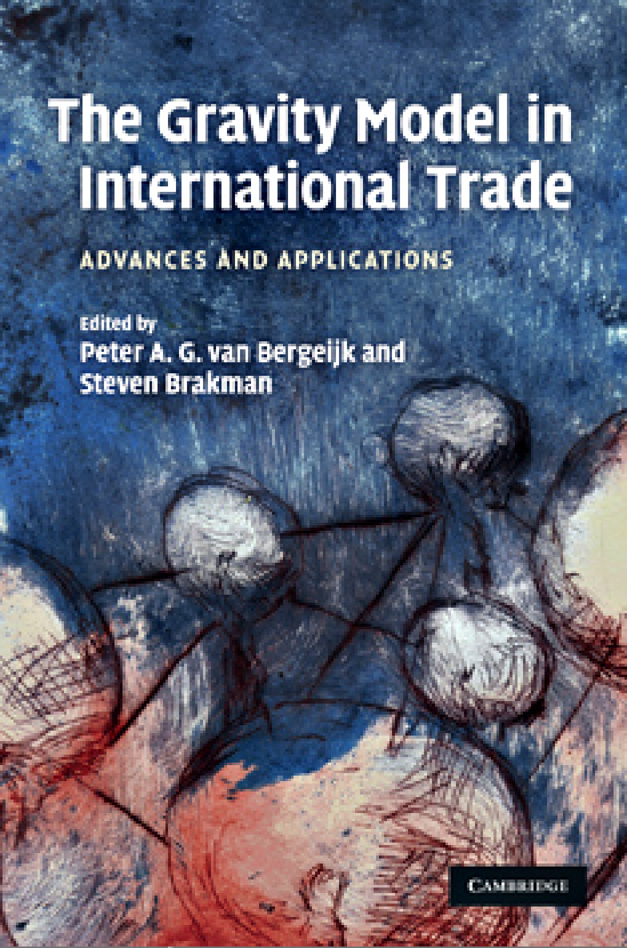
The Gravity Model in International Trade

ADVANCES AND APPLICATIONS

Edited by
Peter A. G. van Bergeijk and
Steven Brakman



CAMBRIDGE

CAMBRIDGE www.cambridge.org/9780521196154

This page intentionally left blank

The Gravity Model in International Trade

How do borders affect trade? Are cultural and institutional differences important for trade? Is environmental policy relevant to trade? How does one's income or wage relate to the fact that trade partners are nearby or far away?

These are just some of the important questions that can be answered using the gravity model of international trade. This model predicts and explains bilateral trade flows in terms of the economic size and distance between trading partners (e.g. states, regions, countries, trading blocs). In recent years, there has been a surge of interest in this model and it is now one of the most widely applied tools in applied international economics. This book traces the history of the gravity model and takes stock of recent methodological and theoretical advances, including new approximations for multilateral trade resistance, insightful analyses of the measurement of economic distance, and analyses of foreign direct investment.

PETER A. G. VAN BERGEIJK is Professor of International Economics and Macroeconomics at the Institute of Social Studies, the Hague. He is also Deputy Director of CERES, the coordinating body for development economics research in the Netherlands. His most recent book is *Economic Diplomacy and the Geography of International Trade* (2009).

STEVEN BRAKMAN is Professor of International Economics at the University of Groningen, the Netherlands. His most recent book is the *The New Introduction to Geographical Economics* (Cambridge University Press, 2009), coauthored with Harry Garretsen and Charles van Marrewijk.

The Gravity Model in International Trade

Advances and Applications

Edited by

Peter A. G. van Bergeijk

Institute of Social Studies/Erasmus University Rotterdam and CERES, Utrecht

and

Steven Brakman

University of Groningen

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore,
São Paulo, Delhi, Dubai, Tokyo

Cambridge University Press
The Edinburgh Building, Cambridge CB2 8RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521196154

First published 2010

Printed in the United Kingdom at the University Press, Cambridge

A catalogue record for this publication is available from the British Library

ISBN 978-0-521-19615-4 Hardback

Cover design based on Peeter Burgeik, “Gravitas,” etching 2009.

Contents

Figures

Tables

Contributors

EMIEL R. AFMAN, Ministry of Finance, the Netherlands

JAMES E. ANDERSON, Boston College and NBER, USA

SCOTT L. BAIER, Clemson University and Federal Reserve Bank of Atlanta, USA

PETER A. G. VAN BERGEIJK, International Institute of Social Studies of Erasmus University, the Hague and CERES, Utrecht, the Netherlands

JEFFREY H. BERGSTRAND, University of Notre Dame, USA

JACOB A. BIKKER, De Nederlandsche Bank and Utrecht School of Economics, the Netherlands

E. MAARTEN BOSKER, University of Groningen, the Netherlands

HERVÉ BOULHOL, OECD, Paris, France

STEVEN BRAKMAN, University of Groningen, the Netherlands

SJEF EDERVEEN, Ministry of Economic Affairs, the Netherlands

PETER EGGER, University of Munich and Ifo Institute, Germany

GUS GARITA, Erasmus University Rotterdam, the Netherlands

HARRY J. GARRETSEN, University of Groningen, the Netherlands

HENRI L. F. DE GROOT, Department of Spatial Economics, VU University Amsterdam, Tinbergen Institute, and CPB Netherlands Bureau of Economic Policy Analysis, the Hague, the Netherlands

KEITH HEAD, University of British Columbia, Canada

GERT-JAN M. LINDERS, Department of Spatial Economics, VU University Amsterdam, the Netherlands

CHARLES VAN MARREWIJK, University of Utrecht, the Netherlands

MATHILDE MAUREL, Centre d'Economie de la Sorbonne, CNRS, University of Paris 1, France

THIERRY MAYER, CERAS, CEPII, CEPR, University de Paris 1, France

L. JAN MÖHLMANN, Department of Spatial Economics, VU University Amsterdam, and CPB Netherlands Bureau of Economic Policy Analysis, the Netherlands

ANDREW K. ROSE, University of California, USA

ALAIN DE SERRES, OECD, Paris, France

MARK M. SPIEGEL, Federal Reserve Bank of San Francisco, USA

1 Introduction

The comeback of the gravity model

P. A. G. van Bergeijk and S. Brakman

1.1 Introduction

The gravity model describes one of the most stable relationships in economics: interaction between large economic clusters is stronger than between smaller ones, and nearby clusters attract each other more than far-off ones. This formulation of the model admittedly is vague. What is meant by large economic clusters, or “far-off”? In fact, this ambiguity reflects the success of the gravity model in economics. Although the model is probably best known in the context of international trade and capital flows between countries, it has also been successfully applied to describe how consumers flow between different shopping malls, patients between hospitals and much more. Also “distance” is a very broad concept. It might reflect actual distances in miles, as an approximation of transportation costs, but over the years more subtle elements of distance-related factors have been considered. Economic factors such as tariffs and non-tariff barriers have been included in applications of the gravity model, but also “non-economic” factors have been included, such as cultural differences, differences in religion, language (dis)similarities, the presence or absence of former colonial ties, institutional differences, differences in technological development, and so on.

The list of applications is long and, most remarkably, empirical tests show that this simple idea is very successful from an empirical point of view and is able to show that many economic phenomena between different locations can empirically be described by a gravity equation.

The papers in this volume were presented in October 2007 at a conference at the Faculty of Economics, University of Groningen, the Netherlands and coorganized with the Ministry of Economic Affairs, Directorate-General for International Economic Relations (DG-BEB). We would like to thank Harry Garretsen, conference participants, and the referees – selected by Cambridge University Press – for their comments and suggestions on earlier versions of the papers and this chapter in particular. We would also like to thank Chris Harrison and his team for editorial support and the Ministry of Economic Affairs, Directorate-General for International Economic Relations (DG-BEB) and the Faculty of Economics of the University of Groningen for financial support for the conference.

Remarkably and in contrast to popular belief, recent estimates of the gravity equation show that distance-related variables have become *more* instead of *less* important. A recent special issue of the *Cambridge Journal of Regions, Economy and Society* (2008), for example, was titled "The World is not Flat," in order to indicate that distance is still one of the most salient characteristics to describe economic interaction in the world economy.[1]

Table 1.1 summarizes (most of) the contributions in this book. The table shows the versatility of the model and provides a good impression of the many questions that can be studied with a gravity model. Why do borders matter (and by how much)? Are cultural and institutional differences important for trade? What is the contribution of an ambassador to bilateral trade? Is environmental policy relevant for trade? How does one's income or wage relate to the fact that trade partners are nearby or far away? These are just a small sample of all the questions that can be answered by the gravity equation. The collection of papers in this book also provides a rich sample of different empirical approaches with regard to the gravity model. Indeed, many chapters contain comparisons and tests of the different modeling strategies including – extended versions – of Tinbergen's straightforward OLS, the method proposed by Anderson and van Wincoop, and more practical variants thereof. One finding stands out – as Table 1.1 testifies: although the magnitude of the distance effect on trade may differ, its sign and significance endure the hardship of scientific scrutiny.

Since the gravity equation was introduced by Tinbergen in 1962 in his *Shaping the World Economy* (where it actually was developed in one of the appendices), the gravity model has always been around in policy circles, because of its robustness and as a versatile tool to analyze all kinds of (trade) policy issues. The academic popularity of the gravity analysis, however, waned in the 1970s and 1980s, especially since the gravity equation could be derived from almost any international trade model, thus offering no scope to test between theories. The lack of a convincing and unambiguous micro-economic foundation gave the gravity model a somewhat ambivalent reputation: perhaps useful as an empirical tool, but unsatisfactory from a theoretical point of view. In the last twenty years the model has again become fashionable due to seminal contributions by Anderson and Bergstrand. The chapters in this book vividly reflect this renewed interest.

[1] In Chapter 2 Bergstrand and Egger attack another myth related to the assumed flatness of the world, namely the idea that increased exports and direct investments from developing countries dominate international trade and investment flows.

Table 1.1. *Overview of gravity analyses in this book*

Chapter	Subject	Period	Sample	Range of the distance parameters		Significant positive regional trade agreements
2 Bergstrand, Egger	GE theory bilateral FDI, final goods and intermediate goods	1990–2000	160 countries	−0.58[a] −0.83[b]	−1.08[a] 0.31[b]	n.a.
4 Baier, Bergstrand	Replication Anderson and van Wincoop (2003) and application of GE approximation	1993	10 Canadian and 20 US states	−0.79	−1.26	n.a.
5 Bikker	Trade creation *versus* trade diversion	2005	178 countries	−1.30	−1.70	Yes, not for EU
6 Head, Mayer	Interpretation of border effects in relation to measurement of distance	1997 1993–95	US states European countries	−0.47 −0.74	−0.73 −1.20	n.a.
7 Bosker, Garretsen	Market access	1996	97 countries	−0.72	−0.0	n.a.
8 Möhlmann, Ederveen, de Groot, Linders	Cultural and institutional distance by product group	2000	55 countries (27 OECD)	−0.53	−1.42	Yes
9 Rose, Spiegel	Trade impact of international environmental arrangements	2001–03	68 source and 221 host countries	−0.49	−1.07	Yes
10 Afman, Maurel	Diplomacy and trade in the context of East–West Trade	1995, 2000, 2005	56 countries (26 OECD)	−1.13	−1.44	Yes
11 Brakman, Garita, Garretsen, van Marrewijk	Cross-border mergers and acquisitions	1986–2005	211 countries	−0.26[b]	−0.36[b]	n.a.

[a] Final goods and intermediate goods.
[b] FDI.

This introductory chapter is organized as follows. First we give a brief historical account of developments surrounding the gravity model. Next we turn to the empirical applications and highlight some of the problems that arise when estimating the model. The main aim of this introduction is to place the contributions of this volume into perspective, rather than providing a full-fledged survey.[2] The final section formulates some of the challenges that lie ahead.

1.2 The gravity model

1.2.1 The history of gravity

The gravity model has a long history as many authors have noted the relationship between, on the one hand, flows between different locations and on the other hand, the "weight" of these locations and the inverse of distance. An early cogent formulation of the gravity narrative is Ravenstein (1885, pp. 198–99) who explains how "currents" of migration are driven by the "absorption of centers of commerce and industry" but "grow less with the distance proportionately." Noteworthy is also the discontent with the classic trade model's neglect that was voiced in the first half of the twentieth century by Ohlin and in the German trade and location school that comprises Weber, Furlan, Engländer, Predöhl, and Lösch. Building on these theories and motivated by a desire to bring multilateral trade and distance into the trade economist's common toolkit, Isard and Peck (1954) empirically demonstrate the negative impact of distance for different modes of both domestic and international transport.[3] Isard (1954), in fact, comes close to formulating a gravity equation, but he uses a somewhat different metaphor from physics than Newton's gravity (electric potential rather than gravity). Still, Isard had already envisioned many of the issues that engage researchers of the gravity model today as he stresses the importance of measurement issues, the composition of trade, cultural factors, and politics for empirical research into the determinants of bilateral trade flows that takes distance seriously. Although the gravity narrative thus has many fathers, the first mathematical formulation and empirical application of the gravity model is due to a group of Dutch economists headed by Tinbergen who were the first to actually publish a gravity model and an empirical application (Tinbergen 1962,

[2] The reader may wish to consult Anderson and van Wincoop (2004), or Combes *et al.* (2008).

[3] Interestingly the same desire to integrate gravity into the much wider class of general equilibrium trade models is still a strong motivation. See, for example, Chapter 3 by Anderson.

Appendix VI).[4] Tinbergen supervised the Ph.D. thesis of Linnemann (1966) that has become the standard reference to the early version of the gravity equation.

These early contributions started the first wave of applications in the early 1960s (see for critical discussions of this early literature, Taplin 1967, and Leamer and Stern 1970, chapter 6). Although the model itself can be applied to many phenomena, most applications involved bilateral trade flows and in our discussion we will thus concentrate on trade. The basic form of the gravity equation is as follows:

$$T_{ij} = \frac{GDP_i^{\alpha} GDP_j^{\beta}}{D_{ij}^{\theta}}, \tag{1.1}$$

where: T_{ij} indicates bilateral trade between country i, and j; GDP_i indicates the economic size of i, measured by GDP; and D_{ij} indicates the bilateral distance between the two countries. The parameters α, β, and θ are often estimated in a log-linear reformulation of the model. This equation explains bilateral trade using economic size and distance: the larger the two trading partners, the larger the trade flows; the larger the distance between the two countries, the smaller bilateral trade. Usually the model explains 70–80 per cent of the variance in bilateral trade flows.

What was missing during the early days of the gravity equation, despite its popularity, was a convincing micro-economic foundation. Tinbergen (1962 p. 263) actually introduced his trade flow equation in a straightforward manner as "a turnover relationship in which prices are not specified," providing only a common sense rationale for this quasi-postulated reduced form equation: trade is determined by supply potential (exporter GDP), market demand potential (importer GDP), and transportation costs (distance). Linnemann (1966) provided some theoretical arguments to justify the formulation of the trade flow equation, deriving it in the context of a quasi-Walrasian model.[5] Neither his attempt nor those of Pöyhönen (1963) and Pullianen (1963) provided a solid

[4] Other members of the group at the Nederlands Economisch Instituut that developed the model were R. Rijken van Olst, F. Hartog, H. Linnemann, and A. van Oven. It is noteworthy that a group of Finnish economists at the same time worked along the same lines, but their scientific publications appeared a year later (Pöyhönen 1963 and Pullianen 1963).

[5] Leamer and Stern (1970, p. 158) who were probably the first to explicitly refer to these models in a standard textbook as "gravity models" mention two other non-economic rationales for the model, simply referring to Newton and pointing out that bilateral trade is the expected value of trade between two partners based on the probability that they meet on the world market.

micro-foundation. Leamer and Stern (1970, p. 169) for example conclude: "The significance of such research must be found in the context of seeking a broader understanding of the empirical base of the pure theory of international trade. This is something a number of studies cited have failed to make clear."

The lack of a sound theoretical foundation gave the gravity model a somewhat dubious reputation among academics although its empirical strength was recognized as it was able to "identify extreme cases of artificial barriers to trade, the role of distance and the effects of membership in various customs union and trade preference groups" (Taplin 1967, p. 442). According to the survey of Leamer and Levinsohn (1995, p. 1387) this state of affairs was still present in the early 1990s: "The gravity models are strictly descriptive. They lack a theoretical underpinning so that once the facts are out, it is not clear what to make of them." They further note that the model is not linked with any issue or raises only further problems (p. 1397): "... the basic proposition that free trade is beneficial doesn't seem obviously at risk if distance is added to the model, so why bother? ... once we admit distance between countries into our theories, it begs the question: what about the role of distance within them?" Leamer and Levinsohn (1995) were clearly not impressed by the gravity model, but also noted that the fact that distance is found to be important creates a tension "that needs to be remedied by a much closer association of the descriptive gravity models with the theory ... of international economics."[6] Deardorff (1998), taking a serious look at the micro-foundations of the gravity model at the same time as Leamer and Levinsohn (1995), is also critical with respect to the theories behind it, but ironically not because of a lack of micro-economic foundations, but of too many. He shows that the model is consistent with a large class of trade models. This implies that the use of the gravity model as a tool to distinguish between theories becomes problematic.[7]

Paradoxically the fact that the gravity model could be derived from Heckscher-Ohlin, as well as increasing returns to scale, Ricardian models,

[6] Somewhat ironically they note that (p. 1387) "... a theoretical foundation by Anderson (1979) is formally fruitful but seems too complex to be part of the everyday toolkit." At the time of the publication of the Handbook of International Economics (Grossman and Rogoff 1995), the gravity model was again gaining ground, and the work of Anderson was pivotal for the revitalization of the gravity model.

[7] See also Evenett and Keller (1998). Feenstra *et al.* (2001, p. 446) agree with Deardorff (1998), but take the discussion a step further: "... the theoretical foundations for the gravity equation are actually quite general, but the empirical performance is quite specific ... different theories lead to measurable different home market effects..." These differences in parameter values can be used as guides in empirical tests.

and so on, was offering great confidence to policy makers. Advice based on the gravity analysis did not seem to depend on a particular vision of the economic process, that is: policy makers argued that the gravity equation was robust with respect to economic theories. Actually, in the early and mid 1990s the model was very fashionable in policy institutions especially for analyzing the big changes in the world trade system when the Iron Curtain fell and a substantial empirical literature developed on this policy-relevant question.[8] These studies predicted substantial improvements of the world trade potential and important shifts in global trade patterns. Havrylyshyn and Pritchett (1991), for example, predicted that the geographic pattern of trade by the mid-European countries would change dramatically. In the 1950s to 1980s 60–80 per cent of their activities focused on the other communist countries and only about 20–30 per cent on Northern and Western Europe, but their calculations showed that the natural "undistorted" trading pattern would be exactly the reverse. The gravity predictions regarding this major systemic shock were spot on, as is illustrated by the discussion of the trade transition in European East–West trade in Chapter 10 by Afman and Maurel.[9] The case of East–West trade is an example that fits in a number of more general-oriented evaluations of the predictive power of the gravity model, such as the analysis of predictive out of sample performance by Ward and Hoff (2005).[10]

All in all, the general consensus in the literature was that the ambiguous or "lack of theoretical underpinnings significantly weakens the credibility of a model, as it introduces a certain degree of subjectivity in the interpretation of the estimated coefficients" (Piermartini and Teh 2005, p. 37). The empirical performance of the model in itself could not convince theorists and also on this account the search for a sound micro-foundation for the model continued.

1.2.2 The micro-foundation of gravity

The discussion above illustrates that starting from the earliest applications of the gravity equation, and stimulated by the empirical success of the model, the search for its micro-foundations were always high on the research agenda.

[8] Examples are: van Bergeijk and Oldersma (1990), Havrylyshyn and Pritchett (1991), Wang and Winters (1991), Döhrn and Milton (1992), Ezran *et al.* (1992), and Hamilton and Winters (1992).

[9] The development of post-1989 trade between East and West continues to inspire evaluations with the gravity model, see for example Bussière *et al.* (2005).

[10] Ward and Hoff (2005) also find that a substantial and persistent amount of trade variation is attributable to country-specific importer and exporter effects.

Anderson (1979) provided a sound micro-economic foundation, but his achievement initially did not get the attention that it deserved (see note 7). Using standard economic techniques, he showed how the gravity model fits into an optimizing framework and inter alia shows some of its limitations. He assumes a (weakly) separable social utility function with respect to traded and non-traded goods. Each region produces both types. In the first round of utility maximization, the share of j's income which is spent on traded goods, a_j, can vary among different regions, and depends on income, and population size in j. In the second round region j maximizes a homothetic (Cobb–Douglas) utility function which is identical across all regions. This implies that, ignoring price discrimination, the share that country j spends on the exports of tradeable goods from i, s_i, is equal for all j (i.e. s_i varies only with i). Country j's imports from country i can be expressed as:

$$T_{ij} = s_i a_j GDP_j \tag{1.2}$$

Equilibrium on the traded goods market implies: $a_i GDP_i = s_i \sum_j a_j \times GDP_j$. Solving for r_i and substituting in the bilateral trade equation gives:

$$T_{ij} = a_i GDP_i a_j GDP_j \Big/ \sum_i \sum_j T_{ij}$$

$$\text{note that } \sum_i \sum_j T_{ij} = \sum_j a_j GDP_j \tag{1.3}$$

This is already a simple variant of the gravity model in which economic masses determine the trade flows. The applicability to real world problems, however, is limited to years with balanced trade. Furthermore, countries must have similar demand structures. It is relatively easy to extend the model with population variables and, most importantly, trade barriers. Bergstrand (1985, 1989) is the second author to provide a theoretical foundation for the model, in which he highlights (complicated) price terms, which are absent in the derivation above.[11] Bergstrand (1985, 1989, 1990) develops a relationship between trade theory and bilateral trade, and includes the supply side of the economy explicitly. The income of destination countries enters the equation because of demand;

[11] The appendix in Anderson (1979) is interesting, as all modern variants of the gravity equation are based on the derivation in this appendix – it anticipates the famous multilateral resistance terms introduced explicitly in Anderson and van Wincoop (2003). These are also acknowledged in Bergstrand (1989, 1990).

income of the exporting countries because it reflects the supply capacity of the exporting country; and distance because it reflects transportation costs which are passed on to the consumers in the destination countries (cif/fob difference between prices in countries of destination and countries of origin). With the benefit of hindsight the contributions of Anderson and Bergstrand initiated a renewed interest in the gravity model, culminating in Anderson and van Wincoop (2003) that extends Anderson (1979), but also introduces a method to deal with the complicated price (index) terms (which are already present in the appendix of Anderson 1979 and in Bergstrand 1985, 1989, 1990).[12]

Anderson and van Wincoop (2003) has become the main reference for subsequent work on the gravity equation. This is why we give a simplified derivation of the model, in six steps, that is based on Baldwin and Taglioni (2006):

Step 1: The first step is a supply equals demand equation. The expenditure share identity, that includes the relevant prices, says that the *value* of trade flow from country i to j, $p_{ij}x_{ij}$, should equal the share country i has in expenditure of j.

$$\begin{aligned} p_{ij}x_{ij} &= s_{ij}E_j, \text{where } p_{ij} = \text{import price, from } i \text{ to } j, \\ s_{ij} &= \text{share of } i \text{ in } j\text{'s expenditure}, E_j. \end{aligned} \tag{1.4}$$

Step 2: Next s_{ij} follows from a familiar CES demand structure and it is easy to derive an explicit expression for the imported goods share in E_j. Assuming that all goods are traded this share depends on the bilateral prices relative to a price index.

$$s_{ij} = \left(\frac{p_{ij}}{P_j}\right)^{1-\sigma} \qquad \text{where } P_j = \left(\sum_{i=1\ldots N} n_i \left(p_{ij}\right)^{1-\sigma}\right)^{1/(1-\sigma)} \tag{1.5}$$

is the exact price index associated with the CES demand structure; $\sigma > 1$ is the elasticity of substitution between "varieties"; N is the number of nations; n_i is the number of varieties supplied by nation i. Note that varieties are defined symmetrically, which allows us to ignore a variety index.

Step 3: Adding trade costs. A crucial element in all gravity equations is the presence of trade costs. These are easily introduced. If t_{ij} indicates

[12] Bergstrand (1989, 1990) is the first to give explicit attention to the price index terms, but in the empirical applications he uses existing price indices. This is also the approach taken in Baier and Bergstrand (2001).

bilateral trade costs, the price in market j equals:

$$p_{ij} = p_i t_{ij}, \tag{1.6}$$

where p_i is the mill price of a variety in country i (note again the absence of an index for varieties; varieties are defined symmetrically). After transportation the price in market j becomes p_{ij}.

Step 4: The gravity equation describes total trade between two countries; this implies that we have to aggregate across varieties

$$T_{ij} = n_i s_{ij} E_j = n_i \left(p_i t_{ij}\right)^{1-\sigma} \frac{E_j}{P_j^{1-\sigma}}, \tag{1.7}$$

where the second equality follows from combining equations (1.5) and (1.6) in the bilateral trade equation (1.7).

Step 5: All goods are traded, so the budget constraint says that total output of country i, Y_i, equals total sales to all destination countries j (including country i itself).

$$Y_i = \sum_j T_{ij} = n_i p_i^{1-\sigma} \sum_j \left(t_{ij}^{1-\sigma} \frac{E_j}{P_j^{1-\sigma}} \right), \tag{1.8}$$

where the second equality follows from combining equation (1.7) with (1.8). We can rewrite equation (1.8) as:

$$n_i p_i^{1-\sigma} = \frac{Y_i}{\Pi_i^{(1-\sigma)}}, \quad \text{where } \Pi_i = \left(\sum_j \left(t_{ij}^{1-\sigma} \frac{E_j}{P_j^{1-\sigma}} \right) \right)^{1/(1-\sigma)} \tag{1.9}$$

Step 6: A gravity equation can now be derived from inserting equation (1.9) into equation (1.7) which gives:

$$T_{ij} = Y_i E_j \left(\frac{t_{ij}}{\Pi_i P_j} \right)^{1-\sigma} \tag{1.10}$$

Equation (1.10) is identical to equation (9) in Anderson and van Wincoop (2003).[13] Note that the main difference between equation (1.1)

[13] Note that Anderson and van Wincoop (2003) use income shares. To see that the two expressions are identical, multiply and divide equation (1.10) and Π_i by world income, y^W.

and equation (1.10) consists of the price indices P and Π, which are the so-called multilateral resistance terms.[14] These are intuitively easy to understand. Bilateral trade between two countries does not only depend on bilateral variables related to these two countries alone, but also on their position relative to the world economy. Consider for example the ranking of OECD countries by Boulhol and de Serres in Chapter 12 by different measures of centrality. They empirically distinguish remote countries, such as Australia and New Zealand, peripheral countries, such as Korea and North America, and centrally located and dense economies such as Belgium and the Netherlands. In the context of multilateral trade resistance a pair of countries in the EU, for example, faces tougher competition from nearby trading partners than a similar pair of countries that are more isolated from the world economy. So trade between two countries in the EU is, ceteris paribus, smaller than trade between the two similar countries that are more isolated (and for which P and Π are higher).[15] Equation (1.10), and variants thereof, has become the standard formulation in recent applications of the gravity equation (see also Anderson and van Wincoop 2004).

1.2.3 Empirical applications of the gravity equation: empirical strategies

Recent empirical research discusses four main topics that have to be dealt with in applications of the gravity equations: how to deal with the multilateral resistance terms, how to deal with the large number of zeroes in trade statistics, how to measure distance and what is the appropriate level of (dis)aggregation for the gravity analysis.

1.2.3.1 Multilateral resistance Although equation (1.10) looks simple it is also difficult to estimate. In essence this is caused by the fact that the multilateral resistance terms depend on trade costs and the multilateral resistance terms themselves, which are part of the estimation. This causes a circular dependence in the estimation of equation (1.10). This complication is a drawback of empirical applications of equation (1.10). Anderson and van Wincoop solve this by making the additional assumption of symmetrical trade costs, $t_{ij} = t_{ji}$, and a custom programmed system of (non-linear) equations.

14 Anderson and van Wincoop (2003) introduce a further simplification: symmetry of trade costs, $t_{ij} = t_{ji}$ that allows a further simplification.

15 See for an extensive discussion of the multilateral resistance terms, Anderson and Neary (2005, Chapter 10) and Chapter 3 by Anderson in this volume.

Three alternatives have been proposed in the literature to deal with this problem: including fixed effects as an approximation of the multilateral resistance terms, a linearization, and an analytical solution.

First, the multilateral resistance terms are unobserved, but can be estimated using fixed effects. This is a method used by, for example, Rose and van Wincoop (2001) or Redding and Venables (2004). For this purpose equation (1.10) can be rewritten as (where the income terms are used as correction factors for the bilateral trade flows):

$$\ln(T_{ij}/Y_iE_j) = (1-\sigma)t_{ij} + \alpha_1^i D^i + \alpha_2^j D^j + \alpha_3 \varepsilon_{ij}, \tag{1.11}$$

where D^i is a dummy that is unity when i is the exporter and zero otherwise, and D^j a dummy that is unity when j is the importer and zero otherwise, and ε_{ij} is the error term. The coefficients measure the multilateral resistance terms: $\alpha_1^i = \ln(\Pi_i)^{\sigma-1}, \alpha_2^j = \ln(P_j)^{\sigma-1}$. Feenstra (2004) expresses a preference for this method as it is easy to implement and gives consistent estimates of the average effects. A disadvantage of the fixed effects method, however, is that the estimates cannot be used in order to calculate comparative-static effects involving changes in trade costs, for which the gravity equation is often used (the multilateral resistance terms, in principle, have to be recalculated for each exercise as transport costs are an element of the multilateral resistance terms).

Second, the solution proposed by Straathof (2008) shows that, using the same assumptions as Anderson and van Wincoop (2003) the nonlinear problem that results from them can be linearized and solved analytically. The advantage of this method is that no approximations are necessary, and that standard econometric estimates can be used in order to derive identical results as in Anderson and van Wincoop (2003). However, as Straathof observes, his method results in an endogeneity problem.

Third, Baier and Bergstrand (2009) recently proposed an elegant method that addresses both the issue of the estimation of the multilateral resistance terms as well as the applicability in comparative-static exercises and does not suffer from Straathof's endogeneity problem. Using the same assumptions as Anderson and van Wincoop (2003) they first apply first-order Taylor-series expansion to the multilateral resistance terms (that depend on – weighted – trade costs), and substitute these in (a variant of) equation (1.10). Baier and Bergstrand further develop this theory for the case of symmetric trading costs in Chapter 4 in this book.

All these methods illustrate the relevance of the contribution of Anderson and van Wincoop (2003), and since this seminal publication, applications of the gravity model have to deal with the multilateral resistance terms, one way or another, in order to be taken seriously.[16]

1.2.3.2 Zero trade flows Another important problem with the analysis of trade flows is the occurrence of zero bilateral trade flows. Helpman *et al.* (2008, p. 443) indicate that 50 per cent of the 158 countries in their sample do not trade with one another. If the gravity equation is applied to FDI flows this number increases to more than 80 per cent. This can be caused by rounding errors, missing observations or truly zero trade flows. The standard procedure (following Linnemann 1966) in empirical studies is simply to drop the zero flows from the sample (Chapter 10 by Afman and Maurel is an example), or add a small constant to all trade flows in order to be able to estimate a log-linear equation (Chapter 11 by Rose and Spiegel is an example); both these procedures – among additional alternatives – have been used in, for example, Santos Silva and Tenreyro (2006).

These approaches are correct as long as the zero values are randomly distributed; however, if they are not random, as is most often the case, it introduces selection biases. Until recently, this problem has been ignored in gravity studies, but it can be handled by means of sample selection correction.[17] In this light, Helpman *et al.* (2008) propose a theoretical model rationalizing the zero trade flows and propose estimating the gravity equation with a correction for the probability of countries to trade. In order to estimate their model, they apply a two-step estimation technique (similar to sample selection models commonly used in labor economics). To implement the new estimator, one needs to find an appropriate exclusion restriction for identification of the second-stage equation, which can be quite difficult. Indeed, zero flows need to be taken more seriously. For one thing the problem occurs more often when a great many developing

16 Although the multilateral resistance terms received a lot of attention in the literature it is not the only empirical problem with the gravity equation. Baldwin and Taglioni (2006), for example, stress three problems. First, they show that the multilateral resistance terms change over time, which implies that the gravity equation developed by Anderson and van Wincoop (2003) can only be estimated in cross-section studies. Second, they point out that if the gravity model is applied to trade between country pairs one should take the average of the logs instead of the log of the average. In the case that trade between countries is balanced the error is small; however with unbalanced trade the error is large. Third, trade data are often deflated by a price index that is the same for all bilateral trade flows (often the US aggregate price index). This can create spurious correlation if there are global inflationary trends in the data that affect both the left-hand side of the equation as well as the right-hand side.

17 The fact that zero flows contained potentially useful and relevant information was also recognized by Bikker (1982, pp. 399–411) in an early application of the Tobit model to international trade.

countries are included from different continents (see Chapter 7 by Bosker and Garretsen), when lower levels of aggregation are used, for example in the analysis of trade at the level of individual product groups (see Chapter 8 by Möhlmann *et al.*) and in FDI (see Chapter 11 by Brakman *et al.*). These chapters provide examples of econometrical solutions to the zero-observation problems.

1.2.3.3 Measurement of distance Finally, without much exaggeration one can claim that the most important contribution of the gravity equation is that it points out the relevance of trade costs. The measurement of these costs is often crude as researchers rely – in addition to actual distance – on the use of dummies to indicate borders, language similarities, cultural differences, colonial ties, membership of Preferential Trading Areas, etc. Only a few studies use actual data on shipping costs, such as Limao and Venables (2001) or Combes and Lafourcade (2005). These studies show that distance might be an inadequate proxy for transport costs. Direct measurement is difficult and differs between commodities. Anderson and van Wincoop (2004) show that on average trade costs are about 170 per cent of the mill price of manufactured goods and consist of 55 per cent of internal costs, 21 per cent transportation costs, and 44 per cent related to border effects ($2.7 = 1.55 * 1.21 * 1.44$). Studies like these confirm that the world is still far from being "flat."

To be more specific it is instructive to look at Figure 1.1, which is based on data underlying the Disdier and Head (2008) study. They performed a so-called meta-analysis of gravity-equation estimates, based on 1,467 estimated distance effects (indicated by θ in their study) in 103 publications using the gravity equation. Their findings can be effectively summarized with the help of Figure 1.1.[18] The higher the estimated parameter θ, the stronger the negative effect of distance on the size of trade flows, and therefore the more important distance and location is for determining these trade flows. The mean effect of distance on trade for the period as a whole is around 0.9 (with 90 per cent of all estimates between 0.28 and 1.55).[19] This implies that a 10 per cent increase in distance leads to a 9 per cent reduction of international trade flows. The estimates in Figure 1.1 suggest that the distance effect became less important between 1890 and 1940. Most striking, however, is the *increased* (not decreased) estimated distance effect in the second half of the twentieth century (also

[18] We are grateful to Anne-Célia Disdier and Keith Head for providing us with the data for this figure. The dates indicate the mid points in a sample.

[19] This can be compared to the estimates in this book that are in the range of 0.47 to 1.70; see Table 1.1.

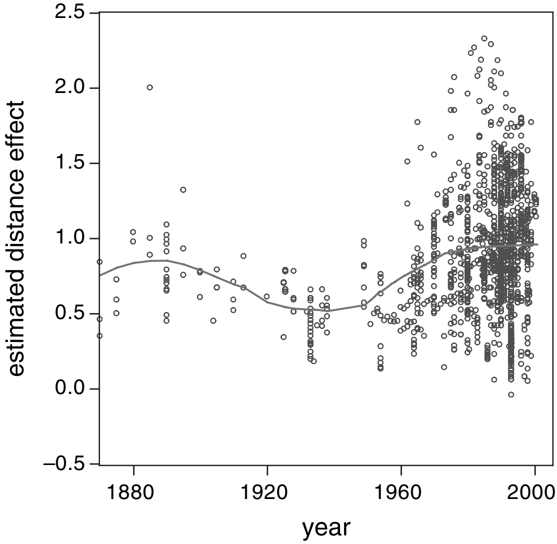


Figure 1.1. Distance effects in gravity-equation studies
Data source: Disdier and Head (2008). Our kernel fit is based on Epanechnikov (h = 19.575).

with some increased variance). In sharp contrast to the opinion of the death-of-distance group, distance is therefore becoming *more* (not less) important for determining international trade flows. See Berthelon and Freund (2008) and Disdier and Head (2008) for further discussion.

1.2.3.4 Micro-data and product groups For a long time, gravity has hardly been studied at disaggregated levels. This may have been due to its originally macro-economic formulation, the importance of country characteristics, limited availability of data or the still limited computational capacity during the early phases of gravity modeling. Whatever the reason may have been, at the turn of the century it was still "surprising how little work has been done on examining disaggregated gravity equations" (Harrigan 2002). Since then a lot of progress has been made mainly in the slipstream of theories about the international activities of essentially heterogeneous firms (Melitz 2003, see also Chaney 2008 for an overview). Krautheim (2007), for example, models a reduction of fixed export costs through information networks that enables firms to share and spread the available knowledge about foreign markets. Krautheim (2007) thus suggests a solution to the distance puzzle since in this setup the fixed costs of exporting are endogenously increasing in trade.

An important issue in the empirical debate is the channel through which gravity works. Does distance influence the extensive margin (the number of firms, the number of products, the number of international trade and/or investment relations) or the intensive margin (value per product, per firm, per bilateral trade and/or investment relationship)? Typically micro-data studies find a strong negative relationship between distance and the numbers of exporting firms, products, and bilateral trade and investment relations (see for example Buch *et al.* 2003 or Hillberry and Hummels 2008). Empirical application of theories that recognize the importance of firm heterogeneity in the context of the gravity model presently often still use macro-data (Bernard *et al.* 2007; Helpman *et al.* 2008), but we assume that with the greater availability of micro-datasets for more countries (see for example International Study Group on Exports and Productivity 2007) more broader empirical testing of theories of characteristics and behavior of heterogeneous firms will become the standard.

A more disaggregated analysis is not only warranted by a better understanding of the actual behavior of the firm; it is also driven by measurement issues related to distance. Border effects, for example, appear to be to a large extent driven by the extent of transportability of goods, as shown by Head and Mayer (Chapter 6). Möhlmann *et al.* (Chapter 8) report significant differences in the distance parameter for different product categories. Both chapters show that an analysis by product type generates both new insights and research questions. The need for further analysis is illustrated by the fact that the findings in these two chapters sometimes agree – for example, that the markets for refined petroleum products are local – but sometimes disagree. Head and Mayer classify tobacco and beverages as products which are difficult to transport whereas Möhlmann *et al.* characterize the markets of these goods as international.[20]

1.3 About this book

The renewed interest in and further development of the gravity equation contrasted with actual policy discussion that assumed whether the world has become "flat." Actually the period following the publication of Friedman's *The World is Flat* in 2005 was a somewhat strange time for the gravity equation, as distance (which is the key feature of gravity) was

20 One possible explanation is that Möhlmann *et al.* also deal with the trade between continents, while Head and Mayer provide an analysis for a national market (USA) c.q. the internal EU market.

increasingly being ignored in policy circles, while at the same time much progress was made on the theoretical and empirical side in academia.[21] This contrast and the wish to take stock of recent advancements provided the reasons to organize the conference "The Gravity Equation or: Why the World is not Flat" in Groningen in October 2007. We have organized the contributions in this book in three coherent parts: methodology, measurement of distance, and applications.

1.3.1 Methodology

In the first contribution Bergstrand and Egger (Chapter 2) develop a general equilibrium model that endogenizes the decision to import, export or to invest abroad. In this model goods can be produced in any country and can potentially be composed of intermediates from all other countries. Flows in goods are further distinguished from FDI. The general equilibrium model is calibrated using a newly created dataset that decomposes the world trade matrix into final and intermediate goods. This tool enables them to theoretically distinguish three gravity specifications for trade in intermediate and final goods and FDI, respectively. This method is a novel approach to arrive at testable unambiguous theoretical rationales regarding the gravity equation: all three types of bilateral flows should be positively related to both the size and similarity of GDP and all own-price effects should be negative, and cross-price effects indicate that barriers to FDI increase flows of final goods (and vice versa: barriers to trade in final goods increase FDI). These hypotheses are confirmed by empirical testing.

Anderson (Chaper 3) extends the applicability of the by now famous multilateral resistance terms that were introduced in Anderson and van Wincoop (2003). Anderson shows that by applying simplifying assumptions a more straightforward derivation is possible by assuming that shipments are to a basket of commodities and demand also comes from this basket; the incidence of trade costs is then divided between supply and demand. This extension of Anderson and van Wincoop (2003) has the potential to get a wider application of the gravity model, in particular in trade and investment theories that so far have not considered distance as an explanatory variable.

Baier and Bergstrand (Chapter 4) take up the issue of how comparative-static analyses that are key to policy makers can be reconciled with the estimation of the multilateral resistance terms. Building on Baier and

[21] See also the review of Friedman by Leamer (2007).

Bergstrand (2009), and extending the analysis for the case of symmetric trade costs, extensive Monte Carlo simulations illustrate that this new method gives almost identical estimates to the method proposed by Anderson and van Wincoop or the fixed effects method. The additional advantage of this method, however, is that the multilateral resistance terms can be retrieved and used in comparative static exercises, as their application to the NAFTA and the European Economic Area illustrate. This solution to the difficulties involved in estimating and simulating along the lines of Anderson and van Wincoop (2003) will be of interest to many applied trade policy analysts.

Bikker (Chapter 5) uses quite a different technology that, however, provides the same sort of result as Anderson and van Wincoop (2003). One of the limitations of the gravity model is that it focuses on bilateral trade, and in fact can only explain an increase or decrease in a bilateral trade flow. Still, the model is applied to issues of trade integration, where trade creations and diversions are key concepts. However, by assumption the model cannot explain substitutions between flows. Bikker extends the standard formulation of the gravity model in such a way that also substitutions between flows can be analyzed by constructing measurement of trade totals in the bilateral trade matrix. Substitution can thus be distinguished from creation.

1.3.2 Distance in the gravity model

The concept of bilateral distance is the main determining characteristic of the gravity model and thus measurement issues related to distance are key to the validity of any empirical application, but also to the interpretation of the results of the econometric findings. Economics is no physics. In the natural sciences distance is well defined and its measurement can be exact and unambiguous. Economic distance, however, is a multifaceted concept, and measurement and interpretation accordingly are subject to continuous debate. Originally distance entered the model because it could be used as an approximation of transportation costs and transport time. Also, distance was used as a measure for the "mental" distance of exporters and importers that increases with distance. New and challenging measures of intangible distances related to different legal and economic institutions, different cultures, and different technologies have recently been added to the gravity model (Martínez-Zarzoso and Márquez-Ramos 2005; Dekker *et al.* 2006). The three chapters in this part provide a perspective on distance in all its manifestations, its importance in describing bilateral economic relations between countries, and the need to address measurement issues.

Head and Mayer (Chapter 6) start from the intuitive notion that border effects are too large to be explained on the basis of the trade barriers that are typically being considered in the literature. This is especially true because large border effects also occur in situations where one would expect that trade barriers are negligible (for example on the European internal market). They show that distances, as they appear in the literature, are mismeasured and lead to a systematic overestimation of the border effect and consequently they develop a new measure of "effective" distance that is used as a benchmark against other measures of distances. Typically the extent of overestimation of distance effects is stronger for countries that are located near each other.[22] They apply this method to the case of the US states and EU member countries, where formal barriers to trade are absent. Although the border effect shrinks for these cases, it remains present.

Bosker and Garretsen (Chapter 7) follow a completely different approach, focusing on different functional forms that so far have been used in the trade literature to describe the relation between distance and costs. They test different specifications for trade costs. Bosker and Garretsen apply the concepts of market access and supplier access in the context of the "New Economic Geography" (NEG) wage equation and show that the specification of trade costs can be very important. Market access is insignificant in four out of six specifications and estimated elasticities may differ up to a factor of six. By way of illustration this chapter also provides numerical insights into the impact of the trade cost specification into the transmission of income shocks, a topic that may gain in relevance due to the financial crisis that may provide a natural experiment to test competing theories.

Chapter 8 by Möhlmann *et al.* includes a measure of cultural and institutional distance into a traditional gravity model thus offering an empirical perspective on the non-economic component of border effects. Building on the seminal work by Rauch (1999) they focus on different product categories. The effect of cultural and institutional distances appears to be small at the level of aggregate trade flows, but significant at the level of individual product groups. Möhlmann *et al.* suggest that the tradeoff between FDI and trade – that is also studied in Chapter 2 by Bergstrand and Egger and Chapter 11 by Brakman *et al.* – can explain this result.

[22] This may actually offer an explanation of an anomaly in Bikker's finding in Chapter 5 of a negative coefficient for EU membership because most EU members typically are located near to each other.

1.3.3 Specific applications

One of the great benefits of the gravity model is that its central notion – economic interaction depends positively on masses corrected for distance – can be applied to many different situations and applications. The contributions in this part both indicate the versatility of the gravity model and put its usefulness into perspective.

A first promising branch of extensions is the inclusion of political and diplomatic interaction that may offer a bridge for the cultural and institutional distances studied by Möhlmann *et al.* in Chapter 8. The importance of these international political non-economic factors is an issue that was investigated (but only for broad classes of political and diplomatic conflict and cooperation) in the late 1980s and early 1990s. Pollins (1989) and van Bergeijk (1992) combined bilateral political events data and trade data in gravity models covering the impact of such events on trade flows in the 1960s to the 1980s inclusive. In recent years, following Rose's (2007) seminal paper, specific aspects of political and diplomatic interaction between countries have become the subject of ongoing research, of which Chapters 9 and 10 provide examples.[23]

Rose and Spiegel (Chapter 9) investigate how bilateral and multilateral spillovers from international treaties can help to build a nation's reputation and how this may exert an impact on economic interaction through the trade credit channel. Their contribution broadens the literature of the 1980s and 1990s that dealt with bilateral impacts only. Rose and Spiegel specifically investigate whether participation in international environmental arrangements (IEAs) is beneficial for economic relations, but argue that the basic logic of the argument can be extended to other facets of international relations such as security. Their model indicates that being part of a political network also benefits economic relations. The other side of the coin is that withdrawing from such a network can have (negative) economic implications as well. A gravity model for cross-border asset holdings uncovers statistically significant and small, but economically meaningful, coefficients both for involvement in bilateral and multilateral treaties.

Afman and Maurel (Chapter 10) investigate the impact of embassies and other diplomatic activity on the trade with Eastern Europe in the post Cold War period. This period provides a unique "natural experiment" in which new diplomatic relationships are established (so the number

[23] Rose (2004) provides a challenging perspective on the impact of multilateral economic diplomacy. This gravity analysis of GATT/WTO membership opened another line of research that is discussed by Piermartini and Teh (2005, pp. 47–49). Van Bergeijk (2009, pp. 5–9, 35–41, and 83–86) discusses gravity studies on the trade impact of bilateral political and diplomatic conflict and cooperation.

of embassies actually shows a lot of variation). Their contribution is to provide both the dynamic counterpart to Rose (2007) and the first application of the Anderson and van Wincoop (2004) model to the question of the impact of political and diplomatic factors on bilateral trade. The paper thus provides an interesting basis for comparison between the methodologies that have been used. Importantly, diplomacy can have beneficial effects on trading relations, as illustrated by their findings that the opening of an embassy is equivalent to a 2–12 per cent reduction of ad valorem tariffs and an increase of bilateral trade by 25 per cent.

The application of Brakman *et al.* (Chapter 11) is different in the sense that the gravity model is applied to mergers and acquisitions, a subset of FDI for which Bergstrand and Egger in Chapter 2 develop a general equilibrium model and provide empirical tests for 1990–2000. Distinguishing empirically between the active group of countries that invest abroad and the non-active group of "zero" investors, Brakman *et al.* show that the phenomenon of a stronger influence of distance over time is not limited to trade flows, but also occurs in mergers and acquisitions over the years 1986–2005. Chapter 11 investigates different hypotheses related to the question of whether FDI acts as a substitute for exports in the context of trade barriers, an issue also raised but now answered by Möhlmann *et al.* in Chapter 8, and furthermore addresses an empirical puzzle, see for instance Neary (2008): why do we see more horizontal M&As while at the same time countries become more and more integrated (where the standard theory predicts a negative relation)? The gravity approach uncovers evidence against the export platform motive and in favor of economic integration as a source of reductions in takeover costs.

Boulhol and de Serres (Chapter 12) offer yet another application in which distance and in particular the proximity to areas of dense economic activity feature prominently. Their approach in a sense fits into the two-stage estimation of the wage equation in the NEG literature that is discussed by Bosker and Garretsen in Chapter 7 and who use per capita GDP as a proxy for wage levels. Boulhol and de Serres develop a framework that directly relates per capita GDP to market and supplier access measures derived from gravity equations as they test these measures in augmented Solow models. Their inclusion of physical and human capital offers a more concrete treatment in a well-known theoretical framework than the exogenous technology parameter that most of the literature is using.[24] Like Bosker and Garretsen they consistently find a significant impact of five measures of proximity to markets on GDP per capita which inter alia suggests the need for testing causality in gravity equations that

24 Bosker and Garretsen, for example, assume that technological differences can be captured by a simple i.d.d. error term that is uncorrelated with the other regressors.

are estimated using panel data over a longer time horizon.[25] Boulhol and de Serres use the same analytical framework also to investigate the impact of transportation and communication costs. These findings generally are quantitatively smaller than the measures derived from gravity equations, offering additional evidence from an alternative methodology that multilateral resistance is an important addition to transportation costs that are a significant element of bilateral trade resistance.

1.4 The road ahead

All in all, this book shows that the gravity approach recovered from its apparent midlife crisis in the 1970s and 1980s. The model has survived almost half a century of economic scrutiny. This is not because it fits the data so nicely, but because it has been part and parcel of a constructive methodological development that provided theoretical answers to the intriguing questions that emerged during and due to its application to real world problems. We expect a bright future for the gravity model.

In particular we see four challenges that will help it to stay alive and kicking. A first challenge is the further analysis of disaggregated trade flows both at the level of product groups (sector data) and at the firm level (micro-data). A stimulus of this line of research could be expected when data at the micro-level become more readily available for increasing numbers of countries. A second challenge consists of a detailed analysis of the absence of trade (or in the words of Harrigan 2002, "instances where gravity fails"). A more detailed analysis of zero trade/FDI flows is necessary when the analysis takes place at increasingly lower levels of aggregation. In addition we expect that the collapse of trade and FDI in 2008–09 will stimulate investigations on different developments in bilateral trade and investment flows during the financial crisis. An interesting topic will be whether the reduction of trade and investment volumes occurs at the extensive or the intensive margin. This is a question that can only be answered by estimations using micro-data. A third challenge is the further investigation into intangible barriers to trade and investment, such as the political co-determinants of trade and new modes of distance. This is also relevant for our fourth challenge, which is identifying and understanding the substantial heterogeneity that exists under the veil of the averages and estimated coefficients. Typically, country experiences diverge, and institutions, economic policies, culture, and even religion have been suggested as an explanation for heterogeneity.

[25] Boulhol and de Serres take endogeneity of trade openness into account in their analysis and use instrument variables in the growth Solow regressions.