumptions, which are specified and explained. The most recent milestone in gravity modeling is characterized by the multilateral resistance terms leading to structural gravity modeling. Thus, this work emphasizes on the assumptions, derivation and interpretation of the former. In addition, the newly acquired importance of structural gravity models is put into the context of international trade and economics.

This work is structured as follows. Chapter 2 introduces the origin of gravity models and its application in international economics. Chapter 3 presents the three classic approaches of explaining international trade relating to the gravity model. Reaching back to Anderson (1979) who developed a first theoretical foundation, the assumptions of monopolistic competition, differing factor endowments and differences in production technologies illustrate the diverse interpretation and set-up of the gravity equation. Subsequently, Chapter 4 describes Anderson and van Wincoop's multilateral resistance approach which was motivated by the border puzzle of McCallum (1995). This chapter derives and interprets the structural gravity equation. Chapter 5 indicates other research areas apart from international trade where gravity modeling is of importance. Migration flows, foreign direct investment and international portfolio capital movement illustrate representatively the variety of flow modeling where gravity equations and gradually multilateral resistance are applied. Finally, chapter 6 concludes.

2 The gravity model in international economics

The gravity equation has been used for decades in international trade. Its origin goes back to 1687 and it was first applied in the field of international trade in 1962. This chapter gives a short overview on its primary origin and its standard application in international trade. The section concludes with a more detailed description of my research question.

2.1 Origins of the gravity model

Gravity modeling has experienced a rising popularity over the years and is a commonly used method of modeling trade flows in international economics today. Its origin goes back to the *law of universal gravitation* in physics which was developed by Isaac Newton in 1687. The law of universal gravitation describes the gravitational force between two masses in relation to the distance that lies between these two masses (Newton, 1687), that is

$$F_{ij} = G\ \frac{M_i M_j}{d_{ij}^2}. \tag{1}$$

The gravitational force F_{ij} relates hereby proportional to the product of the two masses M_i and M_j and inversely proportional to the square of the distance d_{ij} that keeps the two masses apart from each other. The gravitational constant G is an empirically determined value.

This relationship is applicable to any context where the modeling of flows or movements is demanded. The gravity model in international economics puts the law of universal gravity into an economic context. Assuming that the trade volume between country i and j is X_{ij} and furthermore economic masses defined as Y_i and Y_j, trade flows can be characterized by

$$X_{ij} = A \frac{Y_i \, Y_j}{D_{ij}}. \tag{2}$$

This is a strongly simplified version of the gravity equation applied in international trade. Often, X_{ij} is interpreted as the export volume from country i to j, the economical masses are represented by each country's GDP and distance is the geographical distance between the two countries' capital cities (Head, 2003).

The extraordinarily high goodness of fit is one of the reasons which accounts for the empirical success of the gravity equation. Often, explanatory power shows values of 80 per cent and above (McCallum, 1995). Former reason and a relatively simple adaptation of gravity equations to any kind of flow modeling is why research has used gravity modeling on a diverse scope of topics despite its lack of theoretical foundation for many years. Within the last years, especially foreign direct investment, equity flows and migration have been at the attention of research (Head and Ries, 2008; Portes and Rey, 2005; Grogger and Hanson, 2011).

2.2 Standard specification of the gravity equation

The gravity equation was first applied to international trade flows by Tinbergen in 1962. He assumed the following relationship

$$X_{ij} = A \frac{Y_i^{\alpha} Y_j^{\beta}}{D_{ij}^{\gamma}}. \tag{3}$$

Tinbergen notes that direct proportionality between explanatory variable and the variable to be explained is not necessarily implied. The exponents α, β and γ can therefore take values different than 1. They refer to the elasticity of the exporting country's GDP (α), the elasticity of the importing country's GDP (β) and the elasticity of distance (γ). In the case of $\alpha = \beta = 1$ and $\gamma = 2$, this will correspond to the universal gravitation equation of Isaac Newton.

Tinbergen (1962) justified the incorporation of both countries' economical masses by stating that the amount of exports country i is capable to supply depends in the first place on its economic size (Y_i). Equally, the amount of goods sold to the importing country j also depends on its purchasing power/ income (Y_j). Economic size is often defined as GDP, gross national product, income per capita or the country's population size. On the one hand, this enables conclusion on the production capacity of each country. On the other hand, it indicates the market potential for sales of goods to each country. Taken together, economic size variables characterize the supply as well as the demand force affecting each country's market. In the end, these forces are essential when determining trade flows.

Distance is defined as either the geographical distance between the economic hubs of the trading partners or distance between capital cities measured in nautical or land miles. Effects are comparable with the function of a tax, since trade barriers are expensive to overcome and are linked to transportation costs. This stems part of the trade and usually results in the decline of equilibrium trade flows. Tinbergen states that "The factor of distance may also stand for an index of information about export markets." (Tinbergen, 1962, p. 263). Hence, distance is a proxy for various factors that can influence trade such as transportation costs, time elapsed during shipment, synchronization costs, communication costs, transaction costs or cultural distance (Head, 2003). Since it is difficult to measure these factors and they are subject to high complexity, geographical distance is often used as an approximation to these costs. Factors encouraging trade can be the same cultural heritage or similar political systems. During the years, research has specified a multitude of factors expected to affect trade. Current topics of discussion are whether countries in the same currency union, facing the same borders or being members of the WTO have advantages in trading (Rose and van Wincoop, 2001; McCallum, 1995; Rose, 2004).

Gravity models are usually estimated with ordinary least squares (OLS) regression analysis. By taking the natural logarithm of equation (3) and adding an error term ε_{ij}, one obtains a linear relationship which allows to interpret coefficients as elasticity

$$log(X_{ij}) = log\,A + \alpha\, log(Y_i) + \beta\, log(Y_j) - \gamma\, log(D_{ij}) + \varepsilon_{ij}\ . \qquad (4)$$

The coefficient can be interpreted as follows: if the exporting country's GDP (Y_i) increases by 1 per cent, export volume will increase by α per cent everything else held constant. Likewise, if the distance between countries *i* and *j* increases by 1 per cent, then trade flows will decrease by γ per cent ceteris paribus. One assumes that the error term ε_{ij} is independent and log-normally distributed.

2.3 Research question

The underlying research question of this work is:

The theoretical foundation of gravity modeling: What are the developments that have brought gravity modeling into mainstream economics?

My thesis aims at explaining the derivation, incorporation and interpretation of multilateral resistance which contributed to a new theoretical underpinning of gravity modeling. Likewise, light is shed on its importance and application in international economics. The work focuses on assumptions associated with this new foundation and will be delimited from previous theoretical foundations. My approach is to review existing literature and to subtract the developments that led to the popularity of structural gravity. Therewith, my contribution is to explain developments within the foundation of gravity modeling from a theoretical scientific perspective.

I review four main concepts of gravity modeling that endow the reader with background knowledge enabling a more profound and insightful understanding of multilateral resistance. These four concepts of traditional gravity modeling are represented by Anderson (1979), who was the first to develop a theoretical rationale for gravity explaining bilateral trade flows, Helpman and Krugman (1985), who analyze gravity under the assumption of monopolistic competition, and gravity in a Heckscher-Ohlin and Ricardian framework. With this traditional gravity understanding, the focus of my thesis then shifts on Anderson and van Wincoop's multilateral resistance. I subtract important features of multilateral resistance, discuss its assumptions and explain what is so essentially new and innovative about it. Since gravity equations are not only used for modeling international trade flows, recent interest in other research areas broadens the gravity model's scope of application and justifies its existence in mainstream economics nowadays.

Eventually, my thesis aims at giving the reader a review of the theoretical developments within gravity modeling which results in an explanation of its popularity in research today.

3 Theoretical foundations of the gravity model

The classic models of international trade and their belonging theoretical approach of the gravity model are discussed below. The Ricardian model focuses on differences in technology and therewith heterogeneous production functions which results in a country's comparative cost advantage in the production of a particular good. The Ricardian model was first published in 1817 (Krugman and Obstfeld, 2006, p. 26). The approach developed by Heckscher and Ohlin assumes that every country holds a unique factor endowment which is either capital or labor intensive and thus specializes in the production of the either capital or labor abundant good. It was developed from 1919 to 1924 (Heckscher and Ohlin, 1991). It is often used to explain inter-industry trade. Helpman and Krugman (1985) developed an oppositional model explaining intra-industry trade. It is based on monopolistic competition and thus increasing returns to scale.

Prior to the three classic models of international trade, the microeconomic inspired theoretical foundation of Anderson (1979) is presented. This discussion is rather detailed due to the following elaboration by Anderson and van Wincoop (2003, 2004). It is elementary for the derivation of the multilateral resistance term which is reviewed in the following chapter.

3.1 Product differentiation by place of origin

Anderson (1979) was one of the first economists developing a sound theoretical foundation of the gravity model. In his model, products are differentiated by their place of origin, also called the Armington assumption. Armington (1969) distinguishes goods not only by their kind (e.g. merchandise, chemicals, wooden product etc.) but also by their place of production. Hence, the residence of the supplier is crucial for the characteristics of the product. Armington supposes that two goods of the same kind but originating from different countries are imperfect substitutes in demand. In the context of gravity modeling, this is feasible since the place of production is crucial with respect to the trade costs implied. Anderson's theoretical derivation of the gravity modeling develops gradually. His model is amplified by

i. the kind of good which is traded. All goods are tradable in the beginning, and then he distinguishes between tradable and non-tradable goods.
ii. the inclusion of distance.
iii. the number of tradable goods countries produce. Starting with a single differentiated product per country, Anderson expands the model framework to multiple traded goods classes.

Initially, Anderson's gravity model is based on two countries, i and j, each producing one differentiated good. These differentiated goods can be traded among the two countries. Countries have identical Cobb-Douglas preferences. This means that the share of income spent on tradable goods from country i by country j, denoted by b_i , is the same for both of the countries.[1] Prices are chosen to be constant at equilibrium value. In a frictionless world, i.e. no transportation costs or other trade barrier, exports X_{ij} from country i to country j are given by the importing country's income (Y_j) multiplied with the share of income spent on tradable goods from the exporting country i (b_i) which is

$$X_{ij} = b_i Y_j \,. \quad (5)$$

This is subject to the budget constraint

$$Y_i = b_i \sum_j Y_j \quad (6)$$

[1] Cobb-Douglas utility functions take the form $U(T, NT) = k\, x_1^{\alpha} x_2^{\beta}$ where k is a constant and x_1 and x_2 denote tradable and non-tradable goods. Cobb-Douglas utility functions always imply homothetic preferences. If preferences are homothetic, then $\alpha + \beta = 1$. That is, countries with a rising income invest the same proportion in tradable goods in relation to their total income due to a constant marginal rate of substitution (Varian, 2004, pp. 61-65; Pepall, Richards and Norman, 2008, pp. 78-81).

implying that income equals expenditure. Solving the equation for b_i and substituting into equation (5) yields the gravity equation in its simplest form which is often called the standard or traditional gravity equation

$$X_{ij} = \frac{Y_i Y_j}{\sum_j Y_j}. \tag{7}$$

Due to the restrictive assumptions, i.e. only two trading countries each producing one differentiated good, this scenario is unlikely. Striving for a more accurate representation of trade flows, Anderson relaxes these restrictions by assuming that each country produces a tradable and a non-tradable good. Countries maximize their utility function $u = u\ (g(traded\ goods), non - traded\ goods)$. Trade separability allows to maximize solely the extracted utility from traded goods. Due to homothetic preferences, this maximization is subject to traded goods prices only. The fraction of traded goods originating from country *i* on total expenditure for tradable goods by country *j* is θ_i which is comparable to b_i in the former case. Furthermore, Φ_j is the expenditure share on tradable goods of total expenditure of country *j*. Therewith, the exports of tradable goods from *i* to *j* are explained by the fraction spent on tradable goods by *j* (Φ_j) times the share spent on tradable goods originating from *i* (θ_i)

$$X_{ij} = \Phi_j \theta_i Y_j . \tag{8}$$

The new budget constraint follows from the assumption that income from tradable goods in country *i* equals expenditure on tradable goods in country *j* for products originating in *i*

$$Y_i \Phi_i = \theta_i \sum_j Y_j \Phi_j . \tag{9}$$

The budget constraint indicates that expenditure from country *j* spent on tradable goods from country *i* equals income of country *i*. The other way around, country *i* can only spend what it gained through international exports. Solving for θ_i and substituting into (8), yields

$$X_{ij} = \frac{Y_i \Phi_i Y_j \Phi_j}{\sum_j Y_j \Phi_j} = \frac{Y_i \Phi_i Y_j \Phi_j}{\sum_i \sum_j X_{ij}}. \tag{10}$$

Since Φ_i can be a function of income Y_i and population size N_i (or other factors), Anderson sets $\Phi_i = F\ (Y_i, N_i)$. Rewriting equation (10), adding an error term with $E(\ln U_{ij}) = 0$ and a constant *c*, this results in

$$X_{ij} = c\ \frac{Y_i\ F\ (Y_i, N_i) Y_j\ F\ (Y_j, N_j)}{\sum_j Y_j\ F\ (Y_j, N_j)}\ U_{ij}\ .^{2} \tag{11}$$

Assuming a linear form for *F( )*, and replacing the denominator with *K*, yields

$$X_{ij} = \left(\frac{c}{K}\right) Y_i^{\alpha_1} N_i^{\alpha_2} Y_j^{\beta_1} N_j^{\beta_2} U_{ij}\ . \tag{12}$$

K can be interpreted as world trade expenditure (or world trade income). Equation (11) shows that the distance variable is suppressed in comparison to equation (2). Reason is that Anderson (1979) assumes at this point of his derivation a frictionless world, where no trade barriers exist and thus distance is absent. By taking the natural logarithm, one obtains the gravity equation that can be estimated with OLS regression analysis

$$\ln X_{ij} = \ln\left(\frac{c}{K}\right) +\ \alpha_1 lnY_i + \alpha_2 lnN_i + \beta_1 lnY_j + \beta_2 lnN_j +\ lnU_{ij}\ . \tag{13}$$

However, the assumption of a borderless world is not very helpful when determining the influence of trade barriers. Subsequently, Anderson (1979) extends his model by transportation costs. Hereby, $X_{ij}\tau_{ij}$ is the export value landed in country *j* of products manufactured in country *i*. The transport factor is denoted by τ_{ij}. The fraction of traded goods from country *i* on total expenditure for tradable goods in country *j* (θ_i) is now dependent on all τ_{ij}'s from country *i* which results in $\theta_i(\tau_j)$ and hence

$$X_{ij}\tau_{ij} = \Phi_j\theta_i(\tau_j)Y_j \ \text{ or } \ X_{ij} = \frac{1}{\tau_{ij}}\Phi_j\theta_i(\tau_j)Y_j\ .^{3} \tag{14}$$

The budget constraint in a world with barriers is similar to (9) but in addition amended by the transport factor

$$Y_i\Phi_i =\ \theta_i(\tau_j)\sum_j \frac{1}{\tau_{ij}} Y_j\Phi_j\ . \tag{15}$$

When solving for $\theta_i(\tau_j) = \frac{Y_i\Phi_i}{\sum_j \frac{1}{\tau_{ij}} Y_j\Phi_j}$ and plugging this into (14), setting $D_{ij} \equiv \frac{1}{\tau_{ij}}$, this yields

$$X_{ij} = \frac{1}{\tau_{ij}}\Phi_j \frac{Y_i\Phi_i}{\sum_j \frac{1}{\tau_{ij}} Y_j\Phi_j} Y_j = D_{ij}\ \frac{Y_i\Phi_i\Phi_j Y_j}{\sum_j D_{ij} Y_j \Phi_j}. \tag{16}$$

[2] In Anderson (1979), a capital account factor is added. He notes that trade imbalance can occur due to long-term capital account transactions (Anderson, 1979, p. 109). However, one can think of this as any possible term multiplied to equation (11). In this illustration it is neglected for simplification reasons. A constant *c* summarizes all the possible constant terms that reside in $F\ (Y_i, N_i)$ and $F\ (Y_j, N_j)$.

[3] τ_j is a vector of all the transportation costs τ_{ij} for country *j*. τ_{ij} is an increasing function of distance $\tau_{ij} = f(d_{ij})$ and $f(0) = 1$.

By adding an error term and a constant, writing the denominator as K, substituting in the functions for Φ_i and Φ_j and taking the natural logarithm, one will obtain a function ready for OLS regression analysis estimation that is similar to equation (13).

Anderson (1979) derives a second scenario under the assumption of transportation costs, where there are not only two types of goods - one tradable and one non-tradable good - but countries produce multiple differentiated traded goods classified by the product class *k*. Export volume for the specific good *k* from *i* to *j* is given by

$$X_{ijk} = \frac{1}{\tau_{ijk}} \Phi_j \theta_{ik}(\tau_j) Y_j \, .$$

Aggregated trade flows over all goods classes are

$$X_{ij} = \sum_k X_{ijk} = \Phi_j Y_j \sum_k \frac{1}{\tau_{ijk}} \theta_{ik}(\tau_j) \, .$$

The budget constraint is

$$Y_i \Phi_i = \left(\sum_k \frac{1}{\tau_{ijk}} \theta_{ik}(\tau_j) \right) \sum_j Y_j \Phi_j \, .$$

Since transportation does not depend on the class of goods, i.e. *k*, and transport costs solely increase in distance the good is shipped with, one can express $\sum_k \frac{1}{\tau_{ijk}}$ as a function of distance $\frac{1}{f(d_{ij})}$ between country *i* and *j* and hence assume that

$$X_{ij} = \left(\sum_k \theta_{ik} \right) \Phi_j Y_j \frac{1}{f(d_{ij})} \tag{17}$$

$$Y_i \Phi_i = \left(\sum_k \theta_{ik} \right) \sum_j Y_j \Phi_j \frac{1}{f(d_{ij})} \, . \tag{18}$$

Solving equation (18) for $(\sum_k \theta_{ik}) = \frac{Y_i \Phi_i}{\sum_j Y_j \Phi_j \frac{1}{f(d_{ij})}}$, and substituting this into (17), the equation of aggregated trade flows, this results in Anderson's final gravity equation

$$X_{ij} = \frac{Y_i \Phi_i}{\sum_j Y_j \Phi_j \frac{1}{f(d_{ij})}} \frac{\Phi_j Y_j}{f(d_{ij})} = \frac{Y_i \Phi_i \Phi_j Y_j}{f(d_{ij})} \left[\sum_j Y_j \Phi_j \frac{1}{f(d_{ij})} \right]^{-1} \tag{19}$$

which can be rewritten as (20)

$$X_{ij} = \underbrace{\frac{Y_i \Phi_i \Phi_j Y_j}{\sum_j \Phi_j Y_j} \frac{1}{f(d_{ij})}}_{} \quad \underbrace{\left[\sum_j \frac{Y_j \Phi_j}{\sum_j \Phi_j Y_j} \frac{1}{f(d_{ij})} \right]^{-1}}_{} \, .$$

Economic distance from *i* to *j* in relation to world trade expenditure

Economic distance from *i* to all other potential trading partners in relation to world trade

Equation (20) is an extension of equation (19). The term outside the bracket of equation (19) is divided by the sum of world trade expenditure which is $\sum_j \Phi_j Y_j$. This cancels out by multiplying the term in brackets with $\sum_j \Phi_j Y_j$, since $\frac{1}{\sum_j \Phi_j Y_j} \left(\frac{1}{\sum_j \Phi_j Y_j}\right)^{-1}$ cancels out and is equal to 1. It is helpful to divide by the sum of world trade expenditures (world trade income), since this allows for a more coherent interpretation. Trade flows from country *i* to *j* depend "(…) on economic distance from *i* to *j* relative to a trade-weighted average of economic distance from *i* to all points in the system" (Anderson, 1979, p. 113). Thus,

i. Anderson (1979) states that the first term in the aggregated gravity equation symbolizes the economic distance between *i* and *j* in relation to world trade expenditure.
ii. The second term breaks down as follows: $\frac{Y_j \Phi_j}{\sum_j \Phi_j Y_j}$ indicates every country's expenditure on tradable goods in relation to world expenditure on traded goods. Since $\frac{1}{f(d_{ij})}$ indicates the *direction* of these flows, it fixes the economic distance from country *i* to each possible trading partner. The term $\sum_j \frac{Y_j \Phi_j}{\sum_j \Phi_j Y_j} \frac{1}{f(d_{ij})}$ then represents the sum of all countries' trade expenditure on tradable goods in relation to world expenditure and therewith implies *all* trading partners country *i* could possible trade with.
iii. Merging the two terms together describes the economic distance from *i* to *j*, in relation to the trade-adjusted economic distance of *i* to all other possible trading partners.

Therewith, Anderson enables estimating bilateral trade flows under the consideration of all other possible trade flows. Thus, the gravity equation can no longer be describes as an *absolute* construction of interacting bilateral trade partners but reflects rather a *comparative* analysis of a bilateral trade interaction in a multitude of possible trade flows.

As mentioned above, Anderson sets $\Phi_i = F\ (Y_i, N_i)$. There are many other variables of interest that can be crucial. Leamer (1974) adds resource endowments. Bergstrand (1985) includes price variables. He also assumes that products are differentiated by their place of origin. Under this scenario it is most likely that also prices vary from country to country. Anderson (1979) checks for prices in a separate derivation with a constant elasticity of substitution (CES) utility function. Derivation however is beyond the scope of this work.[4] CES preferences are discussed in the following chapter.

In summary, Anderson (1979) chooses a gradual modification of his theoretical derivation which enhances in complexity. The concept of this derivation relies on the countries' expenditure systems

[4] See Anderson, 1979, pp. 114-115 (Appendix) for derivation in the case of CES preferences.

meaning that expenditure equals income. The expenditure system constitutes the posterior derivation of the gravity equation. The gravity model tries to compare bilateral trade to a world trade average.

3.2 Trade and monopolistic competition

The Helpman and Krugman (1985) approach is the third of the classic models explaining international trade. Preceding theories focus on comparative advantages and therewith differences in production technology (Ricardian approach) and factor endowments (Heckscher-Ohlin approach). However, the two latter are not considered until the proceeding sections. Hence, ordering the theories in this way reflects the chronological publication of research papers in the past few years.

The Helpman and Krugman approach assumes increasing returns to scale and a state of monopolistic competition between firms. It is especially used to explain intra-industry trade which is the trade of products belonging to the same goods class (e.g. clothing). Products in this goods class distinguish themselves by small characteristics from other products in that class. Economies of scale incentivize the countries to not produce every differentiated good by themselves, but focus on a selection of goods and trade with other countries producing some other differentiated goods (Krugman and Obstfeld, 2006, p. 127).

Krugman (1979, 1980) bases his work on a Chamberlinian monopolistic competition resulting in imperfect competition.[5] To rule out differences in endowment and technology, Krugman assumes a one-factor model where the only factor of production is labor. Furthermore, there are identical preferences and technologies in the two trading countries. Wage rates in the two countries are equal due to symmetric productivity (same technology in both countries) and hence, prices for all products will be the same. When trading, each country experiences an effect as if its labor force increased compared to the state of no trading. Due to the expanding production and the assumption of economies of scale, unit production costs subsequently fall. Additionally, a rise in welfare occurs because the number of differentiated products available on the market increased. For the consumer, this results in a higher freedom of choice. Thus, by exploiting economies of scale, trade among industrialized countries - that is countries with similar demands - can be explained. The same accounts for trade among developing countries. Intra-industry trade aims at explaining trade flows between countries of similar needs and comparable economic development stages.

Since Krugman (1979, 1980) assumes identical preferences among countries, his findings are comparable to the Lindner hypothesis. Lindner (1961) proposes that countries similar to each other

[5] In the model of Chamberlinian monopolistic competition, firms have monopoly power since no identical substitute exists for the product sold. However, due to market entry of new firms, profits are driven to zero (Chamberlin, 1962).

focus on the same industries. Accordingly, they trade products belonging to the same goods class which however are differentiated. Each country's welfare increases since consumers benefit from the expanded product choice due to the *love of variety*.[6]

Helpman (1987) contributed with an empirical test on the characteristics of intra-industry trade. He calculates bilateral intra-industry trade shares for 14 industrialized countries for each year from 1970 to 1981, all of them being members of the Organization for Economic Co-operation and Development (OECD). Findings demonstrate a negative correlation between the share of bilateral intra-industry trade and dispersion of per capita income in a cross-country analysis. This supports Helpman's hypothesis stating the more similar factor compositions in different countries are to each other (i.e. per capita income levels), the larger will be the share of intra-industry trade. The same can be found when performing a time-series analysis to a within-group trade flow analysis. Helpman (1987) concludes that the relationship between the share of intra-industry trade and differences in income levels has weakened over time.

To exemplify a model with a closer link to gravity modeling, the work of Bergstrand (1985, 1989) is discussed. Likewise, it is based on monopolistic competition with differentiated products and economies of scale. Bergstrand (1989) reviews a gravity equation of the form

$$PX_{ij} = \alpha_0 Y_i^{\alpha_1} \left(\frac{Y_i}{L_i}\right)^{\alpha_2} Y_j^{\alpha_3} \left(\frac{Y_j}{L_j}\right)^{\alpha_4} D_{ij}^{\alpha_5} A_{ij}^{\alpha_6} U_{ij} \tag{21}$$

where PX_{ij} refers to the value of export flows from country *i* to *j*, Y_i and Y_j are GDP of country *i* and *j*, $L_i(L_j)$ is the population size, D_{ij} refers to the distance between economic centers of *i* and *j*, A_{ij} represents any other factors influencing bilateral export flows and U_{ij} stands for a log-normally distributed error term. In equation (21), α_0 represents a constant term, the exponents α_1 to α_4 are assumed to be positive whereas α_5, the exponent of distance, is expected to inhabit a negative sign.

Bergstand's study models a gravity equation which both incorporates factor endowments, referring to Heckscher-Ohlin models (see section 3.3), as well as CES preferences, reflecting Lindner's assumption of similar countries having similar preferences. He assumes that the exporter's income can be interpreted as GDP in units of capital and therefore gives information about the capital-labor endowment ratio of each country. This links back to the Heckscher-Ohlin theorem and reasons for inter-industry trade. Changes in the importer's income on the other hand can be considered as changing expenditure capability which is evident for a change in taste preferences and alludes to the Helpman and Krugman models of intra-industry trade.

[6] The *love of variety* refers to Dixit-Stiglitz preferences, where utility increases solely on the fact that several differentiated varieties of the good exist (Dixit and Stiglitz, 1977, p. 299).

Consumers maximize their utility which depends on manufactured (good A) and non-manufactured (good B) output under a minimum requirement of consumption subject to a budget constraint. The budget constraint consists of the country's nominal income, which takes the adjusted exchange and tariff rate free on board (f.o.b.) of the good into account as well as the quantity manufactured.[7] Bergstrand (1989) assumes that all firms within the same industry in country *i (j)* charge the same price in market *i (j)*. On the firm side, technology among firms is identical and they use capital and labor as input factors. Each country has a fixed supply of labor and capital. Each firm produces either a differentiated manufactured product or a non-manufactured product and interacts in a Chamberlinian monopolistic competition with its counter-players. Firms maximize profits and ship part of their goods to the foreign market given a constant elasticity of transformation function. This means firms in the first stage produce with increasing returns to scale and in the second stage export to foreign markets with decreasing returns to scale because of incurring transportation costs.

Substituting supply by the profit maximizing firm for good A into the market demand of good A, assuming price terms in the market demand for good A are given exogeneously, yields

$$PX_{ij} = \alpha_0 Y_i^{\alpha_1} \left(\frac{K_i}{L_i}\right)^{\alpha_2} Y_j^{\alpha_3} \left(\frac{Y_j}{L_j}\right)^{\alpha_4} C_{ij}^{\alpha_5} T_{ij}^{\alpha_6} E_{ij}^{\alpha_7} P_i^{\alpha_8} P_j^{\alpha_9} U_{ij} \, . \tag{22}$$

In equation (22), the nominal export value PX_{ij} is dependent on the exporter's and importer's GDP (Y_i and Y_j) , the capital-labor ratio of country *i* $\left(\frac{K_i}{L_i}\right)$, the per capita income of the importing country *j* $\left(\frac{Y_j}{L_j}\right)$, and C_{ij} standing for the c.i.f./f.o.b. transport factor, T_{ij} representing the tariff rate, E_{ij} resembling the exchange rate and P_i and P_j stand for the exporting and importing country's price terms.[8] As before, a log-normally distributed error term is added. Comparing equation (22) with (21), the capital labor ratio of country *i* can be approximated by per capita income, the c.i.f./f.o.b. transport factor is another way of expressing geographical distance and the variable A_{ij} in equation (21) represents tariff rate T_{ij}, exchange rate E_{ij} and the two price terms P_i and P_j in equation (22).

Bergstrand (1989) estimates trade flows for the years 1965, 1966, 1975 and 1976 on 16 industrialized countries for 9 different industries. The gravity equation empirically accounts for up to 80 per cent of the variation in trade flows. Most of the time, exporter's and importer's GDP and income per capita have the predicted positive sign, likewise the variables referring to trade agreements or adjacency. The

---

[7] There exist two different ways of how to pass on transportation costs: f.o.b. stands for 'free on board' and denotes that costs of transportation are compensated by the buyer. They are additional to the good's price. The c.i.f. (cost, insurance, freight) price of a good indicates that the price is inclusive of transportation and insurance. Costs of transportation and insurance (from the product's origin to its destination) are paid by the seller.

[8] For exact derivation and composition of the variables, see Bergstand, 1989, pp. 144-146. Due to complexity, the theoretical derivation of the Bergstrand model is neglected since it is of minor interest in the first place.

distance coefficient is negative as expected. Yet, variables reflecting the exporter's and importer's price terms and exchange rate demonstrate mixed findings in terms of predicted signs.

The following section discusses a theoretical derivation of the gravity equation based on Heckscher-Ohlin. The classic Heckscher-Ohlin model of explaining international trade flows was developed before Helpman (1987), Krugman (1979, 1980) and Bergstrand's (1985, 1989) approach of monopolistic competition and economies of scale. These authors disagreed with the Heckscher-Ohlin model and challenged it. Nonetheless, Deardorff (1998) refers back to Heckscher-Ohlin and answers the critique of Helpman, Krugman and Bergstand. According to this, the Heckscher-Ohlin approach is presented after the monopolistic approach, although its theory historically emerged before the Helpman and Krugman (1985) model.

3.3 Trade and differing factor endowments

The Heckscher-Ohlin model explains trade as a result of relative differences in factor endowments between countries. Countries trade with each other due to their unequal labor and capital endowments allowing for a differing productivity in the manufacturing of a good. This can be either capital or labor intensive, and hence economies choose to manufacture the product being intensive in the production factor they are endowed with relatively much. Hence, countries have a comparative advantage in the production of a good. They will tend to export goods which in production use relatively much of the abundant factor the country is endowed with and on the other side they will import goods being relative intensive to manufacture in the scarce factor the country is endowed with. Consequently, the exported goods are relatively cheap to produce, whereas the imported goods are relatively expensive in production due to factor scarceness. Models of the Heckscher-Ohlin type focus on explaining trade between industries - that is inter-industry trade - and are often constrained by constant returns to scale (Krugman and Obstfeld, 2006, pp. 50-61). Deardorff (1982) proved the general validity of the Heckscher-Ohlin theorem, though it is only until later he employs the Heckscher-Ohlin trade theory to derive a theoretically founded gravity equation discusses below.

Deardorff (1998) refutes Helpman and Krugman's thesis that a gravity model cannot deliver a theoretical foundation by applying the Heckscher-Ohlin approach. He assumes two scenarios: frictionless trade and trade with impediments.

In the frictionless case, there are no transport costs or other barriers to trade. Since goods are homogeneous and producers and consumers do not have to pay any transportation costs, they are indifferent to the location they sell and buy their products. Deardorff describes each trade transaction as a choice from a world pool of goods, where producers put their products in the first place and consumers choose their goods from this pool accordingly. Word markets clear at a world price for each good due to perfect competition. The development of the gravity equation below is a more general one and could apply to any trade model which assumes perfect competition.[9] Assuming homothetic and identical preferences as well as balanced trade, Deardorff argues that income must equal expenditure

$$Y_i = p'x_i = p'c_i \tag{23}$$

where p' is a world price vector and x_i and c_i are production and consumption vectors of country *i*. Because countries have identical and homothetic preferences, they spend the same income proportion, β_k, on good *k*. Country *j* consumes good *k* in the following amount

$$c_{jk} = \beta_k \frac{Y_j}{p_k}. \tag{24}$$

Since all countries add to the world pool of good *k*, country *i* contributes the following share to the world market which consist of product *k* from all manufacturing countries *h*

$$\gamma_{ik} = \frac{x_{ik}}{\sum_h x_{hk}} \tag{25}$$

and thus country *j* will purchase from country *i* the following consumption

$$c_{ijk} = \gamma_{ik}\, c_{jk} = \gamma_{ik}\, \beta_k \frac{Y_j}{p_k}. \tag{26}$$

Denoting world output of good *k* as $x_k^w = \sum_h x_{hk}$ and due to identical fractions of income spent on good k among all countries, the world share spent on k in relation to world income Y^w must equal $\beta_k = p_k x_k^w / Y^w$. Hence, country *j*'s value of all imported goods from country *i* is

$$X_{ij} = \sum_k p_k c_{ijk} = \sum_k p_k \gamma_{ik}\, \beta_k \frac{Y_j}{p_k} = \sum_k \gamma_{ik}\, \beta_k Y_j$$

$$= \sum_k \frac{x_{ik}}{x_k^w} \frac{p_k x_k^w}{Y^w} Y_j = \sum_k x_{ik} p_k \frac{Y_j}{Y^w}$$

$$X_{ij} = \frac{Y_i Y_j}{Y^w} \tag{27}$$

Since no transport costs exist, distance plays no role in this model. This equation is comparable to equation (7) derived by Anderson (1979) and shows that multiple starting points in theoretical derivation exist and are valid. Deardorff draws a second scenario of frictionless trade with arbitrary preferences, that is each country spending a different share of its income on the good *k*, then the value of imports from country *i* to *j* is

[9] Deardorff comments that the Heckscher-Ohlin assumption is not necessarily the decisive factor in this case but a model incorporating perfect competition is essential to derive this kind of a model. It could also be a Ricardian model, a specific-factor model or a model which bases on arbitrary technologies (Deardorff, 1998, p. 12).

$$X_{ij} = \frac{Y_i Y_j}{Y^w}\left(1 + \sum_k \lambda_k \alpha_{ik}\ \beta_{jk}\right) \tag{28}$$

where λ_k is the share of world income in product *k*, α_{ik} is the share of country *i* in production of good *k* and β_{jk} is the share of consumption in good *k* by country *j*.[10] Comparison shows the value of exports from *i* to *j* with arbitrary preferences varies around the frictionless trade equilibrium expressed in equation (27).

With barriers to trade, Deardorff assumes complete specialization according to Armington preferences. This is due to invalidity of the factor price equalization which is necessary to overcome the positive trade barriers, i.e. transportation costs.[11] If there is no factor price equalization, a least-cost producer for every good exists; hence, countries specialize in the product they have a relative input cost advantage in and then engage into trade. Therewith, each country produces only one single good. Cobb-Douglas preferences imply that each country dedicates a fixed share of their income to the purchase of goods from country *i*, β_i, country *i*'s income is denoted by

$$Y_i = p_i x_i = \sum_j \beta_i Y_j = \beta_i Y^w\ . \tag{29}$$

With $\beta_i = Y_i / Y_w$ and assuming trade is c.i.f., the gravity equation is equal to the case of zero impediments to trade with homogeneous preferences, see equation (27). Trade flows are neither dependent on distance nor on trade costs. But if exports are measured f.o.b., the gravity equation yields

$$T_{ij}{}^{f.o.b.} = \frac{Y_i Y_j}{t_{ij} Y^w} \qquad\qquad T_{ij}{}^{c.i.f.} = \frac{Y_i Y_j}{Y^w} \tag{30}$$

where t_{ij} represents the transport factor (1+ the transportation cost factor between country *i* and *j*). Thus, trade is reduced by transport costs. This is coherent with the standard implication that trade diminishes when transport costs exist. Considering CES preferences, Deardorff's Heckscher-Ohlin setting yields

[10] More accurately α_{ik} and $\beta_{jk}$are the proportional deviations of country *i*`s production shares and of country *j*'s consumption shares of the world average respectively. For a detailed derivation, see Deardorff, 1998, pp. 13-16.

[11] The factor price equalization theorem states that the relative prices for two factors of production in two different countries will equalize themselves because of competition and trade (Krugman and Obstfeld, 2006, pp. 64-66).

$$T_{ij}^{f.o.b.} = \frac{Y_i Y_j}{t_{ij} Y^w} \left[\frac{\left(\frac{t_{ij}}{P_j^I}\right)^{1-\sigma}}{\sum_h \theta_h \left(\frac{t_{ih}}{P_h^I}\right)^{1-\sigma}} \right] = \frac{Y_i Y_j}{t_{ij} Y^w} \left[\frac{\rho_{ij}^{1-\sigma}}{\sum_h \theta_h \rho_{ih}^{1-\sigma}} \right] . \tag{31}$$ [12]

Here, ρ_{ij} is the relative distance from *j* to its supplier *i*, σ indicates the elasticity of substitution and θ_h stands for the share of country *h* in world income.[13] The term $\frac{\rho_{ij}^{1-\sigma}}{\sum_h \theta_h \rho_{ih}^{1-\sigma}}$ measures the relative distance between importing country *j* to the exporting country *i* in comparison to the average of all importing countries' relative distance to supplier *i*. If these two distances have the same size, the term in brackets equals one and the gravity equation will be consistent with equation (27) in the case of frictionless trade or trade with transportation costs measured in terms of c.i.f.. If the relative distance between *i* and *j* is greater (smaller) than the average, trade between the two countries will be less (more) than trade in the standard gravity equation (Deardorff, 1998, p. 20). The term $\frac{\rho_{ij}^{1-\sigma}}{\sum \theta_h \rho_{ih}^{1-\sigma}}$ describes a possible *multilateral term*, since it sets *more than the geographical distance* between country *i* and *j* in relation to *more than the geographical distance* of all other countries that trade with *i*. *More than the geographical distance* means that the gravity equation not only takes into account the distance between production hubs reflected by the transport factor but also incorporates price indices including information about each country's market.

This is comparable to the bilateral economic distance between trading countries in relation to a trade weighted average of economic distance of the exporting nation to all its other possible trading partners as described by Anderson (1979), see section 3.1 and equation (20).

Factors influencing the gravity equation in this case can be either the elasticity of substitution or transportation costs:

i. The higher the elasticity of substitution, countries close to each other will tend to trade more among each other. This is likely to exceed predicted trade by the standard gravity equation. On the other hand, countries further apart will trade less than the predicted trade level.
ii. In the same way, a reduction in transportation costs due to higher technological advancement will bring trade flows closer to the value of exports that is expected by the standard gravity

[12] The case under trade inclusive of transportation costs (c.i.f.) is the same expression as in equation (31) multiplied by t_{ij}. P_j^I is defined as the CES price index $P_j^I = \left(\sum_i \beta_i t_{ij}^{1-\sigma} p_{ij}^{1-\sigma}\right)^{1/1-\sigma}$. For a detailed derivation, see Deardorff, 1998, pp. 18-20.

[13] The elasticity of substitution is defined as the relative change in the ratio of goods demanded in relation to the change of the goods' price ratio and describes how consumers substitute among products under the assumption of a constant utility (Mittermaier, 2008, p. 3).

equation. Thus, trade will relatively more expand between distant countries, while close trading partners lose some of their *proximity advantage* (Deardorff, 1998, p. 20).

Evenett and Keller (2002) find strong empirical support for a 2x2x2 (2 goods, 2 production factors, 2 countries) unicone Heckscher-Ohlin model. Unlike Deardorff (1998), assuming each country produces only one good (product differentiation due to Armington preferences), they incorporate imperfect product specialization where both countries produce both goods. Evenett and Keller (2002) try to solve what they call the *model identification problem*. Several theoretical assumptions in mind, they test empirically which theory is most consistent with the data. Four model specifactions are tested: a Heckscher-Ohlin and a Helpman-Krugman set-up and in each case models of perfect an imperfect specialization. Both models of complete specialization result in the standard gravity equation equal to $X_{ij} = \frac{Y_i Y_j}{Y_w}$. Imperfect specialization under the Helpman-Krugman assumptions yields

$$X_{ij} = \left(1 - \gamma_j\right) \frac{Y_i Y_j}{Y_w}$$

where γ_j is the share of good *z* in country *j*'s GDP. Here, country *j* is assumed to be relatively capital-abundant and country *i* is relatively labor-abundant. Country *i (j)* will therefore export labor-intensive (capital-intensive) goods *to j (i)*. Good *z* is produced under constant return of scale, is a homogeneous product and is the more labor-intensive good. Therefore, the higher the volume of exports from country *i* to *j*, the lower the share of homogeneous goods in the importing country *j*'s GDP. Good *x* on the other hand is produced under increasing returns to scale and is the differentiated and more capital-intensive product. The gravity equation changes under a Heckscher-Ohlin set-up to

$$X_{ij} = \left(\gamma_i - \gamma_j\right) \frac{Y_i Y_j}{Y_w}$$

with both goods *x* and *z* being homogeneous products and being produced under constant returns of scale. As the capital-labor ratios of both countries approach each other and converge, this is equally reflected in γ_i and γ_j. Finally, when *i* and *j* have equal factor proportions, the Heckscher-Ohlin model theorem predicts no trade between the countries in question. Hence, when Heckscher-Ohlin assumptions are applied to a model, Evenett and Keller (2002) predict greater differences in factor proportions, that is perfect specialization among countries. To distinguish between countries that trade mostly intra-industry or inter-industry specific, the authors apply the Grubel-Lloyd index which indicates the extent of intra-industry trade.[14] They sort bilateral import flows between country pairs

[14] The Grubel-Lloyd Index was developed by Grubel and Lloyd (1975). The index will be equal to 1 if there is only intra-industry trade, whereas it is zero when only inter-industry trade exists.

into two groups. The critical level chosen is a Grubel-Lloyd index of 0.05.[15] The Grubel-Lloyd sample which lies above 0.05 describes intra-industry trade or a Helpman-Krugman model specification and the sample below 0.05 is an indication for inter-industry trade or Heckscher-Ohlin based specification. In the data, 78 per cent of observations fall in the low Grubel-Lloyd sample, and 22 per cent are above the benchmark of 0.05. Evenett and Keller find that both models of perfect specialization overpredict the volume of bilateral trade. They conclude that "(…) models of prefect specialization do not appear to be important in explaining the success of the gravity equation." (Evenett and Keller, 2002, p. 297). Yet, they find support in models with imperfect specialization, both in the case of a Helpman-Krugman models as well as the unicone Heckscher-Ohlin model. However, a comparison between country specific estimates of imperfect specialization models shows that the unicone Heckscher-Ohlin model is to be preferred from an empirically point of view. The established relation between factor endowments and trade is significant for countries which are slightly or intermediate differing in their factor proportions and have the same cone of production.

In sum, the Heckscher-Ohlin trade theory delivers a gravity equation that is theoretically justified. Empirical research calls for a unicone Heckscher-Ohlin model of imperfect specialization. This gives reason to why the Heckscher-Ohlin model has been mentioned in this section, although it is the pioneer of Helpman and Krugman's work. The derivation of a structural gravity model by Deardorff (1998) which is represented by equation (31) is comparable with Anderson's economic distance interpretation (equation (20)), however it additionally incorporates CES price indices which give information about the countries' price levels.

For the sake of completeness, the oldest of all trade theories, the Ricardian approach and its reference to a theoretical gravity equation is subsequently discussed.

3.4 Trade and differences in production technologies

Ricardian trade theory builds on the assumption that trade is beneficial due to comparative advantage. A country that is less productive in absolute terms can nevertheless have a comparative advantage in the production of a good. Comparative advantage arises because of differing costs in production or production technologies (Krugman and Obstfeld, 2006, pp. 24-27).

Eaton and Kortum (2002) develop a Ricardian model of bilateral trade which is based on differences in production technology. They incorporate the state of technology in each country (absolute advantage), the heterogeneity of technology (comparative advantage) and the geographic barriers

[15] Evenett and Keller (2002) use cross-sectional data among 58 countries for the year 1985. The data consists of countries with a GDP greater than US$ 1 billion and to which a comparable estimate of capital per worker is available.

bilateral trade is exposed to. Eaton and Kortum assume that labor is the only production factor that is not mobile internationally. With constant returns to scale and iceberg transportation costs, consumers purchase the worldwide cheapest good produced and maximize a CES utility function.[16] The efficiency of production a country exhibits is random and acts accordingly to a Fréchet distribution. The Fréchet distribution allows the state of technology (absolute advantage) and the degree of efficiency (comparative advantage) to vary across countries. The authors state the Fréchet distribution is suitable to describe "(…) a world of many countries that differ in the basic Ricardian senses of absolute and comparative advantage across a continuum of goods." (Eaton and Kortum, 2002, p. 1747). Also the pricing of goods reflect the state of technology, the input costs and the geographic barriers each county faces.

Eaton and Kortum (2002) assume the fraction of goods country *j* buys from *i* equals the expenditure of country *j* on goods from country *i*. Bilateral trade flows are expressed as

$$X_{ij} = \frac{\left(\frac{d_{ij}}{p_j}\right)^{-\theta} X_j}{\sum_m^N \left(\frac{d_{im}}{p_m}\right)^{-\theta} X_m} Q_i \tag{32}$$

where X_j is the total expenditure of country *j* on goods, Q_i is exporter *i*'s total sales and $\left(\frac{d_{ij}}{p_j}\right)^{-\theta}$ is the geographic barrier (d_{ij}) between *i* and *j* deflated by the price level in the importing country *j* (p_j).[17] The numerator term $\left(\frac{d_{ij}}{p_j}\right)^{-\theta} X_j$ can be thought of as the market size of country *j* as perceived by country *i* and the denominator reflects the total world market from country's *i* perspective respectively. Q_i reflects exporter *i*'s income Y_i and $\sum_m^N X_m$ can be thought of as world income Y^W.

Here, the sensitivity of trade depends on the parameter θ, which reflects the heterogeneity in technology of goods production and hence, comparative advantage. This is a difference in assumptions compared to the Helpman-Krugman and Heckscher-Ohlin models which are described by monopolistic competition and differences in factor endowments respectively (see section 3.2 and 3.3). The models referred to use the preference parameter σ to describe the heterogeneity of products in consumption, not in production as Eaton and Kortum (2002) do in a Ricardian model. Thus, bilateral trade flows depend in this case on production costs and distance. In addition, they show a relation

[16] Samuelson (1952) compares occurring trade costs and the actual value of traded goods to the amount of ice melting during transit depending on the size of the iceberg. According to this, a fraction of the good is lost and so to speak 'melts' during transportation. Iceberg trade costs can be compared to an ad valorem tax equivalent.
[17] For a detailed derivation, see Eaton and Kortum, 2002, pp. 1745-1751.

between trade flows and price levels. The following equation, describes country i's shares in country *j* in relation to its home market presence

$$\frac{X_{ij}/X_j}{X_{ii}/X_i} = \left(\frac{p_i d_{ij}}{p_j}\right)^{-\theta} \quad .$$

As prices in country j fall (or prices in country i rise) which is comparable to an increase of geographical barriers between the two countries, country i's normalized export share in country *j* declines. If θ rises, the comparative advantage declines as relative efficiency in production across goods converge. Consequently, the production of some goods does not overcome the distance related hurdles or differences in price levels (Eaton and Kortum, 2002, p. 1751).

By taking input costs and price levels into account, Eaton and Kortum derive their final gravity equation

$$ln\frac{X'_{ij}}{X'_{jj}} = S_i - S_j - \theta m_n - \theta d_k - \theta b - \theta l - \theta e_h - \delta_{ij} \tag{33}$$

where S_i measures the competitiveness of country *i*, m_n is an overall destination effect for each of the 19 countries the bilateral trade equation is estimated for, d_k is the distance between exporting country *i* and importing country *j* in the *kth* interval with *k*=1,…6, *b* is the border effect when *i* and *j* share one border, l indicates a language effect when the same language in both countries is spoken, e_h indicates the common trade area the two countries belong to and finally δ_{ij} captures the error term.[18] In a sample of 19 OECD countries in 1990, S_i indicates that Japan is the most and Belgium the least competitive country. The distance intervals are all negative and decrease the further apart the two countries are. The border and language coefficient as well as the same trade area coefficient are all positive. The overall destination effect and the competitiveness indicator for each country range from either positive to negative and often (but not always) share the same sign. Eaton and Kortum estimate θ to equal 8.28, which is in the range of comparative studies (see Anderson and van Wincoop, 2003; 2004).

To sum up, a gravity model which is based on the Ricardian trade model focuses on comparative advantage. Comparative advantage arises due to differing technologies in the goods production. Hence, all countries possess a different production technology. Compared to the Heckscher-Ohlin and Helpman-Krugman models, where products are differentiated by endowment factors and by the love of variety respectively, Ricardian trade theory focuses on the state and heterogeneity of technology.

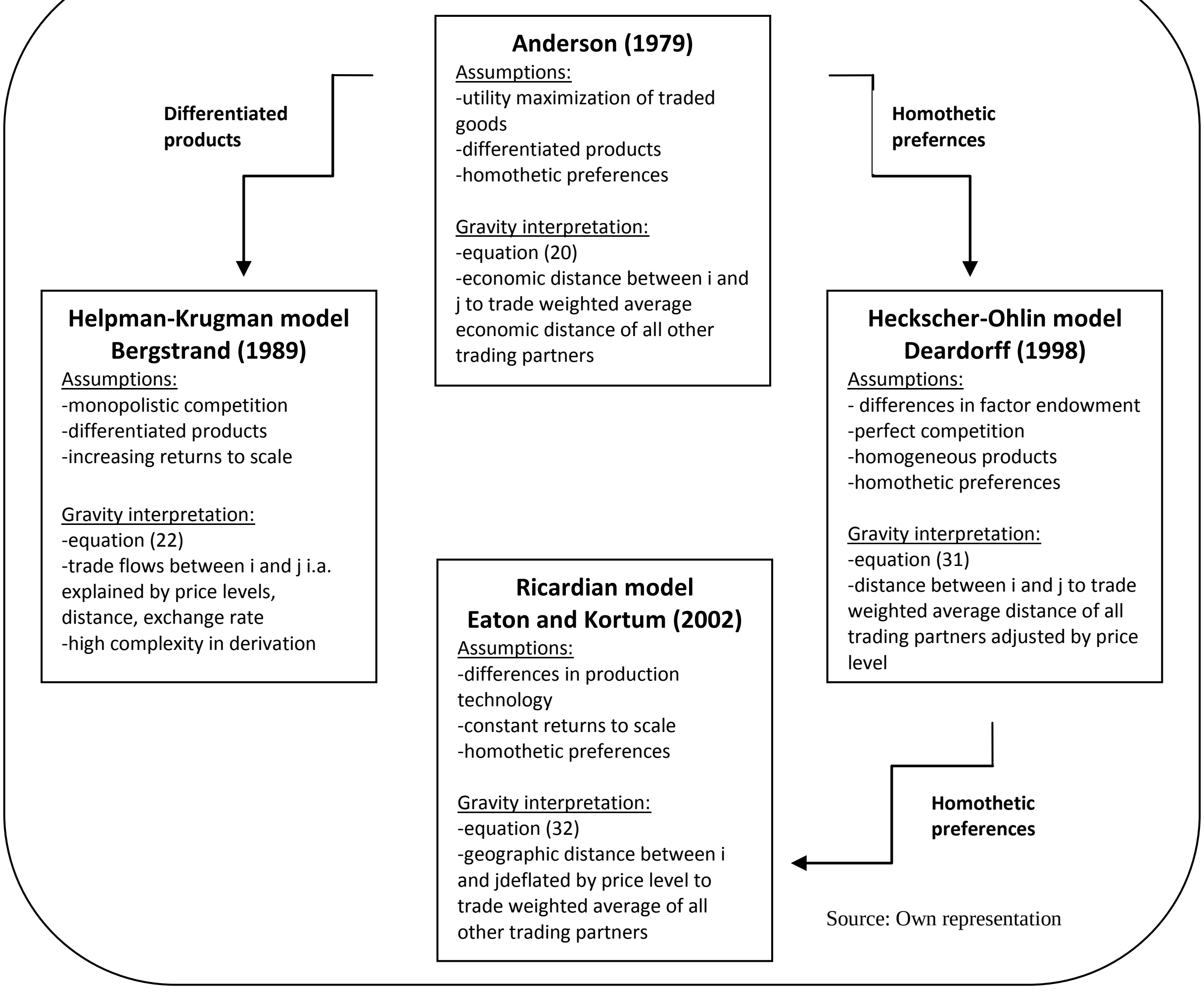
[18] S_i is the competitiveness of country i's state of technology adjusted for its labor costs (see Eaton and Kortum, 2002, p. 1760).

3.5 Summary

The previous sections described the background of the following to come which aims at explaining current developments of theoretical foundation of the gravity equation. From this literature review, one can conclude that the theoretical foundation of the gravity equation cannot be accredited to one trade theory but that a multiple of approaches exist, each with differing assumptions. It is not one precise trade model which accounts for the theoretical derivation, but the assumptions, plausibility and consideration of multiple factors that give credibility to the gravity equation. Figure 1 gives a short overview of the mentioned approaches and links them by their similarities.

Comparing equation (32) to Anderson's (1979) economic distance gravity model in equation (20) and Deardorff's (1998) derived equation (31), one can conclude that all gravity equations imply a comparison between the bilateral trade relationship between country *i* and *j* and the weight adjusted barriers to all possible other trading partners. Whereas Anderson (1979) accounts for economic distance, Deardorff (1998) and Eaton and Kortum (2002) adjust for price levels due to a CES preference structure.

Figure 1: Overview of theoretical gravity approaches before 2003



Source: Own representation

Theoretical origins as in Anderson (1979) and the CES preference structure in Deardorff (1998) and in Eaton and Kortum (2002) are of importance when drawing conclusion on what is essentially new to gravity modeling which is presented next. Thereby, I will develop my answer to the research question by demonstrating what is significantly new to theoretical gravity modeling in comparison to the approaches mentioned.

4 Recent developments in the theoretical foundation

This chapter takes recent developments in the theoretical foundation of the gravity model into account. It focuses on the development of Anderson and van Wincoop's multilateral resistance terms (see section 4.2). Anderson and van Wincoop (2003, 2004) respond to McCallum's 1995 published paper in which he demonstrates trade patterns between Canada and the United States of America (U.S.). Anderson and van Wincoop were motivated by the resulting *border puzzle*. This indicates a tendency for countries to trade and buy domestic products originating from their home country suppliers. They develop a method which allows estimating gravity equations more precisely and consistently due to targeting omitted variable bias (OVB). The multilateral resistance terms allow for a more integral approach to gravity equations since they consider multiple equilibria in international trade.

4.1 Motivation: The border puzzle

McCallum (1995) analyzes the implications of trade patterns between Canadian provinces and U.S. states. His findings reveal that Canadian provinces trade more than 20 times as much among each other than Canadian provinces and U.S. states do. This phenomenon is often described as the *border puzzle* since country borders seem to have a significant effect on trade patterns between countries, even if these are culturally and economically very similar to each other (Mc Callum, 1995, p. 615). McCallum's empirical findings can be considered as initial motivation for Anderson and van Wincoop's development and elaboration on gravity theory to explain the border puzzle effect which is described later in detail.

In McCallum (1995), the trade flow between two entities (either inter-provincial trade or province-state trade) is estimated by the gravity equation equal to

$$lnX_{ij} = a + b\ lnY_i + c\ lnY_j + d\ lndist_{ij} + eDUMMY_{ij} + U_{ij} \tag{34}$$

where X_{ij} is the trade flow from region *i* to region *j*, Y_i and Y_j are the provinces' or states' GDP, $dist_{ij}$ refers to the geographical distance between *i* and *j* and most importantly $DUMMY_{ij}$ reflects a dummy variable that is equal to 1 if the regions are two provinces within Canada (inter-provincial trade) or 0 for province-state trade, that is trade between a Canadian province and a U.S. state. U_{ij} represents the error term. The analysis considers all of the 10 Canadian provinces and is reduced to the 20 biggest U.S. states and all 10 border states. This makes up a total of 90 per cent of trade between the U.S. and Canada in 1988. The North American Free Trade Agreement (NAFTA) came only into effect in 1994, so the analysis refers only to the time before NAFTA.

Regression (1) in Table 1 reports inter-provincial trade estimates between Canadian provinces. The estimates have the expected signs – provincial GDPs have a positive effect on trade whereas distance has a negative effect. Explanatory power is extraordinarily high and equals up to 89 per cent. The second regression takes Canadian province and U.S. state trade into consideration. It is remarkable which importance is assigned to the dummy variable, reflecting trade across a country border. McCallum finds that inter-provincial trade is 22 times larger than province-state trade, other things equal.[19]

Table 1: Estimation of the gravity equation by McCallum (1995)

Independent Variable	(1) CA-CA	(2) CA-US	(3) GDP > US$10 bn	(4) Populations
lnY_i	1.30 (0.06)	1.21 (0.03)	1.15 (0.04)	1.36 (0.04)
lnY_j	0.96 (0.06)	1.06 (0.03)	1.03 (0.04)	1.19 (0.04)
$lndist_{ij}$	-1.52 (0.10)	-1.42 (0.06)	-1.23 (0.07)	-1.48 (0.07)
$lnDUMMY_{ij}$		3.09 (0.13)	3.11 (0.16)	3.07 (0.14)
Number of Observations	90	683	462	683
Standard Error	0.80	1.10	0.97	1.15
Adjusted R^2	0.890	0.811	0.801	0.797

Source: McCallum, 1995, p. 617; Own representation
Results are show in logarithmic values, standard errors in parentheses

Furthermore, regression (3) only takes the biggest provinces and states into account which earn a GDP of U.S. $10 billion and above. This identifies the possibility of heteroscedastic error terms and focuses on a *comparable group of economic entities*.[20] Results show that they are not significantly different to equation estimates from before and thus no inference on size specific patterns of trade can be made. The last equation replaces logarithmic GDP with population variables. Since exports are part of each country's GDP and therefore GDP variables could be correlated with the error term, a different measure is used which is not necessarily linked to province or state GDP. This results in no fundamental deviation from prior estimates. Hence, one can focus on results from regression (1) and

[19] This is because $e^{3.09} = 21.977$ since it is measured in logarithmic values.
[20] Effects of heteroscedastic error terms are commented on in Gujarati and Porter, 2009, pp. 365-401.

(2) since other influential factors have been tested for and do not seem to affect or bias regression results. All models have an explanatory power of 80 per cent or above. McCallum (1995) states that in a borderless world, the gravity model would predict a much stronger north-south trade between provinces and states due to their economic mass. However, the actual patterns of trade indicate a strong inter-provincial trade suggesting national borders do have a large impact on trade and constrain trade among countries, even if these countries are very similar to each other in terms of language, economic institutions and culture as in the case of the U.S. and Canada. McCallum (1995) names this phenomenon the *border puzzle*. In the literature, countries' demand predominance of domestic manufactured goods has been reviewed by multiple researchers and was challenged equally often. Helliwell (1996, 1998) finds high trade density on intra-national level which is around 12 times larger than cross-border trade. On the other hand, Wei (1996) finds home bias to be around 2.5 times as much but mentions a general decline over time. Obstfeld and Rogoff (2001) refer to the border puzzle as *home bias in trade*. They explain a high elasticity of substitution can account for a large home bias in trade, although trade costs might be relatively small. The explanation is as follows: If trade costs are small, this will normally imply an incentive towards imports resulting in a high volume of cross-border trade. However, a strong preference for domestic goods can overcome the incentives towards cross-border trade and causes a cross-border trade diminishing effect. In the end, it may not be the existence of country borders or high trade costs that is responsible for a large intra-national trade in the first place but simply consumer's preferences which are biased towards local goods suppliers.

4.2 The structural gravity equation

Taking the border puzzle as motivation and starting point, Anderson and van Wincoop carry Anderson's (1979) theoretical derivation forward. The latter focused on the economic distance between i and j in relation to a weighted average of economic distance to all other trading partners (after controlling for economic size). Now, Anderson and van Wincoop define multilateral resistance as "(…) the theoretically appropriate average trade barrier (…)" (Anderson and van Wincoop, 2003, p. 170). Since political and economic attention focuses on explaining international trade, the question is which effect the removal of border barriers has on international trade patterns. Anderson and van Wincoop are consequently interested in a comparison of trade equilibrium in a borderless and a border impacted world. The incorporation of multilateral resistance terms allows for such a comparative statistic and helps equally to solve the McCallum border puzzle. The remainder of this chapter outlines the assumptions necessary for a theoretical advancement, the derivation and the interpretation of the gravity equation.

4.2.1 Assumptions of multilateral resistance

Anderson and van Wincoop support a two sided approach. They believe not only a comparison between intra-Canadian and Canada-U.S. trade to be important but additionally to review intra-U.S. trade characteristics. In this case, the $DUMMY_{ij}$ variable in equation (34) is equal to 1 if U.S. states trade with each other, and 0 if U.S. Canadian cross-border trade exists. Table 2 shows estimates based on the year 1993. Regression (1) represents estimates for inter-provincial trade patterns. Estimates have the expected sign. The border effect of 16.44 ($e^{2.80}$) is still remarkably large whereas in regression (2), the U.S. border effect of 1.51 is much smaller. So when controlling for size and distance, U.S. states trade 1.51 times as much with each other than provinces and states. The border coefficients stay close to their prior-estimations when pooling data and allowing inter-provincial, inter-state and inter-province and state trade as shown in regression (3). Anderson and van Wincoop suggest that the large difference between Canadian inter-province and U.S. inter-state trade arises because of OVB - in this context the multilateral resistance terms. Since Canada is a small open economy and relies to a large extent on cross-border trade, even an intermediate trade barrier can result in an increase in provinces' multilateral resistance. On the other hand, a border barrier between the U.S. and its other trading partners has no big effect on states' multilateral resistance, since the U.S. has a large home market and the border barrier does not affect inter-state trade.[21] Striving for a more profound interpretation and theoretical justification of the gravity equation, Anderson and van Wincoop devote their research to the above mentioned OVB. They want to find out what causes OVB.

Tabel 2: Estimation of McCallum's gravity equation by Anderson and van Wincoop (2003)

Independent Variable	(1) CA-CA CA-US	(2) US-US CA-US	(3) US-US CA-CA CA-US
lnY_i	1.22 (0.04)	1.13 (0.03)	1.13 (0.03)
lnY_j	0.98 (0.03)	0.98 (0.03)	0.97 (0.02)
$lndist_{ij}$	-1.35 (0.07)	-1.08 (0.04)	-1.11 (0.04)
$lnDUMMY_{ij} - Canada$	2.80 (0.12)		2.75 (0.12)
$lnDUMMY_{ij} - U.S.$		0.41 (0.05)	0.40 (0.05)
Adjusted R^2	0.76	0.85	0.85

Source: Anderson and van Wincoop, 2003, p.173; Own representation
Results are shown in logarithmic values, standard errors in parenthesis

[21] Anderson and van Wincoop (2003) also calculate equation (34) with a remoteness variable for region i and j, which is $REM_i = \sum_{m\neq j} d_{im}/y_m$reflecting the average geographical distance of region i to all its possible trading partners other than j (Anderson and van Wincoop, 2003, p. 174). Adding these remoteness variables to the regression has little effect on the border effect estimates.

Their model is built on the following three main assumptions:

Goods are differentiated by their place of origin (Armington preferences). Each country specializes in the production of one good in the derivation to follow. However, an expansion of the model also can inhabit several different product classes, and then countries would choose to concentrate on the production of some but not all goods. Since Anderson and van Wincoop's goal is to detect overall trade resistance, they are not interested in good specific trade resistance as this varies depending on the product class under consideration.[22]

Furthermore, trade separability exists. Anderson (1979) also implies trade separability. This means that agents maximize their utility by solely maximizing the consumption of tradable goods (see section 3.1). Anderson and van Wincoop augment and impute that "(…) the allocation of trade across countries can be analyzed separately from the allocation of production and consumption within countries." (Anderson and van Wincoop, 2004, p. 707). This simplifies the determination of trade equilibrium, since it enables to solely focus on trade allocation among bilateral country pairs without accounting for the market structure of the supply and expenditure side within countries. In other words, each country trades with a single world market (Anderson, 2009a).[23] This assumption simplifies the patterns of distribution costs and allows a characterization of trade cost effects independent from regional or national production and consumption patterns. The following requirements must be fulfilled to achieve trade separability.

Countries have identical and homothetic preferences. Identical preferences imply that countries have the same demand structure and spend the same fraction of their income on the good from country ***i***. Homothetic preferences imply that the proportion in demand for any good is dependent on the relative prices of this good and does not depend on the income level. This is best approximated by a CES

---

[22] The model is initially developed in the context of trade between Canadian provinces and U.S. states and refers to regions. However, in the preceding parts, I mentioned trade patterns between countries. To perpetuate continuity, I will not speak of regions but of countries since the model can be applied to different geographical entities. Furthermore, the production of only one good per country is assumed.
Subsequently, in the case of differentiated products indices must be adjusted with the superscript k, however the assumptions of a world with differentiated goods does not change the interpretation and intention of the equation. By arguing for different product classes that are denoted by *k*, Anderson and van Wincoop (2004) argue for estimation with disaggregated data, see also section 4.4.3. They expect a large bias when using aggregated data. Disaggregation and aggregation of data in this context refers to a preferably high digit-number of the Standard International Trade Classification, International Harmonized Commodity Coding and Classification System or Harmonized Tariff Schedule. The more accurate trade flows are classified, the more disaggregated data is available (Anderson and Yotov, 2008; 2011).

[23] This can be amplified by adding different goods classes or product lines. It would result in the determination of trade allocation among country pairs without accounting for the market structure of the supply and expenditure side for each product class within countries. That is each product class of each country trades with a single world market (Anderson, 2009a, pp. 1-5).

utility function where c_{ij} is the consumption of country *j* in goods from country *i* and σ is the elasticity of substitution.[24] The utility function is given by

$$U = \left(\sum_i c_{ij}^{(\sigma-1)/\sigma} \right)^{\sigma/(\sigma-1)} \tag{35}$$

and is maximized by consumers in country *j* subject to the budget constraint

$$\sum_i c_{ij} t_{ij} p_i = Y_j \, .^{25} \tag{36}$$

The budget constraint represents total expenditure of country *j* resulting from the purchasing of goods from country i. In the equation, p_i stands for the supply price (excluding transportation or other costs) and t_{ij} >1 for the trade cost factor between origin *i* and destination *j*. Since $c_{ij} t_{ij} p_i$ is the cost for country *j* to import goods from country *i*, the sum over all goods imported from other countries by *j* is importer *j*'s total expenditure. Since countries can only spend what they earn, Y_j is country *j*'s income.[26]

Trade costs are proportional to the value of traded goods. They are not defined by the volume of trade. This can be understood as an ad valorem tax equivalent or iceberg trade costs. In equation (36), price p_{ij} has been substituted by $p_i t_{ij}$ already since $p_{ij} = p_i t_{ij}$. Thereby, p_i is the price that producers of the good in country *i* receive, whereas t_{ij} is the trade cost factor and $t_{ij} - 1$is the tax equivalent.[27] It is assumed that trade costs are paid for by the exporter, hence trade costs are c.i.f.. Eventually however, they are passed on to importer *j*.

Under these assumptions, the nominal value of exports from *i* to *j* is

$$X_{ij} = c_{ij} \, p_{ij} \, . \tag{37}$$

Total income for country *i* is composed of the sum of exports from *i* to all other countries and is called the market clearing condition

$$Y_i = \sum_j X_{ij} \, . \tag{38}$$

[24] For more on constant elasticity of substitution, see Footnote 13 and Mc Fadden (1963).
[25] In Aderson and van Wincoop (2003) a distribution parameter β is added to the CES utility function which thereafter becomes $\left(\sum_i \beta^{(1-\sigma)/\sigma} c_{ij}^{(\sigma-1)/\sigma}\right)^{\sigma/(\sigma-1)}$. This distribution parameter does not change the interpretation of the resulting gravity equation and - for simplification – is here neglected.
[26] Assuming that a country's total income Y_j is equal to its total expenses E_j since it is a one sector economy.
[27] Due to trade separability, there is no need to determine the market structure of the supply side. However, if a monopoly exists on the market, p_i contains a mark-up. In the case of perfect competition, the supply price will reflect marginal costs.

4.2.2 Derivation of the structural gravity equation

All further applications are based on the utility CES function, budget constraint and the expenditure system of the two trading countries (see equations (35) to (38)). For deriving the gravity model, country *j*'s demand for goods originating from country *i* needs to be determined first with the help of the utility function and the budget constraint. In a second step, this demand needs to be aggregated over all importing countries from exporting country *i*. Therefore, demand is substituted in the market clearing condition to aggregate trade flows and estimate equilibrium trade.

The first step requires maximizing the CES utility function subject to the budget constraint. This yields the maximization following problem

$$L = \left(\sum_i c_{ij}^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}} + \lambda\left(Y_j - \sum_i c_{ij}t_{ij}p_i\right). \tag{39}$$

Deriving with respect to c_{ij} since the aim is to maximize utility (i.e. consumption c_{ij}) and setting the derivation equal to zero gives

$$\frac{\partial L}{\partial c_{ij}} = \frac{\sigma}{\sigma-1}\left(\sum_i c_{ij}^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}-1}\frac{\sigma-1}{\sigma}\,c_{ij}^{\frac{\sigma-1}{\sigma}-1} - \lambda\, t_{ij}p_i = 0$$

$$\lambda\, t_{ij}p_i = \left(\sum_i c_{ij}^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}-\frac{\sigma-1}{\sigma-1}} c_{ij}^{\frac{\sigma-1}{\sigma}-\frac{\sigma}{\sigma}}$$

$$\lambda\, t_{ij}p_i = \left(\sum_i c_{ij}^{\frac{\sigma-1}{\sigma}}\right)^{\frac{1}{\sigma-1}} c_{ij}^{\frac{-1}{\sigma}} \tag{40}$$

Multiplying with c_{ij}, and then summing over all *i* is equal to

$$\lambda\, t_{ij}p_ic_{ij} = \left(\sum_i c_{ij}^{\frac{\sigma-1}{\sigma}}\right)^{\frac{1}{\sigma-1}} c_{ij}^{\frac{-1}{\sigma}+\frac{\sigma}{\sigma}}$$

$$\lambda \sum_i t_{ij}p_ic_{ij} = \left(\sum_i c_{ij}^{\frac{\sigma-1}{\sigma}}\right)^{\frac{1}{\sigma-1}} \sum_i c_{ij}^{\frac{\sigma-1}{\sigma}}. \tag{41}$$

Substituting λ with $\left(\sum_i c_{ij}^{\frac{\sigma-1}{\sigma}}\right)^{\frac{1}{\sigma-1}} c_{ij}^{\frac{-1}{\sigma}}\left(t_{ij}p_i\right)^{-1}$from equation (40), and replacing $\sum_i t_{ij}p_i c_{ij}$ in equation (41) by Y_j due to the transferability of the market clearing condition (equation (38)) on country *j*'s total income, this results in

$$\left(\sum_i c_{ij}^{\frac{\sigma-1}{\sigma}}\right)^{\frac{1}{\sigma-1}} c_{ij}^{\frac{-1}{\sigma}}\left(t_{ij}p_i\right)^{-1} Y_j = \left(\sum_i c_{ij}^{\frac{\sigma-1}{\sigma}}\right)^{\frac{1}{\sigma-1}} \sum_i c_{ij}^{\frac{\sigma-1}{\sigma}}$$

$$c_{ij}^{\frac{-1}{\sigma}}\left(t_{ij}p_i\right)^{-1} Y_j = \sum_i c_{ij}^{\frac{\sigma-1}{\sigma}} .$$

When solving for the price $p_{ij} = t_{ij}p_i$, one obtains

$$t_{ij}p_i = \frac{c_{ij}^{\frac{-1}{\sigma}} Y_j}{\sum_i c_{ij}^{\frac{\sigma-1}{\sigma}}} .$$

The next step is to multiply left and right hand side with the exponent $(..)^{-\sigma}$ which gives

$$\left(t_{ij}p_i\right)^{-\sigma} = \frac{c_{ij} Y_j^{-\sigma}}{\sum_i c_{ij}^{\left(\frac{\sigma-1}{\sigma}\right)^{-\sigma}}}$$

and in a second step the term is multiplied with $t_{ij}p_i$. Finally, this results in

$$\left(t_{ij}p_i\right)^{1-\sigma} = \frac{t_{ij}p_i c_{ij} Y_j^{-\sigma}}{\sum_i c_{ij}^{\left(\frac{\sigma-1}{\sigma}\right)^{-\sigma}}} . \tag{42}$$

These transformations do not have an appropriate meaning so far. However they are important, very clever and sophisticated steps for deriving the structural gravity equation. One of the characteristics of Anderson and van Wincoop's structural gravity equation is its simplicity. The transformations undertaken at this point are preparatory work for later. Like Anderson (1979), Bergstrand (1989) and Eaton and Kortum (2002), they include a consumer price index reflecting product and trade costs from all the supplying countries the buying country *j* imports from. The inclusion of a price index is characteristic for the CES utility function.[28] The CES price index is needed to determine equilibrium prices later on and is an indication of the price level. To obtain a price structure that can be interpreted

[28] See Footnote 4.

as a price index, it is useful to sum over all i on the left and right hand side and additionally replace $\sum_i t_{ij}p_i c_{ij}$ by Y_j, then this yields

$$\sum_i \left(t_{ij}p_i\right)^{1-\sigma} = \frac{Y_j^{1-\sigma}}{\sum_i c_{ij}^{\left(\frac{\sigma-1}{\sigma}\right)^{-\sigma}}} \; . \tag{43}$$

Anderson and van Wincoop (2003) define equation (44) as the consumer price index for country *j*

$$P_j = \left(\sum_i \left(p_i t_{ij}\right)^{1-\sigma}\right)^{1/(1-\sigma)} \tag{44}$$

and equation (43) can be rewritten as

$$P_j^{1-\sigma} = \frac{Y_j^{1-\sigma}}{\sum_i c_{ij}^{\left(\frac{\sigma-1}{\sigma}\right)^{-\sigma}}} \text{ and thus} \sum_i c_{ij}^{\left(\frac{\sigma-1}{\sigma}\right)^{-\sigma}} = \frac{Y_j^{1-\sigma}}{P_j^{1-\sigma}} .$$

By substituting $\sum_i c_{ij}^{\left(\frac{\sigma-1}{\sigma}\right)^{-\sigma}}$ into equation (42), one obtains

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

and $t_{ij}p_i c_{ij}$ can be replaced by X_{ij} (see equation (37) describing the expenditure system), which results in

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

Finally, solving for X_{ij} gives the demand of country *j* for country *i*'s imports, which is

$$X_{ij} = \frac{\left(t_{ij}p_i\right)^{1-\sigma}}{P_j^{1-\sigma}} Y_j . \tag{45}$$ [29]

[29] The derivation of X_{ij} was difficult and my derivation demonstrated frictions at first. I therefore sought advice and contacted Sebastian Krautheim, who is Assistant Professor in International Economics at the Goethe University in Frankfurt, Germany. Due to research on the internet, his course 'Recent Advances in International Trade' caught my attention (http://www.wiwi.uni-frankfurt.de/ei/index.php?id=2278). In this course, the derivation of the gravity model by Anderson and van Wincoop (2003) was discussed. Mr. Krautheim provided some solution for the derivation, which was also sent to my supervisor Pascalis Raimondos-Møller. I confirm that Mr. Krautheim's advice was only sought in the derivation of the gravity equation (4.2.2) and that no other parts have been developed or written with his support.

Now, the export demand of country j is determined. Since the goal is to derive aggregated demand and general equilibrium trade flows, the market clearing condition helps to aggregate over all countries. By substituting equation (45) into the market clearing condition (equation (38)), this obtains the following

$$Y_i = \sum\nolimits_j \left(\frac{p_i t_{ij}}{P_j}\right)^{1-\sigma} Y_j \,.$$

Solving for the price, the equilibrium market price is given by

$$p_i^{1-\sigma} = \frac{Y_i}{\sum_j \left(\frac{t_{ij}}{P_j}\right)^{1-\sigma} Y_j}. \tag{46}$$

By substituting this price back into the import demand equation (45), demand is given by

$$X_{ij} = \left(\frac{p_i t_{ij}}{P_j}\right)^{1-\sigma} Y_j = \left(\frac{t_{ij}}{P_j}\right)^{1-\sigma} Y_j \;\, Y_i \left(\sum\nolimits_j \left(\frac{t_{ij}}{P_j}\right)^{1-\sigma} Y_j\right)^{-1}$$

$$X_{ij} = \left(\frac{t_{ij}}{P_j}\right)^{1-\sigma} \frac{Y_j \;\, Y_i}{Y_w} \left(\sum_j \left(\frac{t_{ij}}{P_j}\right)^{1-\sigma} \frac{Y_j}{Y_w}\right)^{-1}. \tag{47}$$

In equation (47), the terms Y_w cancel each other out since $\frac{1}{Y_w}\left(\frac{1}{Y_w}\right)^{-1} = 1$ and are just an expansion by world income. This expansion is comparable to the one undertaken in equation (20) by $\sum_j \Phi_j Y_j$ based on Anderson (1979). To simplify the equation, the last term within the brackets on the right hand side of equation (47) is defined as $(\Pi_i)^{1-\sigma}$ (Anderson and van Wincoop 2004, p. 708)

$$(\Pi_i)^{1-\sigma} = \sum_j \left(\frac{t_{ij}}{P_j}\right)^{1-\sigma} \frac{Y_j}{Y_w}.$$

Substituting $(\Pi_i)^{1-\sigma}$ back into equation (47), this yields the final structural gravity model

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

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

$$\left(P_j\right)^{1-\sigma} = \sum_i \left(\frac{t_{ij}}{\Pi_i}\right)^{1-\sigma} \frac{Y_i}{Y_w}. \tag{III}$$

The last term $\left(P_j\right)^{1-\sigma}$can be derived by substituting the isolated price term of equation (46) into the CES price index in equation (44) which then results in

$$P_j = \left(\sum_i (p_i t_{ij})^{1-\sigma}\right)^{\frac{1}{1-\sigma}} = \left(\sum_i (t_{ij})^{1-\sigma}\ Y_i \left(\sum_j \left(\frac{t_{ij}}{P_j}\right)^{1-\sigma} Y_j\right)^{-1}\right)^{\frac{1}{1-\sigma}}$$

Expanding this by dividing and multiplying with world income Y_w, yields

$$P_j = \left(\sum_i (t_{ij})^{1-\sigma} \frac{Y_i}{Y_w} \left(\sum_j \left(\frac{t_{ij}}{P_j}\right)^{1-\sigma} \frac{Y_j}{Y_w}\right)^{-1}\right)^{\frac{1}{1-\sigma}}.$$

The term on the right hand side can be substituted by Π_i. Multiplying the superscript with $(1-\sigma)$ gives

$$(P_j)^{1-\sigma} = \sum_i \left(\frac{t_{ij}}{\Pi_i}\right)^{1-\sigma} \frac{Y_i}{Y_w}. \qquad \text{(III)}$$

Anderson and van Wincoop (2003) find that in the case of symmetric transportation costs, that is $t_{ij} = t_{ji}$, a simple solution to the gravity model exists. Then $\Pi_i = P_i$ is a solution to $(\Pi_i)^{1-\sigma}$ and $(P_j)^{1-\sigma}$. Following this,

$$(P_j)^{1-\sigma} = \sum_i \left(\frac{t_{ij}}{P_i}\right)^{1-\sigma} \frac{Y_i}{Y_w}$$

which is a solution for all countries importing goods from i. Therefore, the price terms are dependent on all country-pair trade barriers and income shares. The gravity equation can be expressed as

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

and thus refers to income shares of the trading countries, bilateral trade barriers and their price indices. However, the assumption of symmetric trade costs is crucial for the last finding. Anderson and van Wincoop assume that symmetric barriers exist because "(…) the border barriers we estimate in this paper [can be interpreted] as an average of the barriers in both directions." (Anderson and van Wincoop, 2003, p. 175, Footnote 11). They state that asymmetric barriers exist but are difficult to empirically identify and incorporate in the model.

4.2.3 Interpretation of multilateral resistance

Equation (I) represents the gravity equation and Π_i and P_j in equation (II) and (III) are called the multilateral resistance terms. Thereby, Π_i is defined as the outward and P_j as the inward multilateral resistance term. The resistance terms measure the exporter's and importer's joint average trade resistance (in terms of trade barriers) each of them faces with all of their other possible trading partners. Thus, trade flows between country *i* and *j* are dependent on three resistances:

(i) the bilateral trade resistance existing between country *i* and *j*, reflected by t_{ij},

(ii) outward multilateral resistance Π_i the exporting country *i* faces to all other trading partners. It is an indication of how easily *i* can export goods to other markets.

(iii) inward multilateral resistance P_j the importing country *j* faces to all other trading partners. It is an indication of how easily *j* can import goods from other markets.

When examining the terms in equation (I) separately, the first term $\frac{Y_j Y_i}{Y_w}$ predicts frictionless trade. This is comparable to equation (7) under the set-up of Anderson (1979) and equation (27) referring to a Heckshcer-Ohlin model by Deardorff (1998). The second term $\left(\frac{t_{ij}}{P_j \Pi_i}\right)^{1-\sigma}$ can be interpreted as bilateral trade bias in relation to average multilateral trade impediments when trade barriers exist. It corrects frictionless trade by the presence of relative trade barriers and reflects relative trade resistance. Thus, the structural gravity model expresses that bilateral trade depends on trade impediments between the trading countries of interest *relative* to impediments each faces with other potential trading partners. Additionally, since equation (II) and (III) need to be solved simultaneously, trade flows between country *i* and *j* also depend on bilateral trade costs between third and fourth parties, e.g. country *m* and *n*. Illustrating this, the NAFTA possibly changed the way of how NAFTA-members and European Union (EU) membership countries traded among one another. Consequently, this had an effect of how EU member countries trade with Japan or China (Anderson, 2009b).

When analyzing dynamics of the inward and outward multilateral resistance, following effects on trade are worth mentioning:

- Inward multilateral resistance P_j: If there is a rise in trade barriers between importing country *j* and all its other potential trading partners (multilateral resistance of *j* rises), the relative price of the exporting country *i*'s good will decrease and trade flows between *i* and *j* will increase.
- Outward multilateral resistance Π_i: Likewise, if multilateral resistance of exporter *i* rises - that is *i* finds it difficult to export to other markets - overall demand on *i*'s exported products will slow down thus reducing the price p_i. If the trade barrier between *i* and *j* is constant and does not change, trade will consequently increase between these two countries (see Anderson and van Wincoop, 2003, p.176).

As this example indicates, trade flows are indirectly connected to each other depending on relative barriers of trade. However, to which extent trade barriers influence the exchange of goods between countries also depends on the economic size of the country. According to Anderson and van Wincoop (2003), three implications can be derived when considering a country's economic size and a rise in trade barriers. They find that multilateral resistance of a small country increases more than multilateral resistance of a large country when a rise in trade barriers exists. This is due to the dependence of small countries on international trade, since intra-national trade is constrained by its small size. Larger countries on the other hand most likely possess a big home market which is characterized by vivid inter-regional trade. Therefore, large countries do not depend as much on international trade as small countries do. When a uniform rise in trade barriers exists, then:

- Implication 1 – International Trade: Trade flows *between* small countries decrease relatively less than trade flows between large countries. This is because relative trade resistance between small countries will increase less than between large countries.

  Explanation: Relative trade resistance is reflected by the term $\left(\frac{t_{ij}}{P_j \Pi_i}\right)$. For small countries, multilateral resistance increases more than for big countries because small economies depend on international trade. The increase of multilateral resistance is captured by a rise in the denominator. Large countries feature a small increase in multilateral resistance resulting in a smaller increase in the denominator. Hence, relative trade resistance increases less for smaller countries than for larger countries.
- Implication 2 – Intra-national Trade: Trade *within* small countries rises more than trade within large countries. This is due to the increased multilateral resistance small countries face.

  Explanation: A rising multilateral resistance goes along with a relative decline in intra-national resistance $\left(\frac{t_{ii}}{P_i \Pi_i}\right)$. This decline is more drastic for small countries because their multilateral resistance increases more than the one of large economies.
- Implication 3 – Ratio of Intra-national Trade to International Trade: The ratio of trade within country *i* to trade between country *i* and *j* will increase more, the smaller country *i* and the larger country *j* is.

  Explanation: Intra-national trade increases more for small countries than for large countries (implication 2). Furthermore, trade between small countries drops less compared to large countries (implication 1). Hence, this implies a drop in trade the smaller the small country and the bigger the larger country is, since small countries probably focus on trade with other small countries due to a lower relative trade resistance.[30]

These implications can be applied to the U.S.–Canadian border puzzle and they help to explain this phenomenon in the following section. In this constellation, Canada is the much smaller country in terms of economic power with a smaller number of provinces compared to U.S. states. Therefore, Canada is more dependent on international trade since a home market is to some extent absent. With a uniform increase in trade barriers, intra-Canadian trade rises exceptionally strong in relation to state-province trade. This explains the much larger empirical Canadian border effect of 16.44 in comparison to a U.S. border effect of 1.5 (see section 4.1). Prior to the implications derived above, the interpretation was that Canadian provinces trade 16.44 times more than U.S. states and Canadian provinces trade among each other, ceteris paribus, due to the U.S. Canadian border barrier. However, after analyzing the relationship between trade barriers and a country's size, the significant border barrier effect can be also attributed to the different behavior of relative trade resistance that exists among these two countries.

In general, trade costs can be a function of several trade barriers such as transportation costs, language barriers, if countries share the same language, border, currency or are member of the same trade union. A diversity of factors is discussed in chapter 5. Anderson and van Wincoop (2003) review trade costs as

$$t_{ij} = b_{ij} d_{ij}^{\rho} \tag{48}$$

where b_{ij} indicates whether two regions belong to the same country (than $b_{ij} = 1$) or are located in two different countries (than $b_{ij} = 1 + tariff\ equivalent_{ij}$). The geographical distance between the countries is represented by d_{ij}, and the theoretical gravity equation is transformed into a logarithmic form which finally yields

$$\begin{aligned} ln(X_{ij}) = k + ln(Y_i) + ln(Y_j) + (1-\sigma)\, ln(b_{ij}) + (1-\sigma)\rho\, ln(d_{ij}) \\ - (1-\sigma)\, ln(P_j) - (1-\sigma)\, ln\,(\Pi_i) + \varepsilon_{ij} \end{aligned} \tag{49}$$

where ε_{ij} is the unobserved error term.[31] This equation is similar to the one estimated by McCallum (1995) (see section 4.1, equation (34)), but additionally incorporates inward and outward multilateral resistance terms. Under the assumption of symmetric trade costs $t_{ij} = t_{ji}$ and therewith $\Pi_i = P_i$, the

---

[30] These implications are derived in a profound mathematical and more detailed way in Anderson and van Wincoop, 2003, p. 176-178.

[31] Trade barriers are expressed more generally in Anderson and van Wincoop (2004) as $t_{ij} = \Pi_{m=1}^{M}(z_{ij}^{m})$ where z_{ij}^{m} is a multiple of observable factors (see Anderson and van Wincoop, 2004, p. 710). Since the focus of this work is on the border barrier between Canada and the U.S., the border and distance coefficients b_{ij} and d_{ij} will be sufficient for this analysis.

price indices are determined by all country-pair trade barriers, in this case a function of geographical distance and the tariff equivalent. Accuracy of the gravity equation is improved and tries to prevent OVB. Prior to the incorporation of multilateral resistance, remoteness terms were used similar to $REM_i = \sum_{m \neq j} d_{im}/y_m$ which measure the average distance of country i to all its trading partners except trading partner j. It was solely a measurement of geographical trade resistance and did not take notice of a general equilibrium in trade (see section 4.2.1, Footnote 21). Consequently, remoteness terms neglect border barriers such as tariff equivalents to all other trading partners.

To sum up, multilateral resistance enables to capture relative trade resistance. Relative trade resistance sets bilateral partner specific trade between country i and j in relation to trade possibilities among other potential trade partners. Therefore, a structural gravity model takes all bilateral trade resistances into account when eventually determining general trade equilibrium.

4.3 Solving the border puzzle: Empirical results

Anderson and van Wincoop (2004) report multilateral resistances for each of the 10 Canadian provinces and 30 U.S. states. They assume an elasticity of substitution equal to $\sigma = 8$ and imply that import and export trade costs are the same, which results in $P_i = \Pi_i$. They find that Canadian provinces have by tendency a much higher multilateral resistance than U.S. states (Anderson and van Wincoop, 2004, p. 724).

Referring back to McCallum (1995), Anderson and van Wincoop (2003) estimate multilateral resistance for a two country case (only Canada and the U.S.) and for a multi-country model (Canada, U.S. and the rest of world (ROW)). In the two country case, the same 30 states and 10 Canadian provinces are analyzed as in McCallum (1995). However, the other 21 states are group as one region to simplify dependence and capture multilateral resistance from each of the 30 individual specified states on all other states/provinces.[32] The fact that the group of missing states has an influence on trade relations is neglected by McCallum (1995). Since price variables are non-observable, the price indices are solved as an implicit function of border barriers, distance and income shares (see section 4.4.3 which discusses the interpretation of price indices). Anderson and van Wincoop rely on a slight modification of equation (49) and estimate

[32] The distance is calculated by taking the GDP weighted average distance between region i and each region of the 21 states (Anderson and van Wincoop, 2003, p. 179). Instead of including multilateral resistance terms, country-specific dummies can be added to the gravity equation. This equation can then be estimated with OLS.

$$\ln\left(\frac{X_{ij}}{Y_iY_j}\right) = k + (1-\sigma)\ln(b_{ij})(1-\delta_{ij}) + (1-\sigma)\rho\ln(d_{ij}) - \ln(P_j{}^{1-\sigma}) - \ln(P_i{}^{1-\sigma}) + \varepsilon_{ij}\ . \quad (50)$$

Trade costs are defined by equation (48) as $t_{ij} = b_{ij}\mathrm{d}_{ij}^{\rho}$, where b_{ij} is the border coefficient and $\delta_{ij} = 0$ if one determines the trade flow between U.S. states and Canadian provinces and border barrier exists, or $\delta_{ij} = 1$ if intra-national trade exists. Distance between the trading entities is captured by d_{ij}^{ρ}. The analysis to follow focuses on the multi-country case since it tends to reflect world trade more appropriately.

Table 3 reports average multilateral resistance terms in a state where border barriers exist and when there are no border barriers under the assumption of symmetric trade costs. As expected, multilateral resistance is much higher for Canada than for the U.S.. Canada - being a small open economy - faces a rise in multilateral resistance by the factor 2.44 when comparing border barrier trade to borderless trade. This means that multilateral resistance rises by 144 per cent whereas multilateral resistance increases by only 12 per cent in the U.S. case. Since Canada relies heavily on international trade partners, border barriers impact this relation significantly.

Tabel 3: Estimation of multilateral resistance under border barriers and borderless trade in the multi-country model by Anderson and van Wincoop (2003)

Average $P^{1-\sigma}$	U.S.	Canada	ROW
Border barriers (BB)	1.55 (0.01)	4.67 (0.09)	2.97 (0.07)
Borderless trade (NB)	1.39 (0.00)	1.91 (0.04)	1.54 (0.01)
Ratio (BB/NB)	1.12 (0.01)	2.44 (0.09)	1.93 (0.06)

Source: Anderson and van Wincoop, 2003, p.183; Own representation
Results indicate the value after taking the exponential of the average logarithm of all $P_i{}^{1-\sigma}$. The average is taken over the 30 U.S. states and 10 Canadian provinces in the sample, as well as 20 industrilized countries (ROW), standard errors in parenthesis

The U.S. on the other hand relies to its strong internal trade among states. Border barrier disturb less. Anderson and van Wincoop discover that even under borderless trade Canadian provinces are confronted with a larger multilateral resistance (1.91) since distance between Canadian provinces and their trading partners are presumably greater than distance between U.S. states. Once more, this refers back to the small size of Canada and its peripheral position to potential trading partners. The last column in Table 3 presents the multilateral resistance for the ROW. Multilateral resistance is higher than U.S.' but lower than Canadian multilateral resistance and lies in between the former two (Anderson and van Wincoop, 2003, p. 183).

After reviewing the size of multilateral resistance, it is interesting to determine which effect trade resistance has on trade flows. The effect can be decomposed into the impact border barriers and multilateral resistance has on bilateral trade flows as presented in Table 4.

Tabel 4: Impact of border barriers on bilateral trade in the multi-country model by Anderson and van Wincoop (2003)

Bilateral trade pairs	(1) US-US	(2) CA-CA	(3) US-CA	(4) US-ROW	(5) CA-ROW	(6) ROW-ROW
Ratio (BB/NB)	1.25 (0.02)	5.96 (0.42)	0.56 (0.03)	0.40 (0.01)	0.46 (0.01)	0.71 (0.02)
Due to bilateral resistance($t_{ij}^{1-\sigma}$)	1.0 (0.0)	1.0 (0.0)	0.20 (0.02)	0.19 (0.01)	0.10 (0.01)	0.19 (0.01)
Due to multi-lateral resistance ($P_i^{1-\sigma} P_j^{1-\sigma}$)	1.25 (0.02)	5.96 (0.42)	2.72 (0.12)	2.15 (0.09)	4.70 (0.31)	3.71 (0.25)

Source: Anderson and van Wincoop, 2003, p.184; Own representation
Standard errors in parenthesis

As before, Canada - the small open economy - experiences a tremendous increase in intra-national trade by factor 5.96 or nearly 500 per cent when observing a change in state from a border barrier world to frictionless trade (whereas trade between U.S. states increases by only 25 per cent, see columns (1) and (2) in Table (4)). Due to significant increase in multilateral resistance, relative trade resistance for inter-provincial trade decreases enormously and explains the rise in trade between Canadian provinces under border barriers as can be seen in column (2) (this refers to implication 2 in section 4.2.3). U.S. - Canadian cross-border trade is reduced by 44 per cent to a remaining fraction of 56 per cent (column (3)), bilateral trade barriers accounting for a drop of 80 per cent. Simultaneously, a greater multilateral resistance accounts for a rise in cross-border trade by factor 2.72, meaning that the negative effect of bilateral border barriers between the U.S. and Canada is to some extent balanced. According to Anderson and van Wincoop (2003), the balancing effect is obtained because the border barrier makes U.S. goods more expensive for Canada. However, border barriers make goods of all possible trading partners more expensive to Canada which partly compensates the rising costs in the U.S. Canadian case. Implication 1 of section 4.2.3., that trade between large countries decreases more than between small countries, is captured by a 60 per cent decrease in U.S.-ROW trade (column (4)), whereas Canadian-ROW trade drops only by 54 per cent (column (5)). It shows again that the impact of multilateral resistance between Canada and ROW is once again greater (4.70) than between the U.S. and ROW (2.15) which suggests a greater drop in Canadian-ROW relative trade resistance than relative trade resistance would decrease U.S.-ROW wise (Anderson and van Wincoop, 2003, pp. 183-184).

The last aspect to take into account is the relation of intra-national trade to international trade, since McCallum (1995) estimates the ratio of Canadian inter-provincial trade to province-state trade. This distinguishes itself in the following:

i. Before, trade flows are compared in a state with border barriers and without border barriers (Table 3). Further, the border effect is split down into bilateral and multilateral resistance (Table 4).
ii. At present, the question is how trade within the country relates to trade among countries under the assumptions of trade barriers.
iii. This is not to be mixed up with the effect on international trade under the assumption of removing border barriers (as mentioned above, there is a 44 per cent decrease in U.S.-Canadian trade when border barriers exist compared to the frictionless state).

In the multi-country model, the impact of the border barrier on intra-national trade relative to international trade is a factor 10.7 for Canada and 2.24 for the U.S., meaning that Canadian provinces trade 10.7 times as much than Canadian provinces and U.S. states due to an existing country border.[33] This result is described by implication 3 in section 4.2.3, which predicts a greater ratio of intra-national to international trade for smaller countries. Crucial for this result is the much higher impact of border barriers on bilateral trade because of the greater Canadian multilateral resistance.

To sum up, McCallum (1995) estimates that Canadian provinces trade 22 as much with each other that U.S. states do due to an existing border barrier, whereas Anderson and van Wincoop (2003) under consideration of U.S. intra-national trade estimate a factor of 16.4 on the basis of McCallum's gravity equation (not taking into account multilateral resistance). These large border coefficient estimates in the *incorrectly* specified gravity equation are mainly for two reasons.

First, McCallum (1995) does not account for multilateral resistance terms. The absence of multilateral resistance biases the border effect upwards which is reflected in the significant border estimates just mentioned. In the U.S.-Canada example, omitted Canadian multilateral resistance is the crucial factor for the biased estimation of the border effect. McCallum's border coefficient therewith absorbs the impact multilateral resistance imposes on trade patterns, finally resulting in OVB. Since multilateral resistance is correlated with distance and the border barrier, McCallum's estimation of both the border and the distance coefficient is biased. Additionally, Anderson and van Wincoop (2003) specify a computation bias since McCallum neglects changes in the multilateral resistance terms due to omitted variables.

Second, McCallum (1995) compares inter-province trade to province-state trade and defines this as the border coefficient since a country border exists in the latter case. He believes to estimate the impact of border barriers (BB) to a borderless (NB) trade scenario, but actually evaluates an intra-national to

[33] The impact is calculated from Table 4. For Canada: 5.96/0.56=10.64 that is the impact of border barriers of Canadian inter-province trade divided by the impact of border barriers of province-state trade. For the U.S. accounts the following: 1.25/0.56=2.23 (Anderson and van Wincoop, 2003, p. 185).

international trade comparison under border barriers. The structural gravity equation incorporating multilateral resistance yields a factor of 10.7 for the latter. This means that due to border barriers Canadian provinces trade 10.7 times as much as US states and Canadian provinces would trade among each other. Anderson and van Wincoop find that the U.S.-Canadian border reduces trade by *only* 44 per cent when comparing a state of existing border barriers to a borderless scenario (BB/NB). Hence, McCallum (1995) misses a comparative analysis of the *real border coefficient* since he aims at estimating a BB/NB scenario but in reality compares intra-national to international trade under border barriers.

In conclusion, the high border coefficient of McCallum (1995) can be explained by the following reasons:

i. omitted variables - such as multilateral resistance terms which are incorporated in a structural gravity equation,
ii. neglecting the fact that Canada is a relatively small economy - and thus its size influences perceived trade options with other partners and
iii. missinterpretation of the border coefficient – which is defined as the ratio of intra-national to international trade. However, the correct analysis is the comparison of border barrier to borderless trade patterns.

In the following, the significance of the structural gravity model is discussed in the context of international trade.

4.4 Gravity modeling and international trade

The discussion of multilateral resistance terms in the previous sections shows how complex multilateral resistance is. Multilateral resistance allows examining the core of two frequently asked questions in international trade:

How is general equilibrium determined and affected?

Which opportunity costs are implied with the preferred choice?

Furthermore, gravity can be interpreted as sellers' and buyers' incidence of trade. Another section discusses this approach of structural gravity by Anderson and van Wincoop (2003, 2004) to the incidence of trade.

4.4.1 General equilibrium and opportunity costs

Since international trade focuses on the interaction between multiple entities that trade with each other, it is important to review arising changes when trade barriers or prices shift. Dixit and Norman (1980) note that: "(…), one should bear in mind two important points. The first is that the very concepts of trade theory – relative costs and relative prices – call for consistent use of general equilibrium analysis. This need not always be a Walrasian competitive analysis, but in a problem with several goods and factors, and several producing and consuming units, an approach which constantly reminds us of their mutual relationship is essential if errors of oversight are to be avoided. (…) The second point is that micro-economic theory tells us a great deal about general equilibria (…)." (Dixit and Norman, 1980, p. 1).

A general economic equilibrium is complex to achieve. Anderson and van Wincoop (2003, 2004) manage to reduce complexity by the introduction of a CES price index, identical and homothetic preferences, proportional transportation costs and the assumption of trade separability. Supply side market structures are subordinate and not of interest in the model (see section 4.2.1). Anderson (2011) uses *modularity* as a synonym for trade separability meaning that the distribution of goods *between* product classes is superior to the distribution of goods *within* product classes. This means that if there is a price change for steel products (i.e. *within* a product class), expenditure shares on non-steel products are not affected (i.e. *between* product classes), although the overall purchase might be affected. Modulartiy (or trade separability) helps to reduce complexity by focusing on a constant expenditure share between product classes. It simplifies the achievement of general trade equilibrium.

The derivation of the theoretical gravity equation is often called *structural gravity*. Anderson and van Wincoop (2003, 2004) enable to analyze a contextual trade relationship between the bilaterally trading countries and all of their other possible trading partners. Due to

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

bilateral trade flows are dependent on income shares of country *i* and *j*, the bilateral trade barrier separating them and price indices. Price indices are furthermore dependent on

$$\left(P_j\right)^{1-\sigma} = \sum_i \left(\frac{t_{ij}}{P_i}\right)^{1-\sigma} \frac{Y_i}{Y_w}$$

when $t_{ij} = t_{ji}$, meaning that export and import trade costs are the same (see also end of section 4.2.2).[34] Hence, the gravity equation describes a system of full general equilibrium since it takes in the U.S.-Canadian case all 10 Canadian provinces, the selected 30 U.S. states and the other 21 remaining states (grouped together as one region) into account. Therefore, the price index $(P_j)^{1-\sigma}$ is a generated price index of 41 different market-equilibrium conditions which are solved simultaneously. It changes when any trade barriers shift (in this example especially border barriers, as the distance from entity to entity is not likely to change). This does not explicitly require a change of bilateral trade barriers under consideration (between *i* and *j*) but *any* possible shift between country *i* (*j*) to all of its potential trading partners. Therefore, the structural gravity approach allows to give a more complete and profound picture of international trade. The fact that inference draws from general equilibrium – and not from partial equilibria – and a coherent, relatively simple theoretical derivation brings gravity modeling closer to a reasonably justified standing within the field of international economics. Proving theoretical implications 1 to 3 empirically, fortifies the implications of structural gravity and its multilateral resistance approach even more (see section 4.2.3 and 4.3 with Table 3 and 4).

Another concept of international economics is the comparison of opportunity costs. Different countries have differing costs in the manufacturing of goods. Hence, a comparative advantage arises because of distinct technologies or factor endowments (as the Ricardian and Heckscher-Ohlin model explain in chapter 3). Countries decide on the production of goods in which they have a comparative advantage, and therefore a relatively low opportunity cost. Since countries engage into trade, they are most likely interested in purchasing other goods at a relatively low price. This reflects the willingness to search for the best trade opportunity. Countries strive for the lowest possible opportunity cost in purchasing a desired product (Krugman and Obstfeld, 2006, pp. 25-47). For example, Rose and van Wincoop (2001) investigate the relationship currency unions have on trade patterns. Based on the work of Frankel and Rose (1998), Rose and van Wincoop (2001) expect that currency unions lower opportunity costs of national monetary policies (for a detailed discussion see section 5.4). This gives consideration for a more general reasoning of the importance of opportunity costs within the structural gravity model:

i. The question is whether a country prefers one trade option to another because of differing trade costs, i.e. opportunity costs (in this context understood as national monetary regimes versus currency unions).

[34] This is a strong assumption. Anderson and can Wincoop assume this for simplicity reasons to solve the implicit price index function. In the U.S.-Canadian case $t_{ij} = t_{ji}$ is reasonable, since $t_{ij} = b_{ij} d_{ij}^{\rho}$ and is therewith only dependent on the border coefficient and distance. Distance is not likely to change whereas a border coefficient could capture different tax policies of the countries.

ii. Then, it is not the detached option that determines bilateral trade, but the contextual option that influence trade choices. The contextual options are reflected in multiple trade barriers and the belonging relative trade resistances.

Anderson and van Wincoop (2003) achieve to integrate this choice in differing opportunity costs (trade costs) into their structural gravity model. It accounts for all possible trading options and undertakes the *mutual relationship* which determines the most profitable trade option. Trading is eventually a choice from different opportunity costs in trade.

Since the structural gravity equation allows for general equilibrium and consequently is a choice from differing opportunity costs in trade, Anderson and van Wincoop's approach is an adequate and theoretical underpinned model to explaining international trade flows.

4.4.2 Estimation of the incidence of trade costs

In current work, Anderson (2011) describes multilateral resistance as the measurement of average buyers' and sellers' incidence of trade costs. Outward multilateral resistance, ${\Pi_i}^{1-\sigma}$ or ${P_i}^{1-\sigma}$, can be understood as the mark-up on a shipment the exporter makes to the world market. It is the fee the exporter needs to pay to be able to sell his goods on the world market. Inward multilateral resistance can be understood as the buyers' incidence of trade costs, i.e. it is the cost importers need to pay for being able to purchase goods from the world market (Anderson, 2011, p. 142). It is remarkable that Anderson and Yotov (2008) find outward multilateral resistance (or the sellers' incidence of trade costs) to be 5 times larger than inward multilateral resistance. Remote areas are found to face a higher outward and inward multilateral resistance because access to the world market is more complex and costly. Anderson and Yotov (2008) conclude there is no specific industry facing an extremely high or low outward multilateral resistance, since the sellers' incidence of trade costs varies over goods classes. They also develop a Constructed Home Bias (CHB) which indicates a region's predicted internal trade under border barriers in relation to its internal trade in a frictionless world. Multilateral resistance terms help to capture the CHB. Anderson and Yotov (2008) find that outward multilateral resistance falls over time which results in a significant decrease in multilateral resistance of up to 90 per cent for Canadian provinces. CHB is given by

$$CHB_i = \left(\frac{t_{ii}}{\Pi_i P_i}\right)^{1-\sigma}. \tag{51}$$

It is similar to the internal multilateral trade resistance or to the adjustment for trade costs when border barriers exist which is $\left(\frac{t_{ij}}{\Pi_i P_j}\right)^{1-\sigma}$. A direct comparison of intra-regional trade costs may be inaccurate since regions may face different multilateral resistance and therefore relative trade barriers vary: If regional trade costs are equal, i.e. $t_{ii} = t_{jj}$, the CHB is not necessarily the same since $\Pi_i P_i \neq \Pi_j P_j$. In that sense, the CHB is a continuation of Anderson and van Wincoop's structural gravity model since the latter can also be applied to intra-regional trade (trade within provinces or states). Its transferability on other geographical entities and therewith the optional application on internal (intra) and external (inter) trade relations makes it a useful tool. This expresses the elaborated structural characteristic and modularity of the gravity equation.

4.4.3 Discussion

This section comments on estimation problems that may arise when applying structural gravity. It alludes to the quality of the data set in terms of missing observations and aggregation. Furthermore, country fixed effects replacing multilateral resistance terms and the non-observability of multilateral resistance terms are discussed.

A problem in empirical application of the structural gravity model is a data set which contains many zeroes. Typically, zeroes arise when trade flows between countries do not exist. The reasons may be missing observations, disaggregated data on firm or industry level, or that trade falls below a certain threshold and is not reported. There are various approaches of how these zeroes in the data can be theoretically explained as follows.

One way to explain zero trade flows is that none or very little trade between countries exists. It is therefore not eminent to report these flows since they tend to have no great impact. Hence, zeroes can be neglected when estimating. When heteroscedastic error terms exist, gravity modeling with an OLS estimation can adulterate results because OLS builds on a normally distributed error term. Since a lot of small trade flows are not reported, the error term will potentially be greater for small than for large trade flows. Santos Silva and Tenreyro (2006) advise to let error terms follow a Poisson distribution which they find to result in smaller estimated trade costs than an OLS estimation reports. Heteroscedasticity can be diluted by size-adjusted trade that replaces trade flows from country i to j (X_{ij}) with $X_{ij}/Y_i Y_j$ as the dependent variable (Anderson, 2003; 2004; 2011).

Another approach is to assume a choke price. If a good's price is below the choke price (i.e. demand hurdle), the good is in demand, if the price lies above this hurdle consumers do not purchase the good in question. In the case of natural resources, this behavior of pricing and demand characterization is often found. Novy (2010) contributes by deriving a translog expenditure function making it possible to handle a demand of zero.[35] Findings indicate that trade flows are more sensitive to trade costs if the exporter is responsible for only a small share of total imports in the foreign country.

A third way to explain occurring zeros is by assuming that high fix costs make exporting to other countries unprofitable. Thus, zeroes are generated by selection dividing countries in either exporting or non-exporting economies. Since firms in exporting country *i* face high fix costs of exporting to country *j*, only the most productive firms can overcome these fix costs. More productive firms are able to trade internationally since they tend to overcome high fix costs associated with exporting or monopolistic competition, i.e. mark-ups on prices on the domestic or foreign market. Helpman, Melitz and Rubinstein (2008) develop a gravity model that is capable to yield zero one-sided trade flows between countries - meaning that only one country exports to the other - and at the same time accounts for bilateral trade flows between countries exchanging goods frequently. This effect is modeled with the help of the extensive (decision to export) and intensive margin (how much to export) and results in a volume effect determining how many firms export (Santos Silva and Tenreyro, 2008; also see section 4.5.1). The volume effect is based on the productivity of the exporting firms in relation to a lower bound of productivity that must be achieved. Helpman et al. (2008) argue by not including the volume effect, estimated variable trade costs are biased downwards.

A fourth way of explaining non existing trade flows is based on the assumption of a discrete choice structure. Choosing the most profitable option, trading countries often makes a choice that is not observable to the researcher. Nevertheless, what can be observed is that groups of countries have a special tendency of choosing certain trade partners more likely than others. Hence, researchers can condition the probability of choice on these characteristics and develop proxies representing the true reason of choice. Therefore, it can happen that some countries prefer trade with a special group of countries less due to small economic gain (Anderson, 2011, p. 148).

Although Anderson and van Wincoop's structural gravity model with multilateral resistance terms allows for a more precise estimation of trade costs (equations (I) to (III)), traditional gravity remains one of the most used models in international trade (equation (4) and (7)). Since traditional gravity operates with remoteness terms, results are often not comparable to structural gravity in terms of accurateness as shown before by the estimation of the U.S. – Canadian border coefficient. However, both gravity models have limitations concerning aggregation and data level availability.

[35] A translog expenditure function is a transcendental logarithmic function. The equation can be found in Novy, 2010, p. 5, equation (1). In this case, a translog function allows for endogenous price mark-ups (since the setting takes place in a monopolistic competition framework).

Former relates to GDP variables. GDP is implied as an aggregation among industries and sectors in the gravity equation. When trade flows are estimated for a special product class indicated by *k*, equations I to III change to

$$X_{ij}{}^{k} = \frac{Y_j{}^{k} Y_i{}^{k}}{Y_w{}^{k}} \left(\frac{t_{ij}{}^{k}}{P_j{}^{k} \Pi_i{}^{k}} \right)^{1-\sigma_k} \tag{I'}$$

$$\left(\Pi_i{}^{k}\right)^{1-\sigma_k} = \sum_j \left(\frac{t_{ij}{}^{k}}{P_j{}^{k}} \right)^{1-\sigma_k} \frac{Y_j{}^{k}}{Y_w{}^{k}} \tag{II'}$$

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

Since multilateral resistance, elasticity and trade barriers vary between industries and goods classes, also trade flows will differ and industry or sector specific data is needed. However, performing estimation with aggregated data may result in incorrectly estimated trade costs since these differ at an industry specific level. On the other hand, complexity increases when estimating trade costs on a disaggregated level (Anderson, 2004, p. 723).

When multilateral resistance is not included as in traditional gravity modeling, OVB arises. Country fixed effects for the importing and exporting country can account for the exclusion of multilateral resistance, respectively. Country fixed effects have the advantage to take any other omitted variables into account and incorporate them as a country specific effect. However, significance of trade costs may be reduced. Also, the greater the number of countries incorporated in the trade flow analysis, the higher is the number of degrees of freedom which are used in a statistical analysis (Anderson, 2011, p. 151).

When introducing multilateral resistance into gravity modeling, the non-observability of multilateral resistance terms is another concern. The CES price indices are needed for calculating the size of the multilateral resistance terms. In general, price indices cannot be taken as consumer price levels. Anderson and van Wincoop (2003) give reason to this by arguing that non-tradable goods, not present in their structural gravity derivation, have a major influence on consumer price levels as well. Furthermore, exchange rates can temporarily influence the comparison of consumer price levels (Anderson and van Winccop, 2003, p. 179, Footnote 17). As indicated by equation (44), price indices are determined as the sum of imported goods' prices. This also claims a significant availability of data for estimation.

The following section provides an outlook of the most recent aspects of gravity modeling in international economics.

4.5 Other prospects of gravity models

Recent approaches have concentrated on the division of variable and fixed trade costs as well as integrating heterogeneous firms into the gravity approach. Another point of interest mentioned in the following section is demand functions other than CES preferences which Anderson (1979), Deardorff (1998) and Anderson and van Wincoop (2003, 2004) use.

4.5.1 The intensive and extensive margin of trade

McCallum (1995) and Anderson and van Wincoop (2003, 2004) capture trade costs by the border barrier and geographic distance between economic centers. There are many more possibilities of which decisive factors to include when measuring trade costs. Chapter 5 outlines a diverse spectrum of potential trade barriers existing in empirical research. Most researchers interested in the impact of trade costs make no distinction between variable and fixed trade costs, although fixed costs may indicate whether firms tend to export at all.

Chaney (2008) investigates variable and fixed exporting costs under a heterogeneous firm setting. Variable trade costs are associated with the intensive margin. The *intensive margin* describes the *size of exports* each exporter is willing to ship. If trade costs change, so do exports. It may be more or less profitable for firms to export depending on either decreasing or increasing trade costs. On the other hand, not only the intensive margin varies due to changing trade costs, but also the extensive margin. The *extensive margin* relates to fixed costs incurred *when entering foreign markets*. They depend on the extra supply that is generated by an additional number of firms deciding to export.[36] Crucial for the development of the intensive and extensive margin is the change in trade barriers and the elasticity of substitution. Chaney (2008) gives the following example: If trade barriers decrease, this will incentivize new firms and firms being less efficient in their manufacturing of goods to enter the market for exported goods.

i. Scenario 1: Given a high elasticity of substitution, competition among firms on the market is high since consumers are willing to change to competitors' products or to substitutes in case the price of their presently brought product increases. For less productive firms with unattractive pricing, the change of consumers' choice is a disadvantage and they are less likely to gain significant market shares. According to Chaney (2008), less productive firms gain no or negligent market share and are not attracted to market entry.

[36] The intensive and extensive margin is defined as follows: "The intensive margin is defined by how much each existing exporter changes the size of its exports. The extensive margin is defined by how much new entrants export (in the case of a reduction in trade barriers)." (Chaney, 2008, p. 1716).

ii. Scenario 2: Given a low elasticity of substitution, consumers are not flexible enough to change to competitors' products and do not substitute. They want the good produced by their initial supplier and firms do not have to compete as much with each other as in scenario 1. This leaves a market for less productive firms. They are attracted to market entry and have the possibility to gain and maintain large market shares (Chaney, 2008, pp. 1707-1708).

In the case of a high elasticity of substitution, the extensive margin is less sensitive to a change in trade barriers since few market entrants survive or even take the risk of market entry, whereas the intensive margin is more sensitive because of less productive firms gaining only a small market share or even do not produce any export related goods. Hence, the elasticity of substitution has an opposite impact on the extensive and intensive margin. With heterogeneous firms, Chaney (2008) predicts the impacts of trade barriers on trade flows to be larger than within a model that consists of representative firms. He implicitly assumes that fixed costs in the context of market entry are borne by exporters. In the literature, Evans (2001) also expects fixed costs to be compensated for by exporters. She finds that half of the border effect can be explained by the *availability-hypothesis* suggesting not all varieties of goods which are supplied domestically are exported to international markets. Thus, a smaller fraction of goods is traded internationally. Klenow and Rodriquez-Clare (1997) on the other hand derive a model assuming fixed costs are aligned by importers.

It is worth mentioning that Anderson and van Wincoop do not incorporate fixed costs associated with exporting into their gravity equation. Fixed costs can be interpreted as market entry costs to foreign destinations indirectly associated with costs due to cultural dissimilarity, i.e. language barriers, market research and information costs or legal counseling fees. A consideration of the extensive margin is of importance, since the border coefficient determined in Anderson and van Wincoop (2003, 2004) can also be responsible for absorbing entry costs associated with exporting to foreign destinations. Fixed costs of exporting could possible represent OVB captured by the border coefficient.

4.5.2 Demand functions other than CES preferences

There are other possibilities of demand functions rather than a CES utility maximizing consumer. In recent years, research has focused on an increasing competition due to falling transportation costs and therewith on the cutting of mark-ups on goods' prices. Thus, more flexible demand systems than CES are required. According to Novy (2010), a CES preference structure "(…) translates into a constant elasticity of trade with respect to trade costs. It implies that a reduction in trade costs, for instance a ten percent tariff cut, has the same proportionate effect on bilateral trade regardless of whether tariffs were initially high or low or whether a country pair traded little or a lot." (Novy, 2010, p. 2). He derives a gravity equation under translog homothetic preferences in a monopolistic competition setting since translog preferences are more flexible than CES preferences in the sense that they enable modeling substitution from different product alternatives, therewith accounting for price mark-up adjustments and changes in price elasticity.[37] The more varieties consumers have to choose from, the more competition among firms will occur resulting in decreasing price mark-ups and a higher price elasticity of substitution. Trade barriers are non-constant and vary to the degree of how intensively two countries trade with each other. According to Novy (2010), translog preferences account for this flexibility. Empirical application on OECD country trade flows from the year 1991 to 2000 shows strong contradiction to a CES specification. Novy (2010) finds bilateral trade is more sensitive to trade costs the smaller the exporting country's share in total imports of the importing country is. Hence, changes in trade barriers affect trade flows differently depending on the country pair under consideration.

To sum up, the structural gravity model which introduces multilateral resistance has shed light on the border puzzle phenomenon. Due to omitted variable bias, the border coefficient is biased upwards in McCallum's (1995) estimation. The structural gravity model convinces because of its relatively simple system of equations: General equilibrium is described by only three equations (equation (I) to (III)) that have to be solved simultaneously determining equilibrium bilateral trade flows and inward and outward multilateral resistance. Even though deriving the structural gravity model builds on a common utility maximization problem, transformations during the derivation possess a certain degree of difficulty. Anderson and van Wincoop's approach to gravity is convincing because the theoretical implications of their model are confirmed by empirics and furthermore explain the border puzzle. General equilibrium is achieved because trade is no longer solely dependent on bilateral trade flows but is set into a more contextual framework determining the average trade resistance to all other potential trading partners. This takes into account the trade barriers and hence the opportunity costs that arise when choosing with whom to trade.

In previous chapters, trade costs almost only were determined as geographical distance and the border barrier. Nevertheless, Anderson and van Wincoop (2004) mention several kinds of trade costs linked to national borders. They are reported as ad-valorem tax equivalent: 44 per cent of border related trade costs which roughly break down into (i) policy barriers (which amount to 8 per cent), (ii) language barriers (7 per cent), (iii) currency barriers (14 per cent), (iv) information cost barriers (6 per cent) and (v) security barriers (3 per cent). Additionally, local distribution costs (55 per cent) and transportation

[37] The translog gravity equation is complex in explanation and derivation, it can be found in Novy, 2010, p. 7, equation (7).

costs (21 per cent) exist, so that in total a tax equivalent of 170 per cent is found (Anderson and van Wincoop, 2004, pp. 691-693).[38]

The next chapter indicates other areas of research where gravity modeling has been applied indicating its popularity among economists nowadays.

5 Other applications in economics

Gravity modeling has been applied to various economic research topics. As discussed, it is predominantly used to model international trade flows, but recent research has brought attention to the modeling of migration, foreign direct investment (FDI), international portfolio capital movement, WTO membership and other trade agreements, currency unions or even internet traffic and the decay of colonial linkages. The distance and border coefficient often do not capture the full effect of trade costs and so a range of other variables have been identified to extend the standard gravity model. However, in the following choice of selected models, empirical application is more stressed and leaves a theoretical justification often unfounded.

First, the modeling of migration flows, FDI and portfolio investment is demonstrated which represent alternative dependent variables when applying gravity models. This gives a broader view of the gravity equation in economics today. Secondly, focus is shifted on different forms of trade costs that have been at the spotlight of attention when modeling international trade flows. Exemplarily, the effect of currency unions and other trade agreements is discussed as they are often predicted to have a great influence on international trade flows.

5.1 Migration flows

According to Borjas (1989), the decision to migrate is an investment choice from an economic perspective. Given that workers compare wages across immigration destinations, they choose their destination country on the basis of the discounted income surplus between their emigration and immigration country. Additionally, the investment decision has at least to compensate for moving costs or higher tax rates. The country promising the best possible income after costs of moving, taxation and border formalities will be crucial for the final decision.

Migration flows can be described in relation to the source and destination countries' population size and income. Melkumian (2009) identifies these as push and pull factors of migration and develops a

[38] Any further discussion of these empirically determined trade costs are not undertaken since this work focuses solely on a theoretical foundation. However, estimates may give a feeling which trade costs exist. Tax equivalent of 170 per cent is calculated as 1.21 (transportation costs)*1.44 (border related costs)*1.55 (local distribution costs)-1=1.7, see Anderson and van Wincoop, 2004, p. 692).

gravity model of migration that amongst other factors includes political and social components such as the Index of Economic Freedom. The migration flow M_{ij} from country *i* to country *j* can be described as

$$M_{ij} = \frac{\alpha_0 S_i^{\alpha_1} D_j^{\alpha_2}}{R_{ij}^{\alpha_3}} . \tag{52}$$

In equation (52), $S_i^{\alpha_1}$ stands for the push factor, determining the *supply of migrants* originating from country *i*. Variable $S_i^{\alpha_1}$ is defined as $S_i^{\alpha_1} = \beta_0 y_i^{\beta_1} n_i^{\beta_2}$ and is dependent on the population n and income size y of the source country *i*. Likewise, $D_j^{\alpha_2} = \gamma_0 y_j^{\gamma_1} n_j^{\gamma_2}$ is the pull factor representing *demand in migration* of the destination country *j*. The denominator $R_{ij}^{\alpha_3}$ reflects all other possible influences determining migration flows from *i* to *j*, such as income tax levels or moving costs. Melkumian (2009) bases his research on Karemera, Oguledo and Davis (2000) who examine the factors determining migration to Canada and the U.S.. With the help of a gravity equation they estimate migration flows in the period of 1976 to 1987. Findings of Karemera et al. (2000) show that the emigrants' country population size and the immigrants' country income size both have a significant influence on migration flows to North America.

Melkumian (2009) uses the following gravity equation for estimation after taking logarithms of non-dummy variables which sets migration to country *j* in relation to the origin country *i*'s population dependent on the distance d_{ij}, population and income sizes n_i, n_j, y_i, y_j, if source and destination countries share borders ($Border = 1$ if country *i* and *j* are neighbors, 0 otherwise), $MigrantStock_j$ represents the percentage of foreign-born in country *j*, and $English_i$ indicates whether country *i* uses English as a predominant language ($English_i = 1$ if English is a commonly used language, 0 otherwise). Furthermore, $FreedomIndex_i$ stands for the economic freedom in country *i* and is predicted to be positively correlated with migration to country *j*, $Unemployment_j$ is the destination's unemployment rate and finally D indicates a multiple of region dummy variables depending on the region the country of origin belongs to[39]

$$\begin{aligned} \left(\frac{mig}{pop}\right)_{ij} = {} & \beta_0 + \beta_1 ln\, d_{ij} + \beta_2 ln\, n_i + \beta_3 ln\, n_j + \beta_4 ln\, y_i + \beta_5 ln\, y_j + \beta_6 Border \\ & + \beta_7 ln\, MigrantStock_j + \beta_8\, English_i + \beta_9\, ln\, FreedomIndex_i \\ & + \beta_{10} ln\, Unemployment_j + \beta_{11}\, D . \end{aligned} \tag{53}$$

[39] Melkumian (2009) suggests that $\beta_1, \beta_3, \beta_4, \beta_{10} < 0 \; and \; \beta_2, \beta_5, \beta_6, \beta_7, \beta_8, \beta_9 > 0$. The higher the Index of Economic Freedom, the less economic freedom exists in the migration's source country *i*.

Estimating the equation for the time period ranging from 1996 to 2000 for migration to the U.S. from more than 100 different countries, results indicate that variables of population and income have an important effect on migration, as well as the migration stock in destination country j. Consistent with Melkumian's prediction, countries featuring a lower level of economic freedom are associated with a higher tendency of migrating to the U.S.. Thus, he concludes that government incentives exist to provide, establish and maintain efficient policy regulations and a stable political environment. This would help to reduce immigration of the educated work force from countries with a less economic freedom to the U.S., as in the case for China and India. By enhancing economic freedom in the source country, educated workers might have the tendency to emigrate less.

However, considering a structural gravity equation, Melkumian (2009), Karemera et al. (2000) and Borjas (1989) do not take into account a general equilibrium setting. Although a theoretical foundation has been rolled out by Anderson and van Wincoop (2003, 2004) empirical research often does not account for multiple equilibria and views the targeted research question in a more isolated manner and not in a contextual framework.

Anderson (2011) transfers the multilateral resistance approach onto migration. He assumes that migration arises as a discrete choice from several locations. Migrating workers choose to move to the country promising the biggest wage or utility surplus. They face a cost of moving that has to be smaller than the utility surplus. Anderson (2011) derives a structural gravity equation for migration. Here, inward and outward migration terms measure the frictions occurring in relation to migration. They are comparable with the multilateral resistance term in international trade flow gravity models. He derives the following structural gravity equation

$$M_{ij} = \frac{L_j N_i}{N} \left(\frac{\delta_{ij}}{\Omega_j W_i} \right)^{1-\theta} \tag{54}$$

where M_{ij} is the predicted migration flow from country *i* to country *j*, N_i ist the population of country *i*, N is total world labor supply, L_j is the total labor force in country *j* that comes from any region in the world including its own country, with δ_{ij} standing for the iceberg cost factor of migration and θ representing the relative risk aversion. Finally, Ω_j and W_i are the inward and outward migration frictions.[40] As before, the first term represents migration in a frictionless world, now depending on the

---

[40] Inward migration friction is defined as $\Omega_j = \left[\sum_i \frac{(\delta_{ij})^{1-\theta}}{W_i} \frac{N_i}{N} \right]^{1/(1-\theta)}$ and outward migration friction as $W_i = \left[\sum_j \frac{(\delta_{ij})^{1-\theta}}{\Omega_j} \frac{L_j}{N} \right]^{1/(1-\theta)}$. Both are dependent on bilateral migration costs, population shares and labor supply as well as net wage. For a more detailed definition and derivation of inward and outward migration resistance, see Anderson, 2011, pp. 152-155.

countries' labor and population forces in comparison to total world population. The second term introduces the effects on migration in a world with migration barriers and implicates a general equilibrium approach. When comparing equation (54) with equation (I) featuring the structural gravity model, the first term of (54) equals $\frac{Y_j Y_i}{Y^w}$ in (I) and the second term is similar to $\left(\frac{t_{ij}}{P_j \Pi_i}\right)^{1-\sigma}$ in (I). Thus, migration flows are dependent on labor forces in relation to total world population adjusted by multilateral migration frictions, whereas trade flows are dependent on countries' GDP to world income corrected for multilateral trade bresistance.

There are many more variations in capturing migration costs. Other authors interpret migration costs as monitoring costs for workers in a distant country. Thus, Head, Mayer and Ries (2009) develop a gravity descended model which investigates monitoring costs in the service sector. Given that there is a tendency of outsourcing specific business services due to a rapidly improving communication technology, i.e. information technology or call centers, Head et al. (2009) estimate a gravity equation indicating that distance still has a restricting effect on the service provision.

5.2 Foreign direct investment

FDI can take place in the form of greenfield investment or mergers and acquisitions (M&A), the former involves the upsetting of new operational facilities and the latter requiring a suitable takeover target. Generally speaking, investments abroad are an alternative to trading goods across borders. FDI pays off the more frictions constrain trade flow between countries which is represented by high or increasing trade costs. Thus, the trade-off between exporting to a low trade cost environment and investing in a foreign operation through M&A depends on the extent of trade costs (Lasserre, 2007, pp. 191-198). The decision on FDI is similar to the migration approach, since both decisions are location specific. Anderson (2011) mentions one problem concerning location choice: firms are able to diversify their risk by choosing multiple foreign affiliates whereas migrants do not have the possibility to diversify risk by location. They can normally only work in one location at a time.

Head and Ries (2008) develop a gravity equation which models bilateral FDI stocks. Their paper focuses on M&A since research shows more than 65 per cent of FDI between the years 1987 to 2001 was invested in existing foreign assets through M&A. Frictions in bilateral FDI flows can be interpreted as asymmetric information between the foreign investor and the takeover target. Gordon and Bovenberg (1996) give several reasons of why asymmetric information occur such as real interest differentials between countries, a scarce or missing international portfolio diversification and the tendency to invest in the source country where savings have been generated which is comparable to the home bias in international trade.

By considering other investors' locations, Head and Ries (2008) apply a general equilibrium approach that takes into account multilateral resistance: the M&A of a foreign asset by an investor depends on the distance between the investing and the invested country as well as on the bidder's location compared to the offices of other potential bidders. Thus, a successful M&A depends not only on the bidder's and its target's characteristics but also on the geographical location of other competing bidders. Their general equilibrium approach is designed in a three stage process. First, Head and Ries (2008) undertake a cost-benefit analysis of the new owner monitoring the foreign subsidiary from abroad. Monitoring or costs of control increase in distance. Under this assumption, expected bilateral FDI stock $E[F_{ij}]$ from originating country *i* to destination country *j* is given by

$$E[F_{ij}] = \exp\left(\frac{\mu_i}{\sigma} + lns_i^m + lnK_j - lnB_j - D_{ij}\theta\right) \quad (55)$$

where s_i^m expresses country *i*'s share of internationally existing bidders *m* around the world interested in the target takeover domiciled in country j, K_j represents the asset value of the possible takeover targets in country j, B_j signals the bid competitiveness relating to country j and $D_{ij}\theta$ is a reference term to distance between bidding country i and hosting country j multiplied with a trade limiting parameter θ.[41] The first term $\frac{\mu_i}{\sigma}$ stands for the location parameter μ_i divided by the shape parameter σ of the Gumbel distribution.[42] It indicates a mean ability of the firm to monitor its potential takeover target. Positive and negative signs in the equation above indicate the effect each of the variables is predicted to have on bilateral FDI stocks. The first two terms can be interpreted as outward effect $O_i = \frac{\mu_i}{\sigma} + lns_i^m$, consisting of the ability to control and the share of country *i*'s bidders in relation to bidders from around the world. The outward effect derives from the investing country i's features. Likewise, the inward effect can be summarized by the asset value and the bid competitiveness effect $I_j = lnK_j - lnB_j$ which is determined by the FDI destined country *j*. Then, Head and Ries (2008) draw implications for multilateral resistance FDI. Therefore, worldwide total FDI stocks are obtained, as well as predicted inward and outward FDI stocks of each country. Predicted inward FDI from worldwide investors *w* into country *j* is given by

---

[41] For a more detailed calculation of the parameters in equation (55) refer to Head and Ries, 2008, pp. 8-11. The bid competitiveness is defined as $B_j = \sum_l s_l^m exp[\frac{\mu_l}{\sigma} - D_{lj}\theta]$.

[42] The Gumbel distribution is used to model the maximum of an object under consideration. In this case it is used because the expected bilateral FDI stocks depend in the first place on $E[F_{ij}] = \pi_{ij}K_j$ where π_{ij} is the takeover probability of a successful M&A in country j by a firm residing in country *i*. The probability of a successful takeover (π_{ij}) depends on the maximum of bids that is achieved in country *i*. The Gumbel distribution belongs to the extreme value theory which often features exponential distributions, explaining the functional form of equation (55).

$$E[F_{wj}] = \sum_{i \neq j} E[F_{ij}] = K_j \sum_{i \neq j} \pi_{ij} = K_j(1 - \pi_{jj}) \quad (56)$$

where π_{ij} is the takeover probability of a successful M&A in country *j* by a firm residing in country *i*. Since worldwide FDI stocks are the sum of all inward FDI flows directed at different countries, this can be expressed as

$$E[F_w] = \sum_j E[F_{wj}] = \sum_j (K_j - \pi_{jj} K_j) = K_w - \sum_j \pi_{jj} K_j \quad (57)$$

and finally total outward FDI from country *i* directed to other countries is

$$E[F_{iw}] = \sum_{j \neq i} E[F_{ij}] = \sum_{j \neq i} \pi_{ij} K_j = K_i(A_i - \pi_{ii}).\text{[43]} \quad (58)$$

From equations (56)-(58), the authors derive a benchmark for predicted multilateral inward and outward directed FDI which is

$$f_i^I = \frac{E[F_{wi}]}{E[F_w]} = s_i^K \frac{1 - \pi_{ii}}{1 - H} \text{ and } f_i^O = \frac{E[F_{iw}]}{E[F_w]} = s_i^K \frac{A_i - \pi_{ii}}{1 - H}\text{ .[44]} \quad (59)$$

The index *H* stands for the Herfindahl index which is a measurement of market concentration (Pepall, Richards and Norman, 2008, pp. 44-47). In this context it is used as an indication of worldwide GDP distribution. Head and Ries (2008) state that their FDI benchmark distinguishes itself from conventional benchmarks, such as the one used by the United Nations Conference on Trade and Development (UNCTAD) who only take into consideration each country's share of bidders s_i which is proportional to country GDP in relation to overall world GDP. If no transportation costs exist, meaning $\theta = 0$ and $f_i^I = f_i^O = s_i \frac{1-s_i}{1-H}$, one can see that Head and Ries (2008) amend for the country size s_i which is set into relation with the concentration index $\frac{1-s_i}{1-H}$, making a more accurate estimation possible. UNCTAD only corrects for the country size s_i. This demonstrates again the context-sensitive characteristics of a multilateral resistance approach. The multilateral FDI terms enable to account for country fixed effects and view each country's FDI attractiveness into a contextual framework of other possible FDI options. Head and Ries (2008) suggest that a general conclusion about small economies being an FDI-haven due to remarkable, competitive international enterprises residing in these countries must be reviewed under a multilateral FDI set-up.

[43] $A_i = \sum_j \pi_{ij} \frac{K_j}{K_i}$ represents an advantage for the bidder in country *i* if $A_i > 1$ and a disadvantage if $A_i < 1$.

[44] The probability of a successful M&A in country *i* by a firm residing in the same country is $\pi_{ii} = s_i^m \exp(\frac{\mu_i}{\sigma} - D_{ii}\theta)B_i^{-1}$ and $s_i^K = K_i/K_w$, representing the stock of capital held in country *i* in relation to the world capital stock. See Head and Ries, 2008, p. 11 for detailed derivation.

They estimate results in three categories: for OECD country members, for partners associated with OECD countries and non-OECD countries. In the time period from 1987 to 2001, OECD countries accounted for 80 per cent of GDP, 75 per cent of FDI and around 90 per cent of all M&A transactions. M&A was to 80 and 70 per cent predominant in regards to inward and outward FDI. The FDI outward effect $O_i = lns_i^m + \frac{\mu_i}{\sigma}$ is approximated by $O_i = \alpha_1 lnN_i + \alpha_2 lny_i$, where the bidder's share of country i is reflected by i's population and the ability to monitor the takeover target is represented by GDP per capita of i. Results indicate that especially the ability to monitor the takeover target is significantly different than 1 after controlling for population size. This argues for the importance of a country's development determining the source of FDI in the first place. Both variables account for 88 per cent of unexplained variance. Likewise, the FDI inward effect $I_j = lnK_j - lnB_j$ is approximated by capital stock of country j in the year 1990 and bid competitiveness B_j. As before, the ability to monitor the takeover target is significantly different than 1 and has a dominant effect on inward FDI. In the multilateral case, Head and Ries (2008) compare their outward and inward FDI benchmark to the UNCTAD benchmark, adjusting FDI to world GDP shares and to the country-size adjusted Herfindahl index as discussed above under the assumption of zero transportation costs. By comparing expected and actual outward and inward FDI figures, the Head and Ries calculation gives smaller deviations from actual values in most of the cases than other methods of measurement. This indicates the superior construction and elaboration of gravity models which incorporate multilateral resistance effects.

5.3 Portfolio investment

Portes and Rey (2005) provide research on the relation between international portfolio investment and the market size of the investing and invested country as well as on information and transaction technology. They use a gravity similar equation to explain trade in assets of 14 different countries representing major equity markets in the U.S., Europe and Asia between the years 1989 to 1996. The emphasis is on whether distance inhabits a negative effect regarding portfolio diversification. Empirical results indicate that geographical distance bears an exceptionally high negative effect on foreign portfolio investment. At first, Portes and Rey (2005) find this surprising since theory suggests the diversification of assets could even seek greater distance because countries further apart might have oppositional business cycles to the sourcing economy and rates of return should be negatively correlated or should not be correlated at all. Equally, portfolio investment does not depend on transportation costs as traded goods do. They conclude: "The most natural explanation is that informational frictions are positively correlated with distance. Geographical distance is a barrier to interaction among economic agents and, more broadly, to cultural exchange." (Portes and Rey, 2005, p. 270). Hence, a set of variables have been determined which try to represent and model the information flow between the investing economy and the foreign destination country of asset allocation. Variables include telephone call traffic, overlap in trading hours of the country pairs, the number of multinational bank branches and an index of insider trading that resembles asymmetric information between national and international investors. The gravity model encountered in this approach is motivated by the home bias in portfolio investment as discussed in French and Poterba (1991). They assume investors expect their national equity markets to yield a higher return on investment than a foreign asset allocation can deliver. Obstfeld and Rogoff (2001) also address home bias issues in various facets (see section 4.1). For the asset allocation model developed, Portes and Rey (2005) use the following gravity equation

$$\begin{aligned}\log\left(T_{ij,t}\right) = \alpha_0 + \alpha_1 \log\left(mktcap_{i,t}\right) \\ + \alpha_2 \log\left(mktcap_{j,t}\right) + \alpha_3 \log\left(distance_{ij}\right) \\ + \alpha_4 \log\left(teleph_{ij}\right) + \alpha_5 \log\left(bank_{ij}\right) + \alpha_6 \log\left(overlap_{ij}\right) \\ + \alpha_7 \log\left(insiders_j\right) + \alpha_8 \log\left(soph_i\right) + \alpha_9 \log\left(soph_j\right) + \alpha_{10} \log\left(covar_{ij}\right) \\ + time\ dummies + \varepsilon_{ij,t}\ .\end{aligned} \tag{60}$$

$T_{ij,t}$ represents the transactions in equities from country i to country j at time t, $mktcap$ stands for market capitalization, $distance_{ij}$ measures the distance between the capital in i and j, $bank_{ij}$ is the number of bank offices in country j of banks which have their headquarter in country i, $overlap_{ij}$ accounts for the hour quantity trading hours in the two countries overlapping, $insiders_j$ stands for the extent to which insider trading is executed in the destination country j, $soph_i$ indicates the sophistication of the capital market i, $covar_{ij}$ measures the covariance between stock market returns in the country pair and finally $time\ dummies$ represents time fixed effects and $\varepsilon_{ij,t}$ is the error term. The coefficients α_1, α_2 are expected to be close to 1, α_3, $\alpha_7, \alpha_{10} < 0$ having a negative effect on transactions in equities and variables $\alpha_4, \alpha_5, \alpha_6, \alpha_8, \alpha_9 > 0$ claiming a stimulating effect on international asset allocation. Since all variables are in logarithmic form, coefficients must be interpreted as elasticity. Results show that coefficients have the expected sign except the coefficient on insider trading which to some extent fluctuates between a positive and negative sign. The influence of distance in a prior equation, only featuring market capitalization and geographical distance between capital cities, is dramatically reduced when adding $teleph_{ij}$ to the equation, a first variable that represents the information flows between the two countries of interest. With all other information and transaction variables, up to 70 per cent of variation in the data can be explained. When distinguishing between different forms of capital investment - government bonds, corporate bonds and equity - Portes, Rey and Oh (2001) find that distance does not determine investments in government bonds whereas it plays a role in the latter two forms of asset allocation. Theory suggests this is due to the information content needed for validating the different investment options.

In summary, Portes and Rey (2005) conclude that the diversification motive is suppressed by information asymmetries which are a decisive factor for international asset allocation in non-frictionless capital markets. A determining factor is hereby the home bias in asset allocation as investors expect higher returns from domestic investment classes.

More recent research has focused on the effect which society imposes on equity flows. Aggarwal, Kearney and Lucey (2012) introduce a cultural dimension to foreign portfolio investment modeling. Cultural differences prior to that have been considered primarily from an economic point of view as increasing information and transaction costs. The authors contribute by separating culture from its solely economic understanding and describe culture in relation to Hofstede's concept of individualism, masculinity, power distance and uncertainty avoidance (Hofstede, 2001), and Kogut and Singh's cultural distances (Kogut and Singh, 1988).[45] They argue that besides the well acknowledged impacts of return, risk and time on portfolio investment, culture has a considerable influence due to incomplete financial contracts. Incomplete contracts force financial agents to revert to social norms and manners which are unique to each country or cultural group and likewise determine financial investment. Aggarwal et al. (2012) conduct an empirical study where they focus on long-term debt and equity and include the cultural dimensions mentioned above as well as common language, legal system origin and religion (institutional variables). By testing a culturally extended gravity model against a standard gravity model which incorporates only geographical distance and institutional variables, the authors conclude culture matters in the allocation of foreign portfolio investment. Their main finding is that countries with a high degree of masculinity, individuality and power distance - to which foreign portfolio investment is directed at - tend to increase cross-border debt and equity holdings.

As much as the gravity approach is used in modeling international asset allocation, focus has been on the bilateral equity flow relationship. Multilateral resistance and consequently investment modeling in a contextual framework has been neglected so far. Furthermore, after discussing other research areas of interest where gravity modeling is applied, the focus shifts on the importance of currency unions and trade agreements.

[45] A high degree of individualism describes a society where individual versus group decision making is preferred and that society values the contribution and success of each individual. This results in a number of diverse investment opportunities and encourages a heterogeneous society in terms of wealth distribution. Masculinity implies attributes such as competition, confrontation and independence in contrast to support, cooperation and social engagement. Power distance relates to how society thinks power is distributed. A high power distance puts power in the hands of few members, i.e. localized decision making or bad access to education. Uncertainty avoidance describes the degree of risk society is willing to take in terms of personal achievement, company accounting standards or resistance to change (Aggarwal et al., 2012, pp. 527-529).

5.4 Currency unions

As effects of market size, distance, tariffs and quotas, language barriers or other cultural effects are primarily of interest when determining international trade flows, the list of influential variables has been extended by the fact of whether countries belong to a currency union or possess their own domestic currency. This is of current interest since the recent debt and Euro crisis is not only targeting countries which are primarily affected in terms of economic slowdown, e.g. Portugal, Ireland, Greece and Spain, but it is the entire European Community searching for a new economic identity. Financial help and rescue packages, such as the European Financial Stability Mechanism and the European Stability Mechanism, are at the core of discussion whether the present European monetary union is financially sustainable. Economists claim that the removal from a national and individual monetary policy which enables an idiosyncratic smoothing of economic downturns is one big cost in regard to a currency union establishment.

Rose and van Wincoop (2001) empirically estimate the performance of monetary unions on trade flows. They use 5-year period data from the years 1970 to 1995 of over 200 countries responsible for more than 98 per cent of world trade. Besides a currency union dummy that indicates whether countries share the same currency with any other nations, the logarithms of distance and countries' real GDP and dummies for a common language, free trade agreements, common border, political union, common colonizer and ex-colony dummy add to the gravity equation to estimate trade flows that takes a similar form as equation (4) estimated with OLS. After adding country fixed effects to account for country specific characteristics, results indicate a 230 percent increase in trade for countries in currency unions. Rose and van Wincoop (2001) find this rather implausibly high and do not assume lower transaction costs being the only reason. They estimate the coefficient on currency unions with a structural gravity equation incorporating multilateral resistance terms (see equation (I)). They expect the currency union effect will be smaller between currency union members who traded prior to the joint currency. Reason for relatively high prior trade exchange is the countries' proximity and/or a former trade agreement. Both reasons are plausible in the case of the EU. According to Rose and van Wincoop (2001), the creation of a monetary union will therefore increase trade between member states by only a small fraction. Concerning multilateral resistance, they expect trade costs to decrease due to omitted transaction cost which is expressed in a decline of the multilateral resistance by a fall in the price indices $P_i P_j$. Re-estimation with the help of the structural gravity equation for data on more than 140 countries during the period from 1980 to 1990 yields a decrease from 230 per cent to 58 per cent for countries in the European Monetary Union (EMU), meaning that an EMU increases bilateral trade by 58 per cent. They even explicitly estimate an EMU including Greece yielding a trade increase by 59 per cent and overall welfare increase by 11 per cent.[46]

Rose and van Wincoop (2001) conclude that currency unions have the strength to reduce multilateral resistance and trade barriers respectively and thrive for an enhanced trade between and welfare in all member countries. They allude back to Frankel and Rose (1998), who claim that business cycles of economies with closer trade links seem to approximate each other and align. According to this, the characteristics of symmetric business cycles make members more eligible to join an optimum currency area which on the other hand fosters the positive correlation and synchronization of business cycles even more. Another criterion, especially for small open economies to join a currency union, is the improvement of credibility which can target an undesired high inflation as emphasized by Alesina and Barro (2002).

More than ten years after the EMU has been established, findings are ambivalent. The EMU with its joint monetary policy stands in contradiction to the country specific fiscal policies. Furthermore, the common price policy makes it impossible for prices to adjust on a national level and adapt and harmonize with country individual business cycles. Labor mobility is low contradicting the optimum currency area theory which suggests that member states will benefit from a currency union if factor mobility is high between the members and asymmetric shocks are little to them. To reach a synchronization of business cycles, there is still a far way to go, since synchronization of economic activities has dropped since 2003. Also, financial markets need to be integrated (Centre for European Economic Research, 2008).

With empirical findings promoting a trade and welfare increase for the EMU on the one hand and latest developments of the debt and euro crises on the other hand, the future of the EU will be subject to European policy decisions and a balancing of joint advantages from a currency union or benefits of country specific monetary policies.

5.5 Trade agreements

Another point of interest in the modeling of trade flows has been whether countries belong to a preferential trade agreement. The predominant trade agreement is a membership in the WTO, its ancestor being the General Agreement on Tariffs and Trade. The WTO currently features 155

[46] Results are estimated with country fixed effects representing the individual multilateral resistance of each country under the assumption that the fall in trade barriers complies with the reduction of trade barriers for countries in existing monetary unions. The structural gravity equation allows estimating any country-pair combination, even if these countries are not targeted to form a currency union. Welfare implications are estimated as the average percentage increase in the price index ($\frac{1}{P_i^2}$) under the assumption that elasticity is equal to σ=5 (Rose and van Wincoop, 2001, p. 389).

members (latest figures reported on 10 May 2012; WTO, 2012). Rose (2004) investigates the effect of multilateral trade agreements on international trade flows for 175 countries over a period of 50 years. He uses a standard gravity equation which enables to incorporate a WTO dummy, indicating of whether both or only one country is holding membership. Additionally, Rose (2004) adds a variable that reflects the existence of a Generalized System of Preferences (GSP). A GSP is an exemption of the Most Favored Nation Treatment under WTO regulation. The Most Favored Nation Treatment stipulates member states to apply the negotiated tariffs, belonging to their most preferred trading partners, to all of their other trading partners. The GSP exempts member states from this to benefit the least developed countries and to authorize lower tariffs and more favorable agreements to those. To control for as many possible other effects, distance, market sizes, culture, colonial linkages, geographical features and country fixed effects are included in the gravity model.

Results show the usual gravity effects, i.e. the negative distance coefficient and positive market size effects. Countries being member of a regional trade agreement also tend to trade more with each other. The coefficients of speaking the same language and sharing the same border are also positive as is the linkage to colonial heritage. Trade flows are negatively related to non-coastal countries as well as to relatively large countries in terms of geographic size. Surprisingly, Rose (2004) finds that membership in the WTO of both countries or only one country does not have a significant positive effect. However, the coefficient on GSP is positive and significant; GSP is expected to raise trade by more than 130 per cent. According to that, membership in the WTO does not necessarily increase bilateral trade flows, it however allows least developed nations to fall under the GSP which does more than double country-pair trade. Results are mentioned to be independent from an industry specific decomposition since they only take total trade flows into account. Rose (2004) suggests future research must focus on industry level trade as trade liberalization has been less advanced in the agriculture and textile sector. This might be decisive and influence results. The analysis of trade in services is another research area that experienced little attention so far.

Carrer (2006) focuses on regional trade agreement effects. She specifies variables into *trade creation* and *trade diversion* effects which she estimates for cross-sectional and panel data both yielding different results. In both data sets, a gravity equation is used. Estimation shows that a panel specification achieves more reliable coefficient estimates as they suffer less from OVB. Regional trade agreements are found to have a trade enhancing characteristic. Carrer (2006) bases her gravity equation on Baier and Bergstand (2002) who assume monopolistic competition and Dixit-Stiglitz preferences (see section 3.2). In the case of cross-sectional data, her gravity equation could incorporate multilateral resistance terms which are interpreted as price levels of the exporting and importing country. However, due to the fact that panel data includes time variables, neither fixed effects nor nonlinear systems of estimation for multilateral resistance can be used when aiming at the same method of estimation applied to both cross-sectional and panel data. Hence, multilateral resistance is approximated by remoteness terms which are comparable to the ones mentioned in section 4.2. Unfortunately, price levels of the trading countries are therewith excluded and do not add to the empirical results. Consequently, multilateral resistance seems to break into the theoretical underpinning of many researchers' models, nevertheless empirical estimation has its difficulty with a proper implication of how to integrate and apply multilateral resistance.

Other trade relations of interest are the colonial past of certain countries. It is assumed that former colonizing countries and their newly discovered territories were bonded by agile trading; oftentimes the expansion of European nations was motivated by gaining access to foreign resources and markets. Considering the relationship of former colonies and their colonizers as some kind of special trade agreement, colonial bonding is likewise reviewed as trade determining factor. Thus, Head, Mayer and Ries (2010) investigate the development of trade between former colonies and their authorizing countries suggesting that trade during colonialism is significantly stronger than trade after independence of the colony. Results indicate that more than 40 years later, trade decreased by 65 per cent. Head et al. (2010) analyze the years reaching from 1948 to 2006 for a wide range of colonies and their former empires which results in observation of more than 200 countries. The decreasing trade relation does not depend on the event of a pleasant or antagonistic ending of colonialism and can be observed throughout country pairs. According to Head et al. (2010), mainly four reasons deliver explanations to why the break-up of colonial linkages adds to a decrease in trade: the economic slowdown of the newly independent country, establishment of an own currency, the former colonizer joins other trade agreements such as the WTO and therewith channels import and exports into other countries and finally, the collapse of governmental institutions of the former colony may negatively affect trade costs. To make a comparison between trade in the colonial and post-colonial state possible, a benchmark of trade is established that estimates trade volumes if independence had not been achieved. This benchmark is determined with the help of a gravity equation. Due to the complexity of the data (missing observations and therefore zeroes in the data), Head et al. (2010) develop *tetradic terms* since the estimation of multilateral resistance through country year fixed effects would result in including several thousand dummies. The tetradic terms represent "the ratio of ratios of trade flows" implying trade relations of four countries helping to reflect multilateral resistance terms (Head et al., 2010, pp. 3-4). The country and time specific fixed effects are therewith soaked up by the tetradic terms. This method empirically proves that a decrease in trade has no exporter, importer or time specific reason but rather develops because of the step of the former colony into independence.

The given examples show research strives to implement multilateral resistance into empirical application with a theoretical underpinning taking multilateral resistance terms for granted. Anderson and van Wincoop's development of a profound contextual framework is not only theoretically accepted and used; it also changes empirical research by relying on estimation methods capturing multilateral resistance.

6 Conclusion

This work attempts to give an overview of the theoretical foundation and developments which have brought gravity modeling into mainstream economics in recent years. Specifically, the incorporation of multilateral resistance terms have been of great importance since they enable a general equilibrium setting which derives from the choice and comparison of opportunity costs in international trade settings.

Prior to the development of multilateral resistance by Anderson and van Wincoop (2003, 2004), a diversity of research approaches existed. These refer to the three classic models in international trade. By reviewing the Ricardian model, focusing on differences in production technology and comparative advantage, the Heckscher-Ohlin model, known for differences in factor endowment and explanation of inter-industry trade as well as the Helpman and Krugman approach, which takes monopolistic competition and intra-industry trade into consideration, all approaches focus on different assumptions. Nevertheless, they level the ground for a more profound analysis in the sense that they do not limit the derivation of a theoretical justification to any specific classic trade theory model.

With the help of inward and outward multilateral resistance, Anderson and van Wincoop develop a new approach towards the border puzzle, first described by McCallum (1995). His work suggests that Canadian provinces trade more than 20 times as much among each other than Canadian provinces and U.S. states. McCallum (1995) explains this phenomenon with the existence of country borders. Borders seem to have a significant effect on trade patterns even if the trading countries have a similar economic understanding, social values and stem from the same cultural heritage. Anderson and van Wincoop (2003, 2004) reveal that prior findings were adulterated by omitted multilateral resistance, which they resolve by introducing multilateral resistance terms. Likewise, the fact of Canada being a small open economy influences its export dominant trade behavior due to the restricted size of its potential domestic market. And eventually, the researchers claim a false interpretation of the border coefficient by McCallum defined as the ratio of intra-national to international trade. They recommend a comparison between a border barrier situation to a frictionless scenario. Their findings indicate that trade between Canada and the U.S. drops by 44 per cent when a change from a border barrier to a frictionless world occurs.

It is the theoretical derivation of a structural gravity approach which convinces in simplicity based on CES preferences and balanced trade assuming each country can only spend what it earns by trading goods. Anderson and van Wincoop's assumptions enable the theoretical derivation of multilateral resistance due to a few mathematical transformations. The insight gained on my behalf is that the clever and reasonable incorporation of price indices makes it possible to determine a structural gravity model expressed in equations (I) to (III). These need to be solved simultaneously to describe a general equilibrium. This puts the traditional gravity equation into a more contextual framework. Trade is no longer solely dependent on bilateral trade flows but on multilateral trading relations between countries. Since inward and outward multilateral resistance needs to be determined simultaneously, trade flows between third parties also influence the general equilibrium of trade. Hence, the structural gravity equation suggests a more holistic view, which prior to Anderson and van Wincoop was barely exploited. Derivations prior to the structural gravity model were either incomplete and partial or very complex. Hence, the structural gravity equation is a step towards a comprehensive, all-encompassing and context-sensitive estimation of international trade flows. The reasons just mentioned account for a profound theoretical foundation and its application in economics today and answer the research question of this work.

Although current research is aware of multilateral resistance when developing the underlying theoretical model the research interest is based on, empirical application and estimation is still insufficient. Often, data is either not detailed enough - and thus biasing effects would not occur when analyzing industry or goods specific data - or incomplete due to missing observations and an imperfect coverage of trade relations. Using country fixed effects is another more standardized technique of capturing country specific effects; other researchers develop further instruments permitting the measurement of multilateral resistance by constructing tetradic terms as Head et al. (2010) demonstrate.

The gravity model has been predominantly applied to explain international trade flows. Amongst multiple variables, the effect of currency unions and trade agreements has been of recent interest due to the current European currency and debt crisis. The gravity equation's flexible and far outreach with regards to the variables of interest enables its application to various research areas. It is also used to model other flows in international economics such as migration, foreign direct investment and international capital movement.

The understanding I gained from this work is the exposition to a relatively new research approach and the comprehension and reconstruction of its theoretical derivation in depth. This might not bring forward research in itself. However, acquiring detailed knowledge and not solely considering *what* is the innovative milestone of an existing research approach but also *why* it is innovative, demanded a much broader and at the same time deeper examination of the subject and its consequences on other research areas.

Further research in gravity modeling could focus on manifold assumptions, other than CES preferences and trade separability. Equally, the determination of price indices and overall trade effects estimation could be improved by gaining access to industry specific data. The extensive and intensive margin of trade focusing on the determination of fixed and variable trade costs might be a further research area to spend more attention on. Empirical application is capable of developments to integrate multilateral resistance into the gravity equation more easily. Hence, further work could improve and enhance research findings on trade effects which can be essential and of great interest for international operating companies and international policymakers.

7 Bibliography

Aggarwal, Raj, Kearney, Colm and Lucey, Brian (2012): "Gravity and culture in foreign portfolio investment," Journal of Banking & Finance, vol. 36(2), pp. 525-538, February.

Alesina, Alberto and Barro, Robert J. (2002): "Currency Unions," Quarterly Journal of Economics, vol. 117(2), pp. 409-436, May.

Anderson, James E. (1979): "A Theoretical Foundation for the Gravity Equation," The American Review, vol. 69(1), pp. 106-116, March.

Anderson, James E. (2009a): "Gravity, Productivity and the Pattern of Production and Trade," NBER and Boston College, January.

Anderson, James E. (2009b): "The Incidence of Gravity," NBER and Boston College, January.

Anderson, James E. (2011): " The Gravity Model," The Annual Review of Economics, vol. 3(1), pp. 133-160, September.

Anderson, James E. and van Wincoop, Eric (2003): "Gravity with Gravitas: A Solution to the Border Puzzle," The American Economic Review, vol. 93(1), pp. 170-192, March.

Anderson, James E. and van Wincoop, Eric (2004): "Trade Costs," Journal of Economic Literature, vol. 42(3), pp. 691-751, September.

Anderson, James E. and Yotov, Yoto V. (2008): "The Changing Incidence of Geography," Boston College and NBER, Drexel University, October.

Anderson, James E. and Yotov, Yoto V. (2011): "Spezialization: Pro- and Anti-Globalizing, 1999-2002," NBER and Boston College, April.

Armington, Paul S. (1969): "A Theory of Demand for Products Distinguished by Place of Production," Staff Papers – International Monetary Fund, vol. 16(1), pp. 1159-178, March.

Baier, Scott L. and Bergstrand, Jeffrey H. (2002): "On the Endogeneity of International Trade Flows and Free Trade Agreements," American Economic Association annual meeting, August.

Bergstrand, Jeffrey H. (1985): "The Gravity Equation in International Trade: Some Microeconomic Foundations and Empirical Evidence," The Review of Economics and Statistics, vol. 67(3), pp. 474-481, August.

Bergstrand, Jeffrey H. (1989): "The Generalized Gravity Equation, Monopolistic Competition, and the Factor-Proportions Theory in International Trade," The Review of Economics and Statistics, vol. 71(1), pp. 143-153, February.

Borjas, George J. (1989): "Economic Theory and International Migration," International Migration Review, vol. 23(3), pp. 457-485, Autumn.

Carrer, Celine (2006): "Revisiting the effects of regional trade agreements on trade flows with proper specification of the gravity model," European Economic Review, vol. 50(2), pp. 223-247, February.

Centre for European Economic Research (2008): “Study on economic integration and business cycle synchronization – Project for the EU Commission Bureau of European Policy Advisers BEPA-01/2007-PO,” Zentrum für Europäische Wirtschaftsforschung GmBH, Mannheim and Vienna.

Chamberlin, Edward (1962): “The theory of monopolistic competition: A re-orientation of the theory of value,” Harvard University Press, Cambridge.

Chaney, Thomas (2008): “Distorted Gravity: The Intensive and Extensive Margins of International Trade,” The American Economic Review, vol. 98(4), pp. 1701-1721, September.

Deardorff, Alan V. (1982): “The General Validity of the Heckscher-Ohlin Theorem,” The American Economic Review, vol. 72(4), pp. 683-694, September.

Deardorff, Alan V. (1998): “Determinants of Bilateral Trade: Does Gravity Work in a Neoclassical World?,” in J.A. Frankel (ed.): The Regionalization of the World Economy, pp. 7-22, University of Chicago Press, Chicago.

Dixit, Avinash K. and Stiglitz, Joseph E. (1977): “Monopolistic Competition and Optimum Product Diversity,” The American Economic Review, vol. 67(3), pp. 297-308, June.

Dixit, Avinash K. and Norman, Victor D. (1980): “Theory of International Trade: A Dual, General Equilibrium Approach,” Cambridge University Press, Cambridge.

Eaton, Jonathan and Kortum, Samuel (2002): “Technology, Geography, and Trade,” Econometrica, vol. 70(5), pp. 1741-1779, September.

Evans, Carolyn L. (2001): “ Border Effects and the Availability of Domestic Products Abroad,” Working Paper, Federal Reserve Bank of New York, New York.

Evenett, Simon J. and Keller, Wolfgang (2002): “On Theories Explaining the Success of the Gravity Equation,” Journal of Political Economy, vol. 110(2), pp. 281-316, April.

Frankel, Jeffrey A. and Rose, Andrew K. (1998): “The Endogeneity of the Optimum Currency Area Criteria,” The Economic Journal, vol. 108(449), pp. 1009-1025, July.

French, Kenneth R. and Poterba, James M. (1991): “Investor Diversification and International Equity Markets,” The American Economic Review, vol. 81(2), pp. 222-227, May.

Gordon, Roger H. and Bovenberg, A. Lans (1996): “Why Is Capital So Immobile Internationally? Possible Explanations And Implications For Capital Income Taxation,” The American Economic Review, vol. 86(5), pp. 1057-1075, December.

Grogger, Jeffrey and Hanson, Gordon H. (2011): “Income maximization and the selection and sorting of international migrants,” Journal of Development Economics, vol. 95(1), pp. 42-57, May.

Grubel, Herbert G. and Lloyd, Peter J. (1975): “Intra-industry Trade: The Theory and Measurement of International Trade in Differentiated Products,” Macmillan, London.

Gujarati, Damodar N. and Porter, Dawn C. (2009): “Basic Econometrics,” McGraw-Hill, New York.

Head, Keith (2003): “Gravity for Beginners,“ Mimeography, University of British Columbia.

Head, Keith and Ries, John (2008): “FDI as an outcome of the market for corporate control: Theory and evidence,” Journal of International Economics, Elsevier, vol. 74(1), pp. 2-20, January.

Head, Keith, Mayer, Thierry and Ries, John (2009): “How Remote is the Off-Shoring Threat?,” European Economic Review, vol. 53(4), pp. 429-444, May.

Head, Keith, Mayer, Thierry and Ries, John (2010): “The erosion of colonial trade linkages after independence,” Journal of International Economics, vol. 81(1), pp. 1-14, May.

Heckscher, Eli F. and Ohlin, Bertil (1991): “Heckscher-Ohlin trade theory,” MIT Press, Cambridge.

Helliwell, John F. (1996): “Do National Boders Matter for Quebec’s Trade?,” The Canadian Journal of Economics, vol. 29(3), pp. 507-522, August.

Helliwell, John F. (1998): “How Much Do National Borders Matter?,” Brookings Institution, Washington DC.

Heplman, Elhanan (1987): “Imperfect Competition and International Trade: Evidence from Fourteen Industrial Countries,” Journal of the Japanese and International Economies, vol. 1(1), pp. 62-81, March.

Helpman, Elhanan and Krugman, Paul (1985): “Market Structure and Foreign Trade,” MIT Press, Cambridge.

Helpman, Elhanan, Melitz, Marc and Rubinstein, Yona (2008): “Estimating Trade Flows: Trading Partners and Trading Volumes,” The Quaterly Journal of Economics, vol. 123(2), pp. 441-487, May.

Hofstede, Geert (2001): “Culture’s Consequences: Comparing Values, Behaviors, Institutions and Organizations across Nations,” Sage, Thousand Oaks.

Karemera, David, Oguledo, Victor I., Davis, Bobby (2000): “A Gravity Model Analysis of International Migration to North America,” Applied Economics, vol. 32(13), pp. 1745-1755.

Klenow, Peter J. and Rodriquez-Clare (1997): “Quantifying Variety Gains from Trade Liberalization,” Working Paper, Graduate School of Business, University of Chicago, Chicago.

Kogut, Bruce and Singh, Harbir (1988): “The Effect Of National Culture On The Choice Of The Entry Mode,” Journal of International Business Studies, vol. 19(3), pp. 411-432, Fall.

Krugman, Paul (1979): “Increasing Returns, Monopolistic Competition, and International Trade,” Journal of International Economics, vol. 9(4), pp. 469-479, November.

Krugman, Paul (1980): “Scale Economies, Product Differentiation, and the Pattern of Trade,” The American Economic Review, vol. 70(5), pp. 950-959, December.

Krugman, Paul R. and Obstfeld, Maurice (2006): “International Economics – Theory and Policy,” Pearson Addison Wesley, Boston.

Lasserre, Philippe (2007): “Global Strategic Management,” Palgrave Macmillan, Basingstoke/New York.

Leamer, Edward E. (1974): “The Commodity Composition of International Trade in Manufactures: An Empirical Analysis,” Oxford Economic Papers, New Series, vol. 26(39), pp. 350-374, November.

Lindner, Staffan B. (1961): “An essay on trade and transformation,” John Wiley and Sons, New York.

McCallum, John (1995): “National Borders Matter: Canada-U.S. regional Trade Patterns,” The American Economic Review, vol. 85(3), pp. 615-623, June.

McFadden, Daniel (1963): “Constant Elasticity of Substitution Production Functions,” The Review of Economic Studies, vol. 30(2), pp. 73-83, June.

Melkumian, Arsen V. (2009): “A Gravity Model of Legal Migration Into the United States,” The Journal of Economic Development & Business Policy, vol. 2(1), Special section p. 1-18.

Mittermaier, Ferdinand (2008): “Nice to know about CES functions (preliminary),” Department of Economics, University of Munich retrieved 18 July 2012 from http://www.ecpol.vwl.uni-muenchen.de/downloads/advanced/ss08/uebung/att_ces_functions.pdf.

Newton, Isaac (1687): “Philosophiæ naturalis principia mathematica,” retrieved 16 May 2012 from University of Cambridge - Cambridge Digital Library: http://cudl.lib.cam.ac.uk/view/PR-ADV-B-00039-00001/9.

Novy, Dennis (2010): “International trade and Monopolisitc Competition without CES: Estimating Translog Gravity,” Warwick Economic Research Papers No. 929, University of Warwick, March.

Obstfeld, Maurice and Rogoff, Kenneth (2001): “The Six Major Puzzles in International Macroeconomics: Is There a Common Cause?,” NBER Macroeconomics Annual 2000, vol. 15, pp. 229-412, January.

Pepall, Lynne, Richards, Dan and Norman, George (2008): “Industrial Organization: Contemporary Theory and Empirical Application”, Blackwell, Malden.

Portes, Richard and Rey, Helene (2005): “The determinants of cross-border equity flows,” Journal of International Economics, Elsevier, vol. 65(2), pp. 269-296, March.

Portes, Richard, Rey, Helena and Oh, Yonghyup (2001): “Information and capital flows: The determinants of transactions in financial assets,” European Economic Review, vol. 45(4-6), pp. 783-796, May.

Rose, Andrew K. and van Wincoop, Eric (2001): “National Money as a Barrier to International Trade: the Real Case for Currency Union,” The American Economic Review, American Economic Association, vol. 91(2), pp. 386-390, May.

Rose, Andrew K. (2004): “Do We Really Know That the WTO Increases Trade?,” American Economic Review, American Economic Association, vol. 94(1), pp. 98-114, March.

Samuelson, Paul A. (1952): “The Transfer Problem and Transport Costs: The Terms of Trade When Impediments are Absent,” The Economic Journal, vol. 62(246), pp. 278-304, June.

Santos Silva, Joao and Tenreyro, Silvana (2006): “The Log of Gravity,” The Review of Economics and Statistics, MIT Press, vol. 88(4), pp. 641-658, November.

Santos Silva, Joao and Tenreyro, Silvana (2008): “Trading Partners and Trading Volumes: Implementing the Helpman-Melitz-Robinstein Model Empirically,” University of Essex – Discussion Paper No. 662, Oktober.

The International Bank for Reconstruction and Development/ The World Bank (2009): “World Development Report 2009: Reshaping Economic Geography,” The World Bank, Washington DC.

Tinbergen, Jan (1962): “Shaping the World Economy: Suggestions for an International Economic Policy,” The Twentieth Century Fund, New York.

Varian, Hal R. (2004): “Grunzüge der Makroökonomik,” Oldenbourg, München.

Wei, Shang-Jin (1996): “Intra-national versus international trade: how stubborn are nations in global integration?,” NBER Working Papers No. 5531, Cambridge.

WTO (2012): “Members and Observers,” retrieved 01 July 2012 from http://www.wto.org/english/thewto_e/whatis_e/tif_e/org6_e.htm.