



# Working Papers

[www.cesifo.org/wp](http://www.cesifo.org/wp)

## Gravity, Distance, and International Trade

Scott L. Baier  
Amanda Kerr  
Yoto V. Yotov

CESIFO WORKING PAPER NO. 6357

CATEGORY 8: TRADE POLICY

FEBRUARY 2017

*An electronic version of the paper may be downloaded*

- *from the SSRN website:* [www.SSRN.com](http://www.SSRN.com)
- *from the RePEc website:* [www.RePEc.org](http://www.RePEc.org)
- *from the CESifo website:* [www.CESifo-group.org/wp](http://www.CESifo-group.org/wp)

ISSN 2364-1428

# Gravity, Distance, and International Trade

## Abstract

We review and interpret the main theoretical developments in the gravity literature from its very early, a-theoretical applications to the latest structural contributions. We also discuss challenges and implement methods to estimate empirical gravity equations. We finish with a presentation and examples of numerical simulations with the structural gravity model. Throughout the analysis we attempt to emphasize the links and importance of transportation costs for the trade literature and we outline avenues where we believe interdisciplinary contributions between the international trade and transportation economics fields will be most valuable.

JEL-Codes: F100, F430, O400.

Keywords: structural gravity, estimation, simulation, transportation costs.

*Scott L. Baier*

*John E. Walker Department of Economics  
Clemson University  
USA – Clemson, SC 29634  
sbaier@clemson.edu*

*Amanda Kerr*

*John E. Walker Department of Economics  
Clemson University  
USA – Clemson, SC 29634  
ackerr@clemson.edu*

*Yoto V. Yotov*

*School of Economics  
Drexel University  
USA – Philadelphia, PA 19104  
yotov@drexel.edu*

February 12, 2017

This essay was prepared to become a chapter in the Handbook of International Trade and Transportation with Editors Bruce A. Blonigen and Wesley W. Wilson. We are indebted to Jeffrey Bergstrand, Bruce Blonigen, Mario Larch, and Jeffrey Wooldridge for their thoughtful comments and useful suggestions on earlier drafts of this manuscript. All errors are our own.

# Contents

<b>1</b>	<b>Introduction: Motivation and Goals</b>	<b>3</b>
<b>2</b>	<b>Structural Gravity: Theoretical Foundations</b>	<b>5</b>
2.1	When Trade First Met Gravity . . . . .	5
2.2	The Rise of Structural Gravity . . . . .	11
2.3	All Roads Lead to Rome: A Generalized Gravity System . . . . .	24
2.4	Extensions: Tariffs, Intermediates, and Sectors . . . . .	27
<b>3</b>	<b>Estimating Structural Gravity</b>	<b>33</b>
3.1	Structural Gravity Specification . . . . .	34
3.2	Structural Gravity Estimation . . . . .	36
<b>4</b>	<b>Numerical Exercises and Counterfactuals Analysis</b>	<b>51</b>
4.1	GE Analysis with Structural Gravity . . . . .	51
4.2	On the GE Effects of a FTA between Great Britain and US . . . . .	55
4.3	On the GE Impact of Globalization, 1986-2006 . . . . .	57
<b>5</b>	<b>Conclusion</b>	<b>59</b>

# 1 Introduction: Motivation and Goals

For more than 50 years, the gravity equation of international trade has been the workhorse model for empirical issues in international trade. Supported by an intuitive correspondence between Newton’s universal law of attraction between two objects in physics and the bilateral trade flows, the early applications of the gravity equation in international trade specified a proportional association between bilateral trade flows and the *economic masses* of the two objects and negative association between bilateral trade flows and distance. Given the analogy, trade researchers modeled gravity as a log-linear relationship between gross domestic products, distance, and other variables that induce frictions between trading partners. The results from these early studies were encouraging. For example, many studies found the elasticity of bilateral trade with respect to importer or exporter gross domestic product was *close* to unity, and the elasticity of bilateral trade with respect to distance was negative and close to  $-1$ . Not all trade economists, however, were quick to embrace the gravity equation as an appropriate framework to analyze trade flows. Foremost among the concerns was the lack of a sound theoretical framework to provide guidance for empirical specification. In response to these concerns, there have been a number of important theoretical contributions over the last 40 years that have provided structural basis for the empirical gravity specification and have provided more clarity in terms of model specification. Nearly all of these models lead to a similar relationship between bilateral trade flows, the (relative) economic size of the bilateral pair, some measures of trade costs, and measured indexes for multilateral importer and exporter trade costs.

As with any review essay, the choice of topics covered will be selective and there will be undoubtedly some overlap with other review essays and some important contributions will not be discussed due to space limitations.<sup>1</sup> In Section 2, we provide a brief historical overview of the gravity model and discuss how the model has evolved over time and how the theoretical contributions have helped to shape the empirical specification of the gravity model. We begin with the seminal contributions of Anderson (1979) who employed an Armington demand framework, and Krugman (1979, 1980), Helpman (1981), Bergstrand (1985), Bergstrand (1989), and Dearnorff (1998) who derived the gravity equation from a setting where the market structure was one of monopolistically competitive firms or where factor proportions influenced bilateral trade flows. While these models provided a theoretical framework for the gravity model, the early applications were partial equilibrium, reduced-form specifications that did not incorporate other

---

<sup>1</sup>A partial list of recent surveys and books includes Bergeijk and Brakman (2009), Anderson (2011), Bergstrand and Egger (2011), Benedictis and Taglioni (2011), Head and Mayer (2014), Costinot and Rodríguez-Clare (2014), Donaldson (2015), and Yotov et al. (2016). Indeed these essays have helped shape our approach to gravity and the contributions of these reviews are reflected throughout this essay. Similarly, given the voluminous gravity literature, it is beyond our abilities to review and give credit to all related work. However, we have tried to acknowledge the main contributions that have left significant marks in the development of the gravity literature.

important general equilibrium features of the models that, when introduced, would result in a more unified empirical framework. The structural gravity model, as discussed in Anderson and van Wincoop (2001), incorporated market goods clearing to show how bilateral trade costs could be consistently aggregated to form measures of trade costs that are related to market access for buyers and sellers. We show that the structural gravity specification naturally arises in different settings that include richer supply side features like the the multi-country Ricardian setting in Eaton and Kortum (2001), the monopolistic competition frameworks of Krugman (1979) and Bergstrand (1985), and heterogenous firm models pioneered by Melitz (2003). Section 2 concludes with a brief discussion of how the basic gravity model is modified to include trade costs that create rents, intermediate goods, and multiple sectors.

In Section 3, we discuss different approaches to estimating trade costs given the theoretical models discussed in Section 2. Given that an increasing number of papers estimate the gravity model using a cross-section of bilateral pairs over time, our empirical section presents results for cross-sectional gravity model estimates as well as results using panel estimation techniques. Until recently, nearly all empirical applications of the gravity model estimated a simple log-linear relationship between bilateral trade, economic size, and trade costs.<sup>2</sup> Accordingly, it is natural to begin with the conditions necessary for consistent estimation of the parameters using ordinary least squares (OLS). If these assumptions are satisfied, cross-sectional estimates or pooled OLS will yield consistent estimates as long as one properly controls for multilateral resistance.<sup>3</sup> In many instances, the assumptions required to obtain consistent estimates using OLS are violated. Recent research has focused on accounting for *endogenous* right-hand side variables, selection, and heteroskedasticity that may lead to inconsistent estimates in the log-linear model. We address potential methods to obtain consistent estimates in the presence of endogenous right-hand side variables. We follow the procedures outlined in Baier and Bergstrand (2007) to account for endogeneity. In addition to the endogeneity, researchers must also understand how to work with trade data when there are many instances of zero, missing, or non-reported trade. Throughout section 3 we discuss different approaches to handling zero trade flows. In some instances, OLS on log linear trade flows can still consistently estimate the parameters of interest. In many instances, not appropriately accounting for the zero or missing trade flows can result in inconsistent parameter estimates. Helpman et al. (2008) address this issue by employing a *Heckman-style* approach that can account for zero trade flows and account for the type of firm heterogeneity discussed in Melitz (2003).

Santos Silva and Tenreyro (2006) provide an alternative to OLS that can account for zeros and for the presence of heteroskedasticity in the log-linear trade model, which can result in

---

<sup>2</sup>Notable exceptions include Frankel (1997) and Soloaga and Winters (2001) where the model was estimated using non-linear least squares or a Tobit to account for zero trade flows.

<sup>3</sup>See Baldwin and Taglioni (2006) and Yotov et al. (2016) for more detailed discussions on the importance and benefits of controlling for the multilateral resistance terms.

biased and inconsistent estimates of the impact of trade costs when not accounted for. These authors propose using a Pseudo-Poisson Maximum Likelihood (PPML) estimator. We discuss the PPML estimator and briefly discuss other generalized linear model estimators that yield consistent estimates.<sup>4</sup> In this chapter, we focus primarily on using the Pseudo-Poisson Maximum Likelihood Estimator because of the association between the theoretical model and the empirical estimate.<sup>5</sup> However, we suggest that researchers perform functional form and efficiency tests on other generalized linear models.

In Section 4, we use the estimated trade cost parameters and embed them into the gravity model to simulate the general equilibrium effects of changes in trade costs. In the first simulation, we show the trade and welfare effects from an FTA between Great Britain and the United States. As expected, the Great Britain-US FTA diverts trade from other destinations. However, the effects on countries outside of the agreement are relatively small. For the second simulation, we calculate the trade and welfare effects from *globalization*. In particular, we use the estimates from Section 3 which capture the increased international trade that has occurred over time through globalization and calculate what would happen to trade flows and welfare if *globalization* did not occur. Section 5 concludes.

## 2 Structural Gravity: Theoretical Foundations

The objective of this section is to review the evolution of the gravity theory of international trade from its early foundations to most recent times. The section starts with a review of the first and earliest trade gravity theories. Then we derive, discuss, and compare four of the most influential theoretical foundations and we offer a general representation of the structural gravity system on international trade. The section concludes with several theoretical extensions.

### 2.1 When Trade First Met Gravity

Owing to solid theoretical foundations and remarkable empirical success, the gravity equation is probably the most celebrated empirical model in (international) economics. Another very attractive feature of the trade gravity model, which has contributed significantly to the popularity of the gravity model, is its intuitive appeal. Just like Newton's Law of Universal Gravitation establishes that gravitational force between objects is proportional to the product of their masses and inversely proportional to the square of the distance between them, the gravity theory of trade suggests that the trade flows between two countries will be proportional to the product

---

<sup>4</sup>These pseudo maximum likelihood estimators all yield consistent estimates as long as the *mean* is properly specified. See Egger and Staub (2014) for a more detailed discussion of gravity using GLM.

<sup>5</sup>The online appendix contains additional results.

of their economic sizes and inversely proportional to the trade frictions between them. It was this intuitive appeal that has motivated hundreds of papers to use and apply naive, a-theoretical empirical gravity equations in order to study the impact of various determinants of trade flows and immigration for a century before the introduction of the first theoretical foundations of the gravity theory of trade. Ravenstein (1885) was the first to use a gravity-type relationship in order to study immigration patterns, while Tinbergen (1962) was the first to apply a gravity equation in order to study the impact of trade policy. The empirical success of these and other early gravity applications stimulated policy interest and academic curiosity that led to the development of the theoretical foundations of the gravity model of international trade, which we present next.

## The Armington-CES Gravity Model

Anderson (1979) is widely credited as the first paper to provide a theoretical foundation for the empirical gravity equation.<sup>6</sup> The key building blocks in the Anderson paper are homothetic utility functions and market clearing conditions. Anderson shows that in a world where preferences of the representative agent in each country are characterized by Cobb-Douglas utility functions over each country's distinct (basket of) good(s), bilateral trade flows can be expressed as

$$X_{ij} = \frac{Y_i Y_j}{Y_W}, \quad (1)$$

where  $X_{ij}$  represents bilateral exports from country  $i$  to country  $j$ ,  $Y_i$  ( $Y_j$ ) is the income in country  $i$  ( $j$ ), and  $Y_W$  is world income.<sup>7</sup> While this expression is quite intuitive, (bilateral) trade costs do not enter into this equation. Thus, the intuitive interpretation of specification (1) is that it describes the volume of frictionless trade. Using more elaborate structures, in his early work Anderson (1979) shows how a gravity-like expression emerges when multiple sectors and non-zero trade costs are introduced. The most commonly adopted model from Anderson's framework is the one sector (aggregate) Armington model with Constant Elasticity of Substitution (CES) consumer preferences. The Armington (1969) assumption implies that each country produces a different variety of the aggregate good. For consistency throughout the analysis, we make a small expositional departure from the original notation in Anderson (1979) and, instead of endowment economies, we define the total production of good  $Q_i$  in country  $i$  as given by the production

---

<sup>6</sup>Other early gravity theories, some of which we discuss in more detail later in this section, include Krugman (1979, 1980), Bergstrand (1985, 1989), Helpman and Krugman (1985), and Deardorff (1998). We refer the reader to Leamer and Levinsohn (1995) for reviews of the early gravity literature and its development.

<sup>7</sup>In an online appendix, we take a first look at gravity data assuming frictionless trade. We show a strong correlation between import trade shares and the exporter's share of world GDP. While the correlation is high it is far from perfect. We then allow for a simple measure of trade frictions in the spirit of Head and Ries (2001) and Novy (2013b) and show how the standard gravity covariates are correlated with these measured trade costs.

function  $Q_i = A_i L_i$ , where  $L_i$  is the labor endowment and  $A_i$  is the exogenous technology in country  $i$ .<sup>8</sup> Preferences for consumers in country  $j$  are represented by the CES-utility function

$$U_j = \left( \sum_i \beta_i^{\frac{1-\sigma}{\sigma}} c_{ij}^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}}, \quad (2)$$

where  $c_{ij}$  is consumption of country  $i$ 's goods in country  $j$ ,  $\beta_i$  is an exogenous preference parameter, and  $\sigma > 1$  is the elasticity of substitution across varieties from different countries. Combined with a standard budget constraint, the CES preferences imply that the demand for good  $i$  in country  $j$  is given by

$$c_{ij} = \beta_i^{1-\sigma} \left( \frac{p_{ij}}{P_j} \right)^{-\sigma} \frac{Y_j}{P_j}, \quad (3)$$

where expenditures are given by  $Y_j$ ,<sup>9</sup> and the corresponding CES price index is given by

$$P_j = \left[ \sum_i (\beta_i p_{ij})^{1-\sigma} \right]^{\frac{1}{1-\sigma}}.$$

Assuming iceberg trade costs as in Samuelson (1952), the price of a good shipped from country  $i$  to country  $j$  is given by  $p_{ij} = p_i t_{ij}$  where  $p_i$  is the factory-gate price and  $t_{ij}$  is the amount of country  $i$ 's good that needs to be shipped from  $i$  so that one unit of the good produced in country  $i$  lands in country  $j$  ( $t_{ij} \geq 1$ ). Applying the definition of trade costs and expressing the demand for good  $i$  in country  $j$  in value terms results in the following bilateral trade flow equation:

$$X_{ij} = p_{ij} c_{ij} = \left( \frac{\beta_i p_i t_{ij}}{P_j} \right)^{1-\sigma} Y_j. \quad (4)$$

Equation (4) resembles a gravity equation as it includes an exporter-specific term, an importer-specific term, and a bilateral trade cost term. We refer to this type of gravity specification as a *reduced form* gravity model because it is a solution for bilateral trade flows derived from a partial equilibrium subsystem of equations from a general equilibrium model. However, as we demonstrate later, further improvements in gravity theory have led to an expression that resembles the gravity equation from physics much closer. Deardorff (1998) takes a step in this direction. To this end, choose the quantity of each good in each country so that the  $p_i = 1 \forall i$ ,

---

<sup>8</sup>Due to the endowment or exogenous production structure assumptions, Anderson's framework and consequent extensions that adopt these assumptions are known in the gravity literature as *Demand-side* gravity models. As demonstrated below, a series of frameworks with more elaborate production structures, known as *Supply-side* gravity models, lead to the same benchmark structural gravity representation.

<sup>9</sup>Throughout this section we assume balanced trade, so that aggregate income ( $Y_i$ ) is equal to aggregate expenditures ( $E_i$ ). Alternatively, it is straightforward to allow for exogenous trade imbalances without any implications for our analysis and conclusions. This will become clear at the end of this section, where we present and discuss the properties of the structural gravity system in changes.



and the price index in country  $j$  can be expressed as

$$\delta_j = \left( \sum_i (\beta_i t_{ij})^{1-\sigma} \right)^{\frac{1}{1-\sigma}},$$

Define relative trade costs as

$$r_{ij} = t_{ij}/\delta_j$$

and substitute this expression into equation (4) to obtain

$$X_{ij} = \beta_i^{1-\sigma} (r_{ij})^{1-\sigma} Y_j, \quad (5)$$

Market clearing implies that, at delivered prices, total production in country  $i$  equals country  $i$ 's shipments to all markets, including  $i$ ; that is,

$$Y_i = \sum_k \beta_i^{1-\sigma} (r_{ik})^{1-\sigma} Y_k.$$

Divide the above expression by world GDP and solve for  $\beta_i$ :

$$\beta_i^{1-\sigma} = \frac{Y_i}{Y_W} \left[ \sum_k (r_{ik})^{1-\sigma} Y_k \right]^{-1}. \quad (6)$$

Substitute equation (6) into equation (5) to obtain

$$X_{ij} = \frac{Y_i Y_j}{Y_W} \left( r_{ij} \right)^{1-\sigma} \left[ \sum_k (r_{ik})^{1-\sigma} \frac{Y_k}{Y_W} \right]^{-1}. \quad (7)$$

Equation (7) already resembles a gravity equation more closely because it relates bilateral trade flows to the economic sizes of the two countries, as captured by the term  $\frac{Y_i Y_j}{Y_W}$ , and to the (relative) bilateral trade costs between them  $(r_{ij})^{1-\sigma}$ . In addition, equation (7) enables us to emphasize important features of the economic gravity model. First, comparison between equation (7) and equation (1) reveals that the former specification nests the latter. Combined with our earlier interpretation of the term  $\frac{Y_i Y_j}{Y_W}$  as the volume of frictionless trade, this implies that, in combination, the rest of the terms on the righthand side of equation (7) denote that the total trade costs drive a wedge between realized and frictionless trade. Second, equation (7) reveals that the total trade cost term determines the volume of bilateral trade between two countries. It includes a measure of importer-relative bilateral frictions,  $(r_{ij})^{1-\sigma}$ , as well as a multilateral country-specific component  $\left[ \sum_k (r_{ik})^{1-\sigma} \frac{Y_k}{Y_W} \right]^{-1}$ , which is constructed as a share weighted index of country  $i$ 's trade costs with the rest of the world. The intuition behind this multilateral trade cost term is that bilateral trade between two countries depends on the direct frictions between them but also on the overall remoteness from the rest of the world. Motivated by these early

theoretical foundations, several empirical papers created ad-hoc empirical *remoteness indexes* to control for this share-weighted trade cost measure in gravity estimations.<sup>10</sup>

## Monopolistic Competition

While the Armington model provided a theoretical structure for the gravity equation on the *demand side*, the monopolistically competitive models of trade developed by Krugman (1979, 1980), Bergstrand (1985, 1989), and Helpman and Krugman (1985) offer theoretical foundations of the gravity model with a more elaborate *supply side* with multiple (identical) firms, where each firm produces a different variety, and the number of firms are determined endogenously within the model. The key building blocks of these models are that consumers have Dixit-Stiglitz love-of-variety preferences, the multiple monopolistically-behaved producers face a fixed cost of entry, and entry into the market is free.

As in the Armington model, the Dixit-Stiglitz preferences are characterized by a CES-utility function defined over differentiated goods

$$U_j = \left( \sum_{\omega} c_j(\omega)^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}},$$

where  $c_j(\omega)$  represents consumption of variety  $\omega$  in country  $j$ . These preferences give rise to the following demands for each variety  $\omega$

$$\left( \frac{p_j(\omega)}{P_j} \right)^{-\sigma} \frac{Y_j}{P_j},$$

The production side of this model differs from the Armington model in that there are increasing returns to scale that arise because of the fixed cost of entry. It is typically assumed that producers in country  $i$  have access to the same technology, where production technology is given by

$$q_i(\omega) = A_i(l_i(\omega) - f_i),$$

where  $q_i(\omega)$  is production of good  $\omega$  in country  $i$ ,  $A_i$  is the technology available in country  $i$ ,  $l_i(\omega)$  is the labor in country  $i$  employed to produce good  $\omega$ , and  $f_i$  is the fixed cost of production associated with producing good  $i$  (common across all firms in country  $i$ ).

As in the Armington model, goods shipped from country  $i$  to country  $j$  are subject to *iceberg trade costs*. Since each firm has access to the same technology and because goods enter into the utility function symmetrically, each producer produces a unique variety.

---

<sup>10</sup>Some notable examples include Wei (1996) and Wolf (2000). See Anderson (2011), Head and Mayer (2014), and Yotov et al. (2016) for related discussions.

The total wage bill for producers of good  $\omega$  in country  $i$  can be expressed as

$$w_i l_i(\omega) = w_i \left( \frac{q_i(\omega)}{A_i} + f_i \right),$$

substituting for total demand inclusive of iceberg trade costs,  $q_i(\omega) = \sum_k t_{ik} \left( \frac{p_j(\omega)}{P_j} \right)^{-\sigma} \frac{Y_j}{P_j}$ , implies that the wage bill for producers of good  $\omega$  is given by

$$w_i l_i(\omega) = \sum_k \frac{w_i t_{ik}}{A_i} \left( \frac{p_j(\omega)}{P_j} \right)^{-\sigma} \frac{Y_j}{P_j} + w_i f_i.$$

Given these costs, the firm producing good  $\omega$  in country  $i$  solves the following maximization problem:

$$\pi_i = \max_{p_i} \left\{ \sum_k \left( \frac{p_{ik}}{P_k} \right)^{1-\sigma} Y_k - \sum_k \left( \frac{w_i t_{ik}}{A_i} \right) \left( \frac{p_{ik}}{P_k} \right)^{-\sigma} \frac{Y_k}{P_k} - w_i f_i \right\},$$

where profits of the firm are defined as the difference between total revenues less variable and fixed costs. Profit maximization implies that the price of the good is a mark-up over marginal costs given by

$$p_{ik} = \left( \frac{w_i t_{ik}}{A_i \rho} \right), \quad \text{where } \rho = \frac{\sigma - 1}{\sigma}$$

so that profits can be expressed as

$$\pi_i = \sum_k \left( \frac{w_i t_{ik}}{A_i \rho P_k} \right)^{1-\sigma} \frac{Y_k}{\sigma} - w_i f_i.$$

Free entry implies that profits are zero in equilibrium so that

$$\frac{1}{\sigma - 1} \frac{w_i}{A_i} \underbrace{\sum_k t_{ik} \left( \frac{w_i t_{ik}}{A_i \rho P_k} \right)^{-\sigma} \frac{Y_k}{P_k}}_{q_i} = w_i f_i \Rightarrow q_i = A_i f_i (\sigma - 1).$$

Given the symmetry in demands and the cost structure, each firm in country  $i$  produces the same quantity of output  $A_i f_i (\sigma - 1)$ . The model is closed by assuming balanced trade and the labor market clears. Labor market clearing implies

$$L_i = \sum_i N_i \left( \frac{q_i}{A_i} + f_i \right),$$

which, after substituting  $q_i = A_i f_i (\sigma - 1)$ , pins down the number of firms in each country:

$$N_i = \frac{L_i}{\sigma f_i}.$$

Finally, we turn to bilateral trade flows, which are given by

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

A reduced form gravity equation can be obtained by noting that the value of production in country  $i$  is given by  $Y_i = N_i p_i q_i$ . We substitute it in the previous expression to obtain

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

Finally, employing the definition of  $q_i = A_i f_i (\sigma - 1)$  and assuming that production in all countries is subject to the same technologies,  $A_i = A$ , and the same fixed costs,  $f_i = f$ , the resulting bilateral trade flow equation becomes

$$X_{ij} = B_0 p_i^{-\sigma} \left( \frac{t_{ij}}{P_j} \right)^{1-\sigma} Y_i Y_j, \quad (8)$$

where  $B_0 = \left( (\sigma - 1) A f \right)^{-1}$ . Equation (8) resembles a gravity equation because it implies that trade flows are proportional to the economic sizes of the two countries and inversely proportional to the bilateral trade frictions between them. Equation (8) departs from the intuitive (and naive) gravity representation because it includes two additional price terms, one on the exporter side ( $p_i^{-\sigma}$ ) and one on the importer side ( $P_j^{1-\sigma}$ ). More recent developments in the theoretical gravity literature translate these price terms into intuitive structural gravity components. We present and discuss those developments next.

## 2.2 The Rise of Structural Gravity

In this section, we show how four of the most widely used theoretical models in international trade, which arise from very different underlying assumptions, deliver very similar *structural gravity systems*. These four models include the two gravity frameworks discussed above, i.e. the Armington-CES gravity model of Anderson (1979) and the monopolistic competition models of Krugman (1979, 1980) and Bergstrand (1985, 1989). In addition, we show that the multi-country Ricardian model of Eaton and Kortum (2002) and the heterogeneous firms framework of Melitz (2003) both lead to structural gravity systems that are remarkably similar to those obtained from the Armington-CES and the monopolistically competitive models. For all of these models we show how trade costs are linked to factor payments in a general equilibrium setting. After reviewing the four theoretical structural gravity foundations, we offer a general representation of the structural gravity system, which will serve as the basis for our estimation analysis and numerical exercises in subsequent sections. The section concludes with a presentation of several

useful extensions of the structural gravity model. For simplicity, throughout the main analysis in this section, we assume that there is only one factor of production, but the links between trade costs and factor prices are present when there are multiple factors of production and multiple sectors. We briefly discuss these issues in the concluding sub-section.

### Structural Gravity with the Armington-CES Model

Anderson and van Wincoop (2003) elevate the Armington-CES trade model described above to a new structural level by showing how the market clearing conditions can be used to derive a relationship between trade, bilateral trade costs, the economic size of the two countries, and consistently aggregated measures of the trade frictions for importers and exporters. Anderson and van Wincoop referred to these aggregate measures of trade frictions as *Inward Multilateral Resistances (IMRs)* and *Outward Multilateral Resistances (OMRs)*. To see the impact of their approach, once again consider the Armington model of trade, where bilateral trade is given by

$$X_{ij} = \left( \frac{\beta_i p_i t_{ij}}{P_j} \right)^{1-\sigma} Y_j, \quad (9)$$

and the corresponding CES price aggregator is:

$$P_j = \left[ \sum_i (\beta_i p_i t_{ij})^{1-\sigma} \right]^{\frac{1}{1-\sigma}}. \quad (10)$$

Use the definition of bilateral trade flows from equation (9) and impose market clearing, which implies that, for any country  $i$ , total production equals total consumption, or

$$Y_i = \sum_j X_{ij} = (\beta_i p_i)^{1-\sigma} \left[ \sum_j \left( \frac{t_{ij}}{P_j} \right)^{1-\sigma} Y_j \right]. \quad (11)$$

Define *Outward Multilateral Resistance* as

$$\Pi_i = \left[ \sum_j \left( \frac{t_{ij}}{P_j} \right)^{1-\sigma} Y_j \right]^{\frac{1}{1-\sigma}}, \quad (12)$$

which is a GDP-weighted average of the relative trade frictions that producers in country  $i$  face when they export their goods from country  $i$  to country  $j$ . If large markets are relatively nearby, the OMR term will be lower than if large markets are farther away.

Use equations (11) and (12) to restate the market clearing conditions as

$$(\beta_i p_i)^{1-\sigma} = Y_i \Pi_i^{\sigma-1}. \quad (13)$$

Plug the solution for  $(\beta_i p_i)^{1-\sigma}$  into the CES price index from equation (10) to obtain a new expression, which Anderson and van Wincoop (2003) label as the *Inward Multilateral Resistance*

$$P_j = \left[ \sum_i \left( \frac{t_{ij}}{\Pi_i} \right)^{1-\sigma} Y_i \right]^{\frac{1}{1-\sigma}}. \quad (14)$$

The IMR measures the relative trade cost frictions that importers face when bringing goods into the country. As with the outward multilateral resistance term, the inward multilateral resistance term will be lower when the importing country is relatively close to large markets.

Two steps complete the derivation of the structural Armington-CES gravity system. First, use equation (13) to substitute the solution of  $(\beta_i p_i)^{1-\sigma}$  in the trade flows equation (9) to obtain

$$X_{ij} = \left( \frac{t_{ij}}{\Pi_i P_j} \right)^{1-\sigma} Y_i Y_j. \quad (15)$$

Finally, to highlight how the structural gravity equation provides a link between trade costs and the factor prices, assume that labor is the only input in the production process, the value of output is given by  $Y_i = p_i A_i L_i$ , and the preference parameter  $\beta_i$  is common across all markets and normalized to one for simplicity.<sup>11</sup> Noting that  $W_i L_i = p_i A_i L_i$ , and substituting for prices and income, the market clearing condition can be arranged to yield the following relationship between factor prices, productivity, and trade frictions:

$$W_i = B_A \left( \frac{A_i}{\Pi_i} \right)^{\frac{\sigma-1}{\sigma}} \quad (16)$$

where  $B_A = L_i^{-1/\sigma}$ .

Given the relationship between income and expenditures, the structural gravity system is ex-

---

<sup>11</sup>Alternatively, one can retain the assumption of country-specific preference parameter  $\beta_i$ . In that case, instead of technology, it would be preference-adjusted technology that appears in the market clearing condition.

pressed as

$$X_{ij} = \left( \frac{t_{ij}}{\Pi_i P_j} \right)^{1-\sigma} (W_j L_j)(W_i L_i) \quad \forall i, j \quad (17)$$

$$\Pi_i = \left[ \sum_j \left( \frac{t_{ij}}{P_j} \right)^{1-\sigma} W_j L_j \right]^{\frac{1}{1-\sigma}} \quad \forall i \quad (18)$$

$$P_j = \left[ \sum_i \left( \frac{t_{ij}}{\Pi_i} \right)^{1-\sigma} W_i L_i \right]^{\frac{1}{1-\sigma}} \quad \forall j \quad (19)$$

$$W_i = B_A \left( \frac{A_i}{\Pi_i} \right)^{\frac{\sigma-1}{\sigma}} \quad \forall i. \quad (20)$$

Equation (17) is the structural gravity equation. It implies that bilateral trade flows are proportional to the sizes of the importer and of the exporter as captured by the income terms  $(W_j L_j)$  and  $(W_i L_i)$ , respectively. In addition, equation (17) suggests that bilateral trade flows are inversely proportional to a composite trade cost term,  $\left( \frac{t_{ij}}{\Pi_i P_j} \right)^{1-\sigma}$ , which consists of three components.  $t_{ij}$  denotes bilateral trade costs, which capture the effects of distance, trade policies, cultural ties, etc.<sup>12</sup> Importantly,  $t_{ij}$  also captures the impact of transportation costs and, as discussed in the empirical section below, we view the modeling of transportation costs in the gravity model as an important area for further exploration. The other two components of the total trade cost term are the multilateral resistances of Anderson and van Wincoop (2003), which, as described in Larch and Yotov (2016), are intuitive general equilibrium trade costs that consistently aggregate all bilateral trade costs into two country-specific indexes and decompose the incidence of trade costs and their changes on the consumers and the producers in each country. These aggregated multilateral indexes are given in equation (18) and equation (19).

The last equation from the structural gravity system captures the link between trade costs (and trade liberalization) and factor prices. To ease interpretation, we remind the reader that equation (20) is a restatement of the market clearing condition. It captures two important and intuitive relationships. First, equation (20) captures the inverse relationship between the outward multilateral resistance and factor prices. The intuition for this result is that if producers are faced with lower multilateral resistance to shipping their products, they will internalize part of this favorable change and pass part of it as higher returns to the factors of production. The second intuitive relationship that is captured by equation (20) is that the outward multilateral resistance can be thought of as adjustments to technology. The impact of a given decrease in the OMRs can be mapped uniquely into a corresponding technological improvement that would

---

<sup>12</sup>See Anderson and van Wincoop (2004) for a detailed discussion of the determinants of trade costs in gravity equations.

have resulted in the same effect. Also note that the wage rate is *negatively* related to the size of the population. With perfect competition, an increase in the population increases the amount of production and the increased supply of the country-specific good puts downward pressure on prices (and wages).

## Structural Gravity and Monopolistic Competition

Feenstra (2004, 2015) shows how the Anderson van Wincoop approach can be adapted to the monopolistically competitive models outlined above. As demonstrated above, bilateral trade flows in the monopolistically competitive setting can be expressed as

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

Market clearing implies

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

As before, define the outward multilateral resistance as

$$\Pi_i = \left[ \sum_j \left( \frac{t_{ij}}{P_j} \right)^{1-\sigma} Y_j \right]^{\frac{1}{1-\sigma}} \forall i$$

and substitute it in the market clearing condition to obtain

$$N_i p_i^{1-\sigma} = Y_i \Pi_i^{\sigma-1}. \tag{21}$$

Equation (21) captures the intuitive relationship between output, prices, the number of firms, and outward multilateral resistance. In particular, for a given number of firms, lower *outward multilateral resistance* implies better supplier access which results in higher value of firm's output. Alternatively, for a given value of firm output, a lower multilateral resistance stimulates entry and results in a larger number of firms.

We close the model by substituting for prices and income into equation (21) to obtain the wage equation

$$W_i = B_{MC} \left( \frac{A_i}{\Pi_i} \right)^{\frac{\sigma-1}{\sigma}}$$

where  $B_{MC} = \left( \sigma^{-1} (\sigma - 1)^{\frac{\sigma-1}{\sigma}} \right) f_i^{-\frac{1}{\sigma}}$ .

Substituting the expression for  $N_i p_i^{1-\sigma}$  from equation (21) into the trade flows equation and the corresponding CES price aggregate index and gathering equations gives the structural gravity system



$$X_{ij} = \left( \frac{t_{ij}}{\Pi_i P_j} \right)^{1-\sigma} (W_j L_j)(W_i L_i) \quad \forall i, j \quad (22)$$

$$\Pi_i = \left[ \sum_j \left( \frac{t_{ij}}{P_j} \right)^{1-\sigma} W_j L_j \right]^{\frac{1}{1-\sigma}} \quad \forall i \quad (23)$$

$$P_j = \left[ \sum_i \left( \frac{t_{ij}}{\Pi_i} \right)^{1-\sigma} W_i L_i \right]^{\frac{1}{1-\sigma}} \quad \forall j \quad (24)$$

$$W_i = B_{MC} \left( \frac{A_i}{\Pi_i} \right)^{\frac{\sigma-1}{\sigma}} \quad \forall i \quad (25)$$

System (22)-(25) is remarkably similar to system (17)-(20). In fact the two systems are identical subject to the definition of the parameter  $B_{MC}$ . Equation (25) reveals that in the monopolistically competitive setting, wages are determined in much of the same way as the Armington model. One notable difference is the presence of fixed costs: higher fixed costs in market  $i$  tend to lower the wage. Intuitively, the higher fixed costs reduce the number of varieties produced in country  $i$ . As a result, there will be a lower demand because agents' preferences for varieties; the lower demand for country  $i$ 's goods lowers the wages paid to workers in country  $i$ . With monopolistically competitive markets, wages are not negatively related to the population because as the population increases more varieties are produced. Since consumers have a love of variety, the increase in the demand for the new varieties offsets the supply side effect from more production. As a result, there is no change in the wage rate as the population increases.

## Structural Gravity in a Ricardian Setting

One of the most commonly used theoretical gravity models in the trade literature is the Ricardian model developed by Eaton and Kortum (2002). Unlike the Armington model, where the supply of goods is determined by country of origin, or the monopolistically competitive models, where the fixed costs help pin down the number of firms, the Eaton–Kortum model is a supply-side Ricardian model that extends the two-country Ricardian model with a continuum of goods to a world with many countries. In this setting, bilateral trade flows are determined by probability distributions over technologies and bilateral characteristics of the country. The novel modeling strategy moved the classic Dornbusch et al. (1977) model from the classroom to a practical alternative for computable general equilibrium exercises and allowed researchers and policy makers to better understand how changes in technology and trade costs influence trade, per capita income, and welfare.

As in the Armington model and the monopolistically competitive model, preferences over goods

can be expressed by CES-preferences, which leads to the following familiar demand for varieties

$$c_j(\omega) = \left( \frac{p_j(\omega)}{P_j} \right)^{-\sigma} \frac{Y_j}{P_j}.$$

The point of departure and main contribution of Eaton and Kortum is on the production side. Assuming that labor is the only input and that the production technology exhibits constant returns to scale, the factory gate price for commodity  $\omega$  produced in country  $i$  is given by

$$p_i(\omega) = \frac{W_i}{z_i},$$

where  $z_i$  is the technology of the firm producing good  $\omega$  in country  $i$ . Trade costs are subject to the standard iceberg trade costs assumption and there is no possibility for arbitrage,  $t_{ij} \geq 1$  and  $t_{ij} \leq t_{ik}t_{kj} \forall k \neq j$ . Under perfect competition, consumers in country  $j$  purchase goods from the low cost supplier

$$\mathbf{p}_j(\omega) = \min\{p_{ij}(\omega); i = 1, \dots, N\}.$$

The technical efficiencies of producers of good  $\omega$  in country  $i$  are determined from random draws from a Fréchet distribution given by

$$F_i(z) = \exp\left(-T_i z^{-\theta}\right),$$

where  $T_i$  is a strictly positive country-specific parameter, reflecting the location of the productivity distribution, which in the trade context can be thought of as reflecting comparative advantage, and  $\theta (> 1)$  determines the dispersion of the distribution. Given the assumption about the productivity distribution and that all productivity draws are independent, the distribution for  $\mathbf{p}_j(\omega)$  is given by

$$G_j(\mathbf{p}) = 1 - \exp[-\Phi_j \mathbf{p}^\theta],$$

where

$$\Phi_j = \left\{ \sum_{i=1}^N T_i (w_i t_{ij})^{-\theta} \right\}.$$

The probability that country  $i$  provides a good to country  $j$  at the lowest price is given by

$$\left( \frac{T_i}{\Phi_j} (W_i t_{ij})^{-\theta} \right).$$

Assuming that total expenditures in country  $j$  equal income ( $W_j L_j$ ), then country  $j$ 's bilateral

imports from country  $i$  can be expressed as

$$X_{ij} = \left( \frac{T_i (W_i t_{ij})^{-\theta}}{\Phi_j} \right) W_j L_j. \quad (26)$$

Market clearing implies that

$$W_i L_i = T_i W_i^{-\theta} \left[ \sum_{k=1} \left( t_{ik} \right)^{-\theta} \frac{W_k L_k}{\Phi_k} \right]. \quad (27)$$

Substitute  $T_i W_i^{-\theta} = W_i L_i / \left[ \sum_{k=1} \left( t_{ik} \right)^{-\theta} \frac{W_k L_k}{\Phi_k} \right]$  from the previous expression into equation (26) to obtain

$$X_{ij} = \left( \frac{t_{ij}^{-\theta}}{\left[ \sum_{k=1} \left( \frac{t_{ik}}{\Phi_k} \right)^{-\theta} W_k L_k \right] \Phi_j} \right) W_i L_i W_j L_j.$$

Eaton and Kortum show that  $P_j = \gamma \Phi_j^{-1/\theta}$ , where  $\gamma = \left( \Gamma \left( \frac{\theta - \sigma + 1}{\theta} \right) \right)^{\frac{-1}{\sigma - 1}}$  and  $\Gamma$  is the gamma function. Define the terms in the square brackets of the above expression as the outward multi-lateral resistance, and then substitute for  $\Phi_j$  in the trade flows equation, in the OMR definition, and in the consumer price index. Use the market clearing condition to solve for wages and collect terms to obtain

$$X_{ij} = \left( \frac{t_{ij}}{\Pi_i P_j} \right)^{-\theta} (W_j L_j) (W_i L_i) \quad \forall i, j \quad (28)$$

$$\Pi_i = \left[ \sum_j \left( \frac{t_{ij}}{P_j} \right)^{-\theta} W_j L_j \right]^{-\frac{1}{\theta}} \quad \forall i \quad (29)$$

$$P_j = \left[ \sum_i \left( \frac{t_{ij}}{\Pi_i} \right)^{-\theta} W_i L_i \right]^{-\frac{1}{\theta}} \quad \forall j \quad (30)$$

$$W_i = B_{EK} \left( \frac{\bar{z}_i}{\Pi_i} \right)^{\frac{\theta}{\theta + 1}} \quad \forall i, \quad (31)$$

where  $\bar{z}_i = e^{\frac{0.577}{\theta}} T_i^\theta$  is the geometric mean of  $z_i$ , and it follows that  $B_{EK} = \exp(0.577(1 - \theta))$ .<sup>13</sup> Thus, the Eaton-Kortum system Ricardian model delivers a system of equations that are very similar to the gravity equations generated by the Armington model and the monopolistically competitive models. We note that the first three equations from the Eaton-Kortum system are indeed similar to the previous two structural gravity systems. In this framework, the trade elasticity is associated with the Fréchet distribution parameter ( $\theta$ ) rather than the elasticity of

<sup>13</sup>See Eaton and Kortum (2002) page 1747 for the geometric mean of  $z_i$ .

substitution ( $\sigma-1$ ). Once again, the wage equation (31) captures the intuitive inverse relationship between the outward multilateral resistance and wages. Furthermore, the wage equation that we obtained in the Eaton-Kortum setting only differs from the corresponding Armington and monopolistically competitive equations by the definition of the exogenous composite parameter  $B_{EK}$ .

## Structural Gravity with Heterogeneous Firms

Firms were present in the monopolistically competitive models that we presented and discussed thus far. However, in the standard monopolistically competitive model (i) all firms were assumed to be identical and (ii) they exported to all destinations. These theoretical implications are often rejected in the data. To overcome these difficulties, Melitz (2003) introduced firm heterogeneity into the standard trade model by assuming that firms differ in terms of their productivity and only the most productive firms are able to export. In addition to the differences in productivity across firms, Melitz assumed that there are fixed costs associated with selling in each market as well as a fixed costs associated with free entry. In this section, we demonstrate how the Melitz model delivers a structural gravity system that is remarkably similar to those that we obtained thus far from the Armington, the Monopolistic Competition, and the Ricardian models of trade. Similar to the previous models, the demand structure in the Melitz framework is subject to the standard CES-utility function representation and, as before, this implies that the demand for good  $\omega$  by a consumer in country  $j$  is given by

$$c_j(\omega) = \left( \frac{p_j(\omega)}{P_j} \right)^{-\sigma} \frac{Y_j}{P_j}, \quad (32)$$

where  $P_j$  is the CES price index for a consumer in country  $j$ .

For a producer of good  $\omega$  in country  $i$ , profits from selling in market  $j$  are given by

$$\pi_{ij}(\omega) = \max \left\{ 0, p_j(\omega) \underbrace{\left( \frac{p_j(\omega)}{P_j} \right)^{-\sigma} \frac{Y_j}{P_j}}_{c_{ij}(\omega)} - \frac{W_i t_{ij}}{A_i \varphi} \underbrace{\left( \frac{p_j(\omega)}{P_j} \right)^{-\sigma} \frac{Y_j}{P_j}}_{c_{ij}(\omega)} - W_i f_{ij}^X \right\},$$

where we have assumed, as before, that labor is the only input,  $A_i$  is aggregate productivity in country  $i$ , and  $\varphi$  is firm specific productivity. This implies that the marginal cost for a firm with productivity  $\varphi$  selling goods in market  $j$  is given by  $\frac{W_i t_{ij}}{A_i \varphi}$ . Since the demands for all varieties enter symmetrically into the agent's utility function and the only difference in production techniques is the firm specific technology, it is easier to characterize the firm's profits, the price the firm

charges, and the quantity produced by its productivity. Profit maximization, therefore, implies

$$p_{ij}(\varphi) = \frac{W_i t_{ij}}{\rho A_i \varphi},$$

where  $\rho = (\sigma - 1)/\sigma$ . Profits earned by a firm with productivity  $\varphi$  in country  $i$  that sells to market  $j$  are given by

$$\pi_{ij}(\varphi) = \max \left\{ 0, \left( \frac{W_i t_{ij}}{\rho A_i \varphi P_j} \right)^{1-\sigma} \frac{Y_j}{\sigma} - W_i f_{ij}^X \right\}.$$

Melitz (2003) defines the cutoff productivity as  $\varphi_{ij}^*$  where  $\pi_{ij}(\varphi_{ij}^*) = 0$  or

$$\left( \frac{W_i t_{ij}}{\rho A_i P_j} \right)^{1-\sigma} \frac{Y_j}{\sigma} = (\varphi_{ij}^*)^{1-\sigma} W_i f_{ij}^X \quad (33)$$

for potential firms in country  $i$  selling output in country  $j$ .<sup>14</sup> Expected profits among firms operating in country  $i$  that are active in country  $j$  are given by

$$E[\pi_{ij} | \varphi \geq \varphi_{ij}^*] = \int_{\varphi_{ij}^*} \left( \left( \frac{W_i t_{ij}}{\rho A_i P_j} \right)^{1-\sigma} \frac{Y_j}{\sigma} \varphi^{\sigma-1} - W_i f_{ij}^X \right) \frac{g(\varphi)}{1 - G(\varphi_{ij}^*)} d\varphi,$$

where  $G(\varphi)$  is the cumulative density function and  $g(\varphi)$  is the density function so that  $\frac{g(\varphi)}{1 - G(\varphi_{ij}^*)}$  is the conditional density function. Following Chaney (2008) and assuming a Pareto distribution where the cumulative density is given by  $G(\varphi) = 1 - \left( \frac{\bar{\varphi}}{\varphi} \right)^\kappa$  and is defined on the support  $[\bar{\varphi}, \infty)$ , expected profits can be expressed as

$$E[\pi_{ij} | \varphi \geq \varphi_{ij}^*] = \frac{\kappa}{\kappa - \sigma + 1} \left( \frac{W_i t_{ij}}{\rho A_i \varphi_{ij}^* P_j} \right)^{1-\sigma} \frac{Y_j}{\sigma} - W_i f_{ij}^X.$$

Using equation (33), which implies that  $\left( \frac{W_i t_{ij}}{\rho A_i \varphi_{ij}^* P_j} \right)^{1-\sigma} \frac{Y_j}{\sigma} = W_i f_{ij}^X$ , enables us to express the expected profits for firms in country  $i$  that are active in country  $j$  as

$$E[\pi_{ij} | \varphi \geq \varphi_{ij}^*] = \frac{\sigma - 1}{\kappa - \sigma + 1} W_i f_{ij}^X$$

and the expected profits from sales in market  $j$  among all active firms as

$$E[\pi_{ij} | \varphi \geq \varphi_{ij}^*] = [1 - G(\varphi_{ij}^*)] \frac{\sigma - 1}{\kappa - \sigma + 1} W_i f_{ij}^X = \left( \frac{\bar{\varphi}}{\varphi_{ij}^*} \right)^\kappa \frac{\sigma - 1}{\kappa - \sigma + 1} W_i f_{ij}^X.$$

---

<sup>14</sup>As is standard, we assume that the zero cut-off productivity is lower in the domestic market than any other market; that is,  $\varphi_{ii}^* < \varphi_{ij}^* \forall j \neq i$ .

Aggregating across all markets delivers the expression for the firm's expected total profits

$$E(\pi_i) = \frac{\sigma - 1}{\kappa - \sigma + 1} \sum_j \left( \frac{\bar{\varphi}}{\varphi_{ij}^*} \right)^\kappa W_i f_{ij}^X.$$

Free entry implies that expected profits, conditional on a productivity draw at least equal to  $\varphi_{ii}^*$ , are equal to the *fixed cost of entry*, which is expressed below in terms of domestic labor units

$$[1 - G(\varphi_{ii}^*)]E(\pi_i | \varphi \geq \varphi_{ii}^*) = W_i F_i^E,$$

or equivalently

$$\frac{\sigma - 1}{\kappa - \sigma + 1} \sum_j \left( \frac{\varphi_{ii}^*}{\varphi_{ij}^*} \right)^\kappa W_i f_{ij}^X = \frac{W_i F_i^E}{(\bar{\varphi}/\varphi_{ii}^*)^\kappa}.$$

Labor market clearing implies that the equilibrium number of firms is given by  $M_i = \frac{(\sigma-1)L_i}{\kappa\sigma F_i^E} \left( \frac{\bar{\varphi}}{\varphi_{ii}^*} \right)^\kappa$ .

Given the mass of firms, bilateral trade can be expressed as

$$X_{ij} = M_i \frac{1 - G(\varphi_{ij}^*)}{1 - G(\varphi_{ii}^*)} \int_{\varphi_{ij}^*}^{\infty} \left( \frac{W_i t_{ij}}{\rho A_i P_j} \right)^{1-\sigma} Y_j \kappa \varphi^{-(\kappa-\sigma+2)} (\varphi_{ij}^*)^\kappa d\varphi,$$

substituting  $\left( \frac{W_i t_{ij}}{\rho A_i P_j} \right)^{1-\sigma} EY_j = (\varphi_{ij}^*)^{1-\sigma} W_i f_{ij}^X \sigma$  and integrating yields

$$X_{ij} = \underbrace{\left[ \frac{(\sigma - 1)L_i}{\kappa\sigma F_i^E} \left( \frac{\varphi_{ij}^*}{\bar{\varphi}} \right)^{-\kappa} \right]}_{M_{ij}} \left( \frac{W_i f_{ij}^X \kappa \sigma}{(\kappa - \sigma + 1)} \right). \quad (34)$$

The term in brackets is the extensive margin of trade and expresses how changes in trade costs will impact the mass of exporters from country  $i$  to country  $j$ . Both fixed and variable trade costs impact the extensive margin of trade. The second term captures the intensive margin of trade.<sup>15</sup> As pointed out by Redding (2011) the intensive margin depends only on the fixed cost of trade and is independent of variable trade costs. To express bilateral trade as in Anderson and van Wincoop (2003), we can substitute from the zero cut-off productivity equation

$$(\varphi_{ij}^*)^{-\kappa} = \left( \frac{W_i t_{ij}}{\rho A_i P_j} \right)^{-\kappa} \left( \frac{EY_j}{W_i f_{ij}^X} \right)^{\frac{\kappa}{\sigma-1}}$$

---

<sup>15</sup>More accurately, this intensive margin can be referred to as an *average export margin* as discussed in Head and Mayer (2014). For a given fall in trade cost this margin will include the intensive margin of “originally” active firms and the exports from firms that begin exporting.

and after some rearranging we obtain

$$X_{ij} = \tilde{B}_M (W_j L_j)^{\frac{\kappa}{\sigma-1}} (W_i L_i) W_i^{\frac{-\kappa\sigma}{\sigma-1}} A_i^{-\kappa} \left( \frac{t_{ij}}{P_j} \right)^{-\kappa} (f_{ij}^X)^{\frac{-\kappa}{\sigma-1}-1},$$

where  $\tilde{B}_M = \bar{\varphi} \left( \frac{(\sigma-1)^{\kappa+1} \sigma^{\frac{\kappa\sigma}{\sigma-1}}}{\kappa-\sigma+1} \right)$ .

Market clearing implies

$$W_i L_i = \tilde{B}_M (W_i L_i) W_i^{\frac{-\kappa\sigma}{\sigma-1}} A_i^{-\kappa} \left[ \sum_j \left( \frac{t_{ij}}{P_j} \right)^{-\kappa} (f_{ij}^X)^{\frac{-\kappa}{\sigma-1}-1} (W_j L_j)^{\frac{\kappa}{\sigma-1}} \right]. \quad (35)$$

Define outward multilateral resistance as

$$\Pi_i = \left[ \sum_j \left( \frac{t_{ij}}{\check{P}_j} \right)^{-\kappa} (f_{ij}^X)^{\frac{-\kappa}{\sigma-1}+1} W_j L_j \right]^{\frac{-1}{\kappa}}.$$

where  $\check{P}_j = (W_j L_j)^{\frac{(\kappa-\sigma+1)}{(\sigma-1)\kappa}} P_j$ .

Bilateral trade can be expressed as

$$X_{ij} = \left( \frac{t_{ij}}{\Pi_i \check{P}_j} \right)^{-\kappa} (f_{ij}^X)^{\frac{-\kappa}{\sigma-1}+1} (W_i L_i) (W_j L_j).$$

Inward multilateral resistance ( $\check{P}_j$ ) is given by

$$\check{P}_j = \left[ \sum_i \left( \frac{t_{ij}}{\Pi_i} \right)^{-\kappa} (f_{ij}^X)^{\frac{-\kappa}{\sigma-1}+1} W_i L_i \right]^{\frac{-1}{\kappa}},$$

and the price index is given by

$$P_j = (W_j L_j)^{\frac{-(\kappa-\sigma+1)}{(\sigma-1)\kappa}} \check{P}_j.$$

Substituting the definition of the OMR into equation (35) and rearranging yields the wage equation

$$W_i = B_M \left( \frac{A_i}{\Pi_i} \right)^{\frac{\sigma-1}{\sigma}}$$

where  $B_M = (\tilde{B}_M)^{\frac{\sigma-1}{\kappa\sigma}}$ .

Collect the equations for trade flows, the multilateral resistance terms, and the wage equation

to obtain the structural gravity system from the Melitz model

$$X_{ij} = \left( \frac{t_{ij}}{\Pi_i \check{P}_j} \right)^{-\kappa} (f_{ij}^X)^{\frac{-\kappa}{\sigma-1}+1} (W_j L_j)(W_i L_i) \quad \forall i, j \quad (36)$$

$$\Pi_i = \left[ \sum_j \left( \frac{t_{ij}}{\Pi_i} \right)^{-\kappa} (f_{ij}^X)^{\frac{-\kappa}{\sigma-1}+1} W_j L_j \right]^{-\frac{1}{\kappa}} \quad \forall i \quad (37)$$

$$\check{P}_j = \left[ \sum_i \left( \frac{t_{ij}}{\Pi_i} \right)^{-\kappa} (f_{ij}^X)^{\frac{-\kappa}{\sigma-1}+1} W_i L_i \right]^{-\frac{1}{\kappa}} \quad \forall j \quad (38)$$

$$W_i = B_M \left( \frac{A_i}{\Pi_i} \right)^{\frac{\hat{\theta}}{\theta+1}} \quad \forall i. \quad (39)$$

Once again, the structural gravity system from the Melitz model looks very similar to the systems that we derived in the Armington, the Monopolistic Competition, and the Ricardian settings. As before, the structural gravity equation (36) reveals that bilateral trade depends on bilateral trade costs, the size of the countries, and the importer and exporter's multilateral resistance terms. The expressions for the multilateral resistance terms (37)-(38) as well as the wage equation (39) also carry the same intuition as before. The main difference here is that the price index and inward multilateral resistance are not identical. The price index is the inward multilateral resistance scaled by output; this arises because the fixed costs of exporting to other destinations are incurred by domestic workers.<sup>16</sup> Thus, once again, the differences between system (36)-(39) and the previous structural gravity systems are only in the interpretation of the model parameters, which now reflect the assumptions from the Melitz model.<sup>17</sup>

Specifically, the first difference is that in the Melitz model, both fixed and variable trade costs impact bilateral trade. Changes in the variable or fixed trade costs can impact the value of trade flows on the intensive and extensive margin of trade. That is, changes in trade costs can increase the volume shipped by those firms already exporting and can induce new firms to begin exporting. The variable trade cost elasticity is given by  $-\kappa$  and the elasticity of trade to fixed trade costs is given by  $-\left(\frac{\kappa-\sigma+1}{\sigma-1}\right) < 0$ . The second difference is in the wage equation. As before, equation (39) implies that wages are proportional to the level of national technology and inversely proportional to the outward multilateral resistance, but it differs from the corresponding wage equations from the previous models due to the definition of  $B_M$ .

<sup>16</sup>When the fixed cost of exporting is incurred in the consuming country, as in Arkolakis et al. (2008) and Redding (2011), the inward multilateral resistance is equal to the price index.

<sup>17</sup>The Melitz gravity system can be thought of as a block-recursive system where system 36-39 solves for trade flows, wages, and multilateral resistance. The price index,  $P_j = (W_j L_j)^{\frac{-(\kappa-\sigma+1)}{(\sigma-1)\kappa}} \check{P}_j \quad \forall j$ , can be solved for given the solution to system 36-39.



## 2.3 All Roads Lead to Rome: A Generalized Gravity System

In this section we offer a generalized gravity system, which we use to directly compare and summarize the implications of the four core models that we reviewed in the previous section. The generalized system will be used to motivate various aspects of our estimation and simulation analysis in subsequent sections. The following structural gravity representation nests each of the main theoretical foundations that were discussed so far

$$X_{ij} = \left( \frac{\tau_{ij}}{\Pi_i \check{P}_j} \right)^{-\epsilon} (W_i L_i)(W_j L_j) \quad (40)$$

$$\Pi_i = \left[ \sum_{j=1} \left( \frac{\tau_{ij}}{\check{P}_j} \right)^{-\epsilon} W_j L_j \right]^{\frac{-1}{\epsilon}} \quad (41)$$

$$\check{P}_j = \left[ \sum_{i=1} \left( \frac{\tau_{ij}}{\Pi_i} \right)^{-\epsilon} W_i L_i \right]^{\frac{-1}{\epsilon}} \quad (42)$$

$$W_i = B \left( \frac{A_i}{\Pi_i} \right)^{\frac{-\epsilon}{\epsilon+1}}, \quad (43)$$

where we broadly define  $\epsilon$  as a *trade cost elasticity*,  $I_M = 1$  for the Melitz model and 0 otherwise, and  $\tau_{ij} = t_{ij} (f_{ij}^X)^{I_M \frac{1-\sigma-\kappa}{\kappa(1-\sigma)}}$  as the total bilateral trade cost, which includes the variable trade cost component  $t_{ij}$  that appeared in all four frameworks presented so far, as well as the fixed export cost component,  $(f_{ij}^X)^{\frac{1-\sigma-\kappa}{\kappa(1-\sigma)}}$ , which was introduced in the Melitz framework. Inward multilateral resistance may not equal the price index in the Melitz model; as a result, we can compute the price index after we determine the solution to the above system from  $P_j = (W_j L_j)^{\frac{-(\kappa-\sigma+1)}{(\sigma-1)\kappa}} \check{P}_j$ .

Equation (40) is the structural gravity equation that governs bilateral trade flows. As discussed earlier, equation (40) relates bilateral trade to bilateral trade costs,  $t_{ij}$ , to multilateral trade costs,  $\Pi_i$  and  $\check{P}_j$ , and to country size, measured here by the exporter and importer incomes,  $W_i L_i$  and  $W_j L_j$ , respectively. There are two differences in the trade flows gravity equations across all models that we presented so far. The first difference is in the interpretation of the *trade cost elasticity*  $\epsilon$ . In the Armington model and the monopolistically competitive model  $\epsilon = \sigma - 1$ , where  $\sigma$  is the elasticity of substitution among varieties. In the Eaton-Kortum model ( $\epsilon = \theta$ ) and the Melitz model ( $\epsilon = \kappa$ ), the trade cost elasticity reflects the productivity dispersion across firms. The second difference between the gravity models is in the definition of the bilateral trade costs, which capture iceberg frictions in the Armington model, the monopolistically competitive model, and the Eaton-Kortum models, and are defined as a combination of iceberg and fixed trade costs in the Melitz model. Due to its intuitive appeal, flexible multiplicative structure, and remarkable predictive power, equation (40) has established itself as one of the most popular empirical models in international economics. Theory implies that, in principle, one can decompose the trade

elasticity into its variable cost,  $\sigma$ , and fixed cost,  $\kappa$ , components and identify those separately from specification (40). In the empirical section below, we argue that better knowledge and proper measurement of transportation costs may lead to new developments in this area.

In combination, the next two equations, (41) and (42), define and can be used to solve for the outward and the inward multilateral resistances, respectively. As discussed earlier, the MRs are intuitive general equilibrium trade cost indexes that decompose the incidence of trade costs on the consumers and the producers in each country as if they ship to and buy from a unified world market. In addition, the multilateral resistances provide a theoretical link between the trade flows equation and the equation for factor prices. Thus, the MRs can be viewed as structural vehicles that can translate any changes in bilateral transportation costs or trade policies into general equilibrium effects on various economic outcomes. Similar to the trade flows equation, the expressions for the multilateral resistances are identical across the four core models subject to the interpretation of the *trade cost elasticity* parameter  $\epsilon$  and subject to the definition of bilateral trade costs. It should be noted, however, that very often the power transforms of the multilateral resistances, which do not require a value for the trade elasticity of substitution, are sufficient for many empirical purposes, cf. Redding and Venables (2004) and Anderson et al. (2015a).

The last equation in gravity system (40)-(43) is the wage (or income per capita) equation. The beauty of this equation is that it captures the structural effects of a universe of bilateral trade policies and changes in transportation costs on income per capita in the countries that liberalize their trade policy or improve their transportation costs as well as in countries that have not been directly involved in a particular trade liberalization effort or improvement in bilateral transport channels. These links are captured by the general equilibrium OMR index. Note also that in combination with the corresponding changes in the general equilibrium IMR indexes, the changes in wages captured by equation (43) can be used to evaluate the change in real income per capita in response to trade policy changes or improvements in transportation. Finally, we remind the reader that equation (43) is a restatement of the goods market clearing condition. As such, this equation can be used to directly link changes in the outward multilateral resistance to changes in factory gate prices, cf. Anderson and van Wincoop (2003). Furthermore, together with the definitions of the multilateral resistances, the market clearing equation can be used to nest the gravity model of bilateral trade into significantly more complicated production structures that have been used by a series of papers to study the impact of trade and trade liberalization on income, wages, welfare, unemployment, labor reallocation, and capital accumulation.<sup>18</sup>

---

<sup>18</sup>Building on the the original gravity models of Eaton and Kortum (2002) and Anderson and van Wincoop (2003), more recent structural frameworks used to perform various counterfactual analysis include Anderson and Yotov (2016), Egger et al. (2011), Egger and Larch (2011), Ossa (2011), Fieler (2011), Costinot et al. (2012), Behrens et al. (2014), Eaton et al. (2013), Arkolakis et al. (2013), Felbermayr et al. (2015), Heid (2015), Caliendo and Parro (2015), Anderson and Yotov (2016), Heid and Larch

There are two differences between the wage equations across the four core theoretical foundations. The first difference is in the interpretation of the trade elasticity parameter  $\epsilon$ . We discussed this difference earlier. The second difference is in the interpretation of the scale parameter  $B$ . In the Armington model, as described earlier in this essay, this parameter is equal to  $L^{\frac{-1}{\sigma}}$  because increases in the population increase the supply of the country's good and this results in a lower price and reduces factor prices. If the  $\beta$ 's were not normalized, then the scale parameter  $B$  would have been a function of the country-specific share parameter and of the trade elasticity of substitution. In the monopolistic competition model,  $B_{MC} = (\sigma^{-1}(\sigma - 1)^{\frac{\sigma-1}{\sigma}})f_i^{\frac{-1}{\sigma}}$  reflects the presence of fixed cost of production. In the Eaton-Kortum model,  $B_{EK} = \exp(-\gamma/(1 + \theta))$  reflects the production-side distributional assumptions of this EK model. Finally, in the Melitz model,  $B_M = \left( \bar{\varphi} \frac{(\sigma-1)^{\kappa+1} \sigma^{\frac{\kappa\sigma}{\sigma-1}}}{\kappa-\sigma+1} \right)^{\frac{\sigma-1}{\kappa\sigma}}$  has a more complex structure, which reflects the presence of fixed exporting costs as well as the fact that not all firms in the economy are exporters. The scale parameters  $B$  may be useful for validation purposes, i.e. they can be estimated or calibrated and compared to corresponding moments from actual data or corresponding estimates from other studies. However, these parameters play no role in numerical exercises and simulations of the effects of changes in transportation costs and/or trade policies because of their exogenous nature. We demonstrate this next, where we follow Dekle et al. (2007, 2008) to express our generalized structural gravity system in changes.

Let  $\hat{Z} = Z'/Z$  be the gross change for any variable in our system. Also define country  $j$ 's share of imports from country  $i$  as  $\psi_{ij}^M = X_{ij}/Y_j$ , and country  $i$ 's share of exports to country  $j$  as  $\psi_{ij}^X = X_{ij}/Y_i$ . Then, the gravity trade model (40)-(43) can be expressed as

$$\hat{X}_{ij} = \left( \frac{\hat{\tau}_{ij}}{\hat{\Pi}_i \hat{P}_j} \right)^{-\epsilon} \hat{W}_i \hat{W}_j \quad (44)$$

$$\hat{\Pi}_i = \left[ \sum_{j=1} \left( \frac{\hat{\tau}_{ij}}{\hat{P}_j} \right)^{-\epsilon} \hat{W}_j \psi_{ij}^X \right]^{\frac{-1}{\epsilon}} \quad (45)$$

$$\hat{P}_j = \left[ \sum_{i=1} \left( \frac{\hat{\tau}_{ij}}{\hat{\Pi}_i} \right)^{-\epsilon} \hat{W}_i \psi_{ij}^M \right]^{\frac{-1}{\epsilon}} \quad (46)$$

$$\hat{W}_i = (\hat{A}_i / \hat{\Pi}_i)^{\frac{\epsilon}{\epsilon+1}}. \quad (47)$$

System (44)-(47) clearly demonstrates that, for a given value of the trade elasticity parameter  $\epsilon$ , the response of economic outcomes (e.g. trade, income, welfare, etc.) to trade policy changes does not depend on the scale parameter  $B$ , and it will be the same regardless of the specific theoretical foundation. One of the most widely used general equilibrium indexes in response to

---

(2016), Anderson et al. (2015a), and Eaton et al. (2016). We also refer the reader to the survey articles in Footnote 1.

changes in trade costs, which we also report in the numerical section below, is the change in real income:

$$\frac{\hat{W}_i}{\hat{P}_i}. \quad (48)$$

Arkolakis et al. (2012) generalize this result by demonstrating that for a wide class of trade models the following statistic is sufficient to capture the impact of trade on national welfare:

$$\hat{\lambda}_{ii}^{-\frac{1}{\epsilon}}. \quad (49)$$

Here,  $\hat{\lambda}_{ii}$  is the change in the share of expenditure on home goods,  $\lambda_{ii} = X_{ii}/E_i$ , and  $\epsilon$  is the trade elasticity. Arkolakis et al. (2012) demonstrate that welfare expression (49) is valid for all trade models that share the following four main underlying assumptions: (i) Dixit-Stiglitz preferences; (ii) One factor of production; (iii) Linear cost functions; and (iv) Perfect or monopolistic competition.<sup>19</sup> It is straightforward to demonstrate that equations (48) and (49) are equivalent for the four core models that we presented here.<sup>20</sup> Armed with the powerful theoretical predictions of the structural gravity model, we are now ready to proceed with the estimation and simulation empirical analysis. However, before we do so, we offer three important theoretical extensions with implications for the impact and significance of transportation costs.

## 2.4 Extensions: Tariffs, Intermediates, and Sectors

This section presents three extensions of the gravity model that make it more relevant and realistic for empirical analysis. Specifically, we review extensions of the structural gravity model (i) to introduce tariffs; (ii) to account for trade in intermediates; and (iii) to accommodate multiple sectors. For expositional simplicity, we use the Armington framework to model these extensions, however they can be readily incorporated into the other theoretical gravity models.

### Structural Gravity with Tariffs and Rents

The introduction of tariffs in the structural gravity model is important for at least two reasons.<sup>21</sup> First, all trade costs modeled so far have been of the iceberg type; that is, a portion of the good dissolves and is lost in transport. However, in some instances the trade costs are the result of trade policies that involve rents which through some mechanism get (re)allocated back to agents (or groups) in the model. Second, tariffs represent a special type of trade cost because tariffs are

---

<sup>19</sup>In addition, the following macro-level conditions must be satisfied: (i) Balanced trade; (ii) Aggregate profits are a constant share of aggregate revenues; and (iii) The import demand system is CES.

<sup>20</sup>Melitz and Redding (2015) discuss the implications for welfare in a more general case, where the trade elasticity is no longer constant but instead variable.

<sup>21</sup>We refer the reader to the following papers that have studied the impact of tariffs with the structural gravity framework: Caliendo and Parro (2015) and Caliendo et al. (2015).

a direct price shifter. As we demonstrate below, the implication is that tariffs can be used in an empirical gravity model to directly identify an estimate of the trade elasticity of substitution, which is the most important parameter for welfare analysis of trade, cf. Arkolakis et al. (2012). This analysis is important in the current context because some transportation costs may also be viewed as direct price shifters. To see how rents can be incorporated into the standard trade model, consider the Armington model and assume that *ad valorem* tariffs are levied on imports and that all tariff revenues collected by the government are distributed as lump-sum transfers back to the households. The expression for bilateral trade flows is given by

$$X_{ij} = \tau_{ij}^{-\sigma} \left( \frac{p_i t_{ij}}{P_j} \right)^{1-\sigma} E_j,$$

where  $\tau_{ij} = (1 + \textit{advalorem tariff}) \geq 1$  is the tariff rate, and  $E_j$  are expenditures by country  $j$ . Note that in the presence of tariff revenues, the value of domestic production and expenditure are no longer the same. With labor as the only factor of production, expenditures by households in country  $j$  can be expressed as

$$E_j = W_j L_j + \left\{ \sum_j \left( \frac{\tau_{ij} - 1}{\tau_{ij}} \right) \left( \frac{p_i \tau_{ij} t_{ij}}{P_j} \right)^{1-\sigma} E_j \right\}.$$

This implies that

$$E_j = W_j L_j (\Xi), \tag{50}$$

where

$$\Xi = \frac{\sum_j \left( p_i \tau_{ij} t_{ij} \right)^{1-\sigma}}{\sum_j \left( \frac{1}{\tau_{ij}} \right) \left( p_i \tau_{ij} t_{ij} \right)^{1-\sigma}} \geq 1$$

is the tariff multiplier. Noting that  $P^{1-\sigma} = \sum_j \left( p_i \tau_{ij} t_{ij} \right)^{1-\sigma}$  and substituting this expression into the reduced form gravity equation yields

$$X_{ij} = \tau_{ij}^{-\sigma} \left( \frac{p_i t_{ij}}{\tilde{P}_j} \right)^{1-\sigma} W_j L_j,$$

where

$$\tilde{P}_j = \left[ \sum_j \tau_{ij}^{-\sigma} \left( p_i t_{ij} \right)^{1-\sigma} \right]^{\frac{1}{1-\sigma}}.$$

Market clearing implies that the structural gravity equation with tariffs is given by

$$X_{ij} = \tau_{ij}^{-\sigma} \left( \frac{t_{ij}}{\tilde{\Pi}_i \tilde{P}_j} \right)^{1-\sigma} (W_i L_i) (W_j L_j), \tag{51}$$

where

$$\tilde{\Pi}_i = \left[ \sum_j \tau_{ij}^{-\sigma} \left( \frac{t_{ij}}{\tilde{P}_j} \right)^{1-\sigma} W_j L_j \right]^{\frac{1}{1-\sigma}}, \quad (52)$$

and

$$\tilde{P}_j = \left[ \sum_i \left( \frac{t_{ij} \tau_{ij}}{\tilde{\Pi}_i} \right)^{1-\sigma} W_j L_j \right]^{\frac{1}{1-\sigma}}. \quad (53)$$

Finally, it can be shown that

$$W_i = \left( \frac{A_i}{\tilde{\Pi}_i} \right)^{\frac{\sigma-1}{\sigma}}. \quad (54)$$

## Structural Gravity with Intermediate Inputs

The economic geography literature investigated how forward and backward-linkages impact production and the returns to factors.<sup>22</sup> To see how trade costs can impact these linkages, we consider a one sector Armington model where production is given by

$$Q_i = \left( \frac{l_i(\omega)}{A_i} \right)^\nu \left( \left[ \sum_j (q_{ij}^m)^{\frac{\sigma-1}{\sigma}} \right]^{\frac{\sigma}{\sigma-1}} \right)^{1-\nu}.$$

Total demand for goods produced in country  $i$  is equal to the final demand by consumers ( $X_{ij}^C$ ) and intermediate demand by producers ( $X_{ij}^M$ ); where final demand by consumers is given by

$$X_{ij}^C = \left( \frac{p_i t_{ij}}{P_j} \right)^{1-\sigma} W_j L_j,$$

and intermediate demand is given by

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

Noting that value added is a constant share of output ( $W_j L_j = \nu Y_j$ ), the demand for intermediates can be expressed as

$$X_{ij}^M = \frac{(1 - \nu)}{\nu} \left( \frac{p_i t_{ij}}{P_j} \right)^{1-\sigma} W_j L_j,$$

---

<sup>22</sup>International production fragmentation varies across sectors, cf. Johnson and Noguera (2012), Kraemer and Dedrick (2002), Varian (2007), and Grossman and Rossi-Hansberg (2012). However, intermediate inputs represent more than half of the goods imported by the developed economies and close to three-quarters of the imports of some large developing countries, cf. Ali and Dadush (2011). Recent quantitative trade models that have recognized the importance of intermediate inputs and have incorporated them within the structural gravity model include Eaton and Kortum (2002), Costinot et al. (2012), Caliendo and Parro (2015), and Anderson et al. (2015a).

and total demand for goods from country  $i$  is given by

$$X_{ij} = X_{ij}^C + X_{ij}^M,$$

or the reduced form gravity equation is given by

$$X_{ij} = \left( \frac{p_i t_{ij}}{P_j} \right)^{1-\sigma} W_j L_j,$$

and, under the current assumptions, the *structural gravity equation* and the corresponding expressions for the *multilateral resistance terms* are identical to the structural gravity without intermediates; that is,

$$\begin{aligned} X_{ij} &= \left( \frac{t_{ij}}{\Pi_i P_j} \right)^{1-\sigma} (W_j L_j) (W_i L_i) \\ \Pi_i &= \left[ \sum_j \left( \frac{t_{ij}}{P_j} \right)^{1-\sigma} W_j L_j \right]^{\frac{1}{1-\sigma}} \\ P_j &= \left[ \sum_i \left( \frac{t_{ij}}{\Pi_i} \right)^{1-\sigma} W_i L_i \right]^{\frac{1}{1-\sigma}}. \end{aligned}$$

Thus, all implications of the structural gravity model continue to hold with regard to equations for bilateral trade flows and multilateral resistances. However, intermediates play an important role in these models for the determination of factor prices and, therefore, welfare. To see this, note that given the firm's technology, marginal costs and the factory gate price are given by

$$p_i = \left( \frac{W_i}{A_i} \right)^\nu \left( P_i \right)^{1-\nu},$$

so that market clearing implies

$$W_i L_i = \left( \frac{W_i}{A_i} \right)^{-\nu(\sigma-1)} (P_i)^{-(1-\nu)(\sigma-1)} (\Pi_i)^{-(\sigma-1)},$$

or that wages are given by

$$W_i = \left( \frac{A_i^\nu L_i^{\frac{1}{1-\sigma}}}{\Pi_i P_j^{1-\nu}} \right)^{\frac{\sigma-1}{1+\nu(\sigma-1)}}. \quad (55)$$

Equation (55) differs in two (intuitive) dimensions from the original gravity model. First, we note that technology enters the wage equation as a different power transform. This reflects the fact that, in addition to labor and technology, production now also employs intermediates with a share  $1 - \nu$ . Second, note that, in addition to the outward multilateral resistance, the inward multilateral resistance also appears in the wage equation. The relationship between the IMR and

wages is also inverse. The intuition for this result is that, in addition to bearing the interpretation of being a consumer price index, the inward multilateral resistance in this setting also reflects the cost of intermediates for production and, all else equal, a fall in the price of intermediates will result in higher returns to labor. Motivated by the links captured by equation (55), Redding and Venables (2004) used the standard gravity specification to construct the terms for market access and supplier access, which correspond to the multilateral resistances, and found empirical support for the links between the market access and supplier access indexes and per-capita income.

## Structural Gravity with Sectors

Transportation costs vary widely across products. Similarly, most trade policies are determined and implemented at the sectoral level. Furthermore, trade liberalization may have a quite heterogeneous impact at the sectoral level that often leads to structural changes that may be masked by aggregation in the single-sector models that we presented and discussed so far. Motivated by these arguments, in this section we follow the analysis from Anderson and Yotov (2016) to extend the Armington gravity model of Anderson and van Wincoop (2003) to the sectoral level.<sup>23</sup> Two assumptions, in addition to those made in Anderson and van Wincoop (2003), deliver a sectoral gravity system. First, on the supply side, assume an endowment economy, but this time at the sectoral level, so that:

$$Y_i^k = p_i^k L_i^k A_i^k, \quad (56)$$

where  $L_i^k$  is the amount of labor employed in sector  $k$  in country  $i$ , and  $A_i^k$  is a measure of the labor efficiency in sector  $k$  in country  $i$ . On the demand side, assume that consumer preferences can be a nested utility function where the upper-level is Cobb-Douglas across sectors and the lower level is a CES utility function. Assuming that total expenditures equal total income, the Cobb-Douglas preferences for goods across sectors imply that expenditure on goods from each sector would be a constant share of this country's total income:

$$E_i^k = \alpha^k Y_i = \alpha^k \sum_k Y_i^k. \quad (57)$$

---

<sup>23</sup>Recent papers that develop and quantify sectoral gravity models include Costinot et al. (2012); Larch and Wanner (2014); Caliendo and Parro (2015); and Donaldson (2016). Larch and Yotov (2016) demonstrate that, subject to interpretation of the parameters, the sectoral gravity models that are obtained on the demand side, i.e. based on Anderson and van Wincoop (2003), and those obtained on the supply side are identical.



Following the same steps as in the derivation of the single-sector Armington model, the structure delivers the following structural sectoral gravity system:

$$X_{ij}^k = \frac{Y_i^k E_j^k}{Y^k} \left( \frac{t_{ij}^k}{\Pi_i^k P_j^k} \right)^{1-\sigma_k}, \quad (58)$$

$$(\Pi_i^k)^{1-\sigma_k} = \sum_j \left( \frac{t_{ij}^k}{P_j^k} \right)^{1-\sigma_k} \frac{E_j^k}{Y^k}, \quad (59)$$

$$(P_j^k)^{1-\sigma_k} = \sum_i \left( \frac{t_{ij}^k}{\Pi_i^k} \right)^{1-\sigma_k} \frac{Y_i^k}{Y^k}, \quad (60)$$

$$W_i^k = \left( \frac{A_i^k}{\Pi_i^k} \right)^{\frac{\sigma_k-1}{\sigma_k}}, \quad (61)$$

$$E_i^k = \alpha_k \sum_k Y_i^k \quad (62)$$

$$L_i = \sum_k L_i^k. \quad (63)$$

Note that, apart from the definitions of sectoral expenditure and total labor from equations (62) and (63), system (58)-(61) is very similar to the aggregate Armington gravity system (17)-(20). At first glance, it looks like the appearance of the sectoral superscript  $k$  is the only difference between the two systems. Indeed, most of the implications that we discussed at the aggregate level continue to hold at the sectoral level. Importantly, one nice feature of the new sectoral gravity system is that it implies that the trade flows equation (58) can be estimated exactly the same way separately for each sector as it can be estimated at the aggregate level. Thus, the discussion and all recommendations for gravity estimations that we will make in the next section apply to the aggregate and to the sectoral level. If labor is immobile across sectors, changes in trade costs will only impact prices and there will be no impact on the quantities of goods produced. With mobile labor, changes in trade costs will likely impact prices and quantities. There is also another important implication of the introduction of the sectoral dimension. Specifically, system (58)-(61) captures inter-sectoral linkages, which arise on the demand side due to the fact that consumers not only substitute across different varieties but also substitute across different sectors too. More recently, models with multiple sectors and an input-output structure for intermediate inputs have been extended to also allow for cross-sectoral linkages on the supply side in order to highlight and better capture the impact of changes in trade costs on welfare, e.g. Caliendo and Parro (2015), and on allocations of heterogeneous labor, e.g. Baier et al. (2016). A common feature of these, and other, more complex models is that the industry-level trade flows equation (58) emerges in each of them, and estimation is straightforward. We capitalize on this property in the next section, where we discuss estimations of the structural gravity model.

### 3 Estimating Structural Gravity

The gravity equation has been the empirical workhorse model of international trade over the past 50 years. Empirical applications of the gravity model precede the theoretical developments in the gravity literature.<sup>24</sup> Furthermore, owing to the remarkable predictive success of the gravity model in combination with its intuitive appeal, most of the gravity literature is empirical. The objective of most empirical papers is to quantify the impact of various factors that contribute to higher or lower trade costs.<sup>25</sup> The goal of this section is to review the evolution of estimating the *structural* gravity equation, and introduce the reader to the major challenges and solutions to gravity estimations. We refer the reader to Baldwin and Taglioni (2006), Head and Mayer (2014), and Piermartini and Yotov (2016) for recent complementary studies and reviews of the empirical gravity literature.<sup>26</sup>

---

<sup>24</sup>As noted earlier, Ravenstein (1885) was the first to use a gravity-type relationship in order to study immigration patterns. Tinbergen (1962) was the first to use a gravity equation to study the impact of trade policy. Given the wide gap between these two studies and because the focus of Ravenstein's work is migration, we will use Tinbergen's work as the departing point for our empirical analysis.

<sup>25</sup>It is important to note that estimation is not the only method to identify trade costs. Some authors have used other approaches to measure trade costs by using combinations of ratios, bilateral trade flows, and production data. Notable contributions associated with this approach include Head and Ries (2001), Romalis (2007), Jacks et al. (2011), Novy (2013a), Caliendo and Parro (2015), and Anderson et al. (2015b). Since, by design, the calibration approach of constructing trade costs should fit the data perfectly, it may be able to account better for time varying trade costs. There are two potential drawbacks from this approach. First, this approach may conflate transitory changes in bilateral trade flows resulting from unobserved demand and supply side changes to transitory changes in trade costs. The calibration exercise may attribute the increased trade to a temporary reduction in trade costs. Second, the calibrated trade costs do not allow the researcher to identify the factors responsible for the observed trade frictions. Anderson et al. (2015b) show how it is possible to combine the two approaches to compute the general equilibrium effects of changes in trade costs. Estimation is not without its drawbacks either. In most instances, the variables used to proxy for bilateral trade costs are time invariant. Therefore, while the gravity equation can account for cross-sectional variation in bilateral trade flows, the gravity estimates have more difficulty in accounting for the *growth* in bilateral trade over time, c.f. Lai and Treffer (2002). An open challenge to trade economists and transport economists is to better understand how trade costs are changing over time and how the changes in trade costs have influenced bilateral trade.

<sup>26</sup>Head and Mayer (2014) and Yotov et al. (2016) also offer computer codes for empirical gravity analysis. Some authors have calibrated trade costs by using combinations of ratios of bilateral trade flows and production data in order to isolate the effects that can be plausibly attributed to trade frictions. See for example Head and Ries (2001), Romalis (2007), Novy (2013b), Caliendo and Parro (2015), and Anderson et al. (2015b). Yotov et al. (2016) review the leading methods to calibrate trade costs. By construction, the calibrated measures match the data perfectly. While the calibration approach delivers (subject to measurement error) the total trade cost bill, one shortfall of this approach is that it is harder to identify whether and how specific factors contribute to overall trade costs. Given that the latter has been the primary objective of all empirical gravity studies, we devote the rest of this section exclusively to a discussion of gravity estimations.

### 3.1 Structural Gravity Specification

In this section, we translate the theoretical gravity model into an estimating equation and we discuss some of the challenges associated with estimating the structural gravity equation (*SGE*) derived in the previous section. Throughout the analysis we assume that the researcher has access to panel data; that is bilateral trade flows covering multiple years.<sup>27</sup> In this case, we introduce time subscripts to the empirical gravity equation to obtain:<sup>28</sup>

$$X_{ijt}^{SG} = \left( \frac{t_{ijt}}{\prod_{it} P_{jt}} \right)^{-\epsilon} (f_{ij}^X)^\eta Y_{it} E_{jt}, \quad (64)$$

where we assume that the trade elasticity of variable trade costs is given by  $\epsilon$  and the trade elasticity of fixed costs of exporting is given by  $\eta$ . The difference between actual bilateral trade and the structural gravity equation can be expressed as

$$X_{ijt} = X_{ijt}^{SG} + u_{ijt},$$

where  $u_{ijt}$  is the error term. Equation (64) can be translated into an econometric model in two steps. The first issue the researcher must address is the nature of the trade cost function.<sup>29</sup> A common specification for the bilateral trade cost function is

$$t_{ijt}^{-\epsilon} (f_{ij}^X)^\eta = \exp\left(\mathbf{z}_{ijt}^N \beta^N + \mathbf{z}_{ijt}^H \beta^H + \mathbf{w}_{ijt}^P \psi^P\right), \quad (65)$$

where  $\mathbf{z}_{ijt}^N$  represent natural trade barriers (e.g. distance and contiguity),  $\mathbf{z}_{ijt}^H$  capture historical and cultural linkages that may raise or lower trade costs (e.g. common language, colonial ties, and common legal system), and  $\mathbf{w}_{ijt}^P$  represent (a vector of) trade costs that are potentially endogenous.<sup>30</sup> It is not uncommon that many of the right-hand side covariates are time invariant;

---

<sup>27</sup>As discussed in Piermartini and Yotov (2016), panel data are now widely available to researchers and there are several benefits from the panel dimension in gravity estimations. “First, panel data will lead to improved estimation efficiency; In addition, the panel dimension would enable researchers to apply the pair-fixed-effects methods of Baier and Bergstrand (2007) to address the issue of endogeneity of trade agreements; Third, on a related note, the use of panel data would allow for a flexible and comprehensive treatment and estimation of the effects of time-invariant bilateral trade costs with pair fixed effects.” (p.13, Piermartini and Yotov, 2016). We refer the reader to Head and Mayer (2014), who provide a thorough discussion of empirical issues that may arise with cross sectional data.

<sup>28</sup>Equation (64) naturally arises from a dynamic setting with factor accumulation. See for example Olivero and Yotov (2012), Eaton et al. (2016), and Anderson et al. (2015a)

<sup>29</sup>In some models, fixed exporting costs enter into the *structural gravity equation*. Ideally, we would like to be able to distinguish between fixed and variable trade costs; however, many of the standard gravity covariates could plausibly be related to fixed or variable costs. We refer the reader to Anderson and van Wincoop (2004) for a thorough early survey on bilateral trade costs.

<sup>30</sup>In most applied applications, these types of trade costs are associated with the outcome of a political process (e.g. bilateral tariffs, free trade agreements (FTAs), and WTO membership).

as a result, the gravity specification does a *better job* of explaining cross-sectional differences in bilateral trade than it does in explaining the growth of bilateral trade over time. In particular, the gravity model has a difficult time accounting for the fact that trade has grown faster than income for much of the Post WWII era, cf. Baier and Bergstrand (2001). An open challenge to trade and transport economists is to identify the factors that have contributed to the growth of world trade above and beyond income growth.

Once the trade cost function is specified, researchers need to determine how they will account for the multilateral resistance terms. Recall that in the *naive* gravity specification researchers would construct variables such as the remoteness index, the proxy for the price indexes using GDP or CPI price index. These variables only loosely correspond to the price index implied by the theory. If these proxies are poorly measured and/or omitted, it is likely that the right-hand side will be correlated with the error term, resulting in inconsistent coefficient estimates. Perhaps the most well known example of this is the seminal paper by Anderson and van Wincoop (2003) that showed how the gravity model once correctly specified could help resolve the *border puzzle* in McCallum (1995) where he estimated that the implied border barrier between the United States and Canada, two countries that were seemingly quite integrated, caused internal trade within provinces in Canada to be more than 20 times higher than trade between U.S. states and Canadian provinces. Anderson and van Wincoop (2003) showed that omitting the multilateral resistance terms from the specification may result in biased coefficient estimates. Anderson and van Wincoop employed non-linear least squares to estimate the system of equations given by equations (22)-(24). Since that time, almost every empirical paper that accounts for the multilateral resistance terms uses importer and exporter fixed effects in cross-section settings and importer-time and exporter-time fixed effects with panel data.<sup>31</sup>

Substituting the definition of bilateral trade costs from equation (65) into the structural gravity equation (64), and using exporter-time and importer-time fixed effects to control for the multilateral resistances delivers an estimating structural gravity model:

$$X_{ijt}^{SG} = \exp\left(\mathbf{z}_{ijt}^N \beta^N + \mathbf{z}_{ijt}^H \beta^H + \mathbf{w}_{ijt}^P \psi^P + \delta_{it}^X + \delta_{jt}^M\right). \quad (66)$$

Note that, in addition to controlling for the unobservable multilateral resistance terms, the

---

<sup>31</sup>In their original paper, Anderson and van Wincoop (2003) used custom programming to account for the MRs. Hummels (2001) and Feenstra (2004) proposed the use of exporter and importer fixed effects to account for the MRs cross-section settings. Olivero and Yotov (2012) demonstrated that the MRs should be accounted for by exporter-time and importer-time fixed effects. We note that in a panel setting with multiple sectors, the appropriate (theoretically-consistent) set of fixed effects in the structural gravity equation should include importer-sector-time and exporter-sector-time dummies. At one time, obtaining estimates from empirical models with many countries and time periods was quite time consuming because of the number of dummy variables included in a regression. However, there are now many software packages and user-written programs that allow for high dimensional fixed effects.

exporter-time and the importer-time fixed effects will also absorb the exporter and the importer size variables, respectively. Specifically, the theoretical interpretation of the fixed effects dummies is  $\delta_{it}^X = \left( \ln Y_{it} + (\sigma - 1) \ln \Pi_{it} \right)$  and  $\delta_{jt}^M = \left( \ln E_{jt} + (\sigma - 1) \ln P_{jt} \right)$ . Furthermore, from a purely econometric perspective, the structural fixed effects will also absorb any other time-varying characteristics that may influence bilateral trade flows on the importer and the exporter side. For example, the fixed effects would absorb and prevent identification of the impact of improvements in domestic infrastructure on international trade. We view this as an open challenge to trade and transport economists.

Equation (66) represents a comprehensive version of the structural estimating gravity equation. Versions of it have been used in hundreds of applications. However, in order to obtain sound gravity estimates, researchers must still consider the type of estimator they will employ and how they will address potential endogeneity issues. The following subsections address these issues.

## 3.2 Structural Gravity Estimation

This section presents and discusses the advantages of the latest econometric methods to estimate structural gravity. We begin describing the conditions necessary for the log-linear structural gravity specification to yield consistent estimates. Then, we gradually build on those estimates to address a series of data and estimation challenges with the structural gravity model.

### Traditional, Log-Linear Gravity Models

Following the evolution of the empirical gravity literature, we start our estimation analysis by considering a cross-section of data with  $N$  countries and  $T$  periods so that there are potentially  $TN(N - 1)$  bilateral observations.<sup>32</sup> If the assumptions below are satisfied, the log-linear gravity model will yield consistent estimates. In this case, properly conditioned cross-sectional estimation or pooled OLS will result in consistent estimates. To conserve on notation, we focus on cross-sectional estimation. The initial focus on cross-sectional estimation simplifies notation by dropping time subscripts. We also assume that there are some observations where trade flows are not observed because no bilateral trade took place or because bilateral trade was not recorded or reported. Accordingly, we define  $s_{ij}$  as an indicator variable that equals unity when  $X_{ij} > 0$  and zero otherwise.<sup>33</sup> Finally, for simplicity, we express the gravity equation as

$$X_{ij} = \exp\left(Z_{ij}\beta\right) + u_{ij}$$

---

<sup>32</sup>Initially, we assume that the researcher does not have access to internal trade flows; that is, the researcher does not have a measure of domestic consumption of goods produced within the country.

<sup>33</sup>See Wooldridge (2009a), Wooldridge (2009b), and Wooldridge (2010) for a more general treatment of selection and unbalanced panels.

where  $\exp\left(Z_{ij}\beta\right) = \exp\left(\mathbf{z}_{ij}^N\beta^N + \mathbf{z}_{ij}^H\beta^H + \mathbf{w}_{ij}^P\psi^P + \delta_i^X + \delta_j^M\right)$ .

Below we provide four conditions that are needed to ensure that the parameters of interest are consistently estimated.<sup>34</sup>

**SGE OLS 1** :  $u_{ij} = X_{ij}^{SG}(\tilde{v}_{ij})$

**SGE OLS 2** :  $E(v_{ij} | Z_{ij}) = 0$  where  $v_{ij} = \ln(1 + \tilde{v}_{ij})$

**SGE OLS 3** :  $E(s_{ij} | Z_{ij}, u_{ij}) = E(s_{ij})$

**SGE OLS 4** :  $E(v_{ij}^2 | Z_{ij}) = \sigma_v^2$

- Assumption **SGE OLS 1** implies that the structural gravity equation can be written in multiplicative form; that is,

$$X_{ij} = X_{ij}^{SG}(1 + \tilde{v}_{ij}),$$

or log-linear as

$$\ln(X_{ij}) = \mathbf{z}_{ij}^N\beta^N + \mathbf{z}_{ij}^H\beta^H + \mathbf{w}_{ij}^P\beta^P + \delta_i^X + \delta_j^M + v_{ij}. \quad (67)$$

Equation (67) is the most popular and widely used version of the empirical gravity model. However, as demonstrated below, a straightforward application of this simple log-linear specification may lead to biased and inconsistent gravity estimates.

- Assumption **SGE OLS 2** implies that all right-hand side variables are *exogenous*. This assumption is likely to be violated when the vectors of covariates include trade policy measures. For example, Baier and Bergstrand (2007) argue that if two countries have an extensive amount of domestic regulations that limit trade, the likelihood of a free trade agreement may increase because the trade agreement may provide some harmonization of the regulations. Therefore, trade agreements could be positively correlated with (unobserved) domestic regulations, and if bilateral trade is negatively correlated with the domestic regulations, estimates of the impact of free trade agreements may be *downwardly biased*. Reverse causality may also be present in the case of regional trade agreements, where trading partners that already trade a lot are, all else equal, more likely to sign RTAs. In this case, the estimated impact of free trade agreements may be *upward biased*. Similar intuition and reverse causality concerns would apply to transportation channels, which are built to facilitate trade but are often in response to already strong and demanding existing trade relationship. The issue of endogeneity of trade policy is an important concern that is not new to the trade literature and has gained significant attention already. We will discuss the endogeneity challenges and possible solutions in more detail below.

---

<sup>34</sup>In addition to these conditions, we assume that the standard rank conditions apply.

- Assumption **SGE OLS 3** allows bilateral trade flows to take on zero values, but the realization of zero trade flows are unrelated to the control variables or the error term. This assumes that a reported or recorded zero bilateral trade flow occurs randomly. In the statistical literature this assumption is referred to as *missing completely at random*. Violation of **SGE OLS 3** is one of the reasons an increasing number of papers have explored other estimators to estimate the empirical gravity equation. The *missing completely at random* assumption is likely violated in many practical applications, especially when the structural gravity model is used with disaggregated data and more so in the case of services trade, where consumption is highly localized. However, consistent estimation can be obtained if the following condition holds:

$$\text{SGE OLS 3}' : \mathbf{E}(v_{ij} | \mathbf{Z}_{ij}, s_{ij}) = \mathbf{E}(v_{ij} | \mathbf{Z}_{ij}).$$

This condition implies that selection is a function of the covariates and is uncorrelated with the error term.<sup>35</sup>

- Assumption **SGE OLS 4** implies that that error terms are homoskedastic. In many applications, using heteroskedasticity-robust standard errors can increase efficiency with little additional computational cost. However, heteroskedasticity that is a function of the covariates can result in biased and even inconsistent estimates of the (semi-)elasticities that the researcher is interested in; we discuss alternative estimation strategies below.

Given the above assumptions, cross sectional estimates yield consistent estimates of the trade costs parameters when importer and exporter fixed effects are included in the specification. For pooled OLS, consistent parameter estimates are obtained when importer-year and exporter-year fixed effects are included in the empirical specification.<sup>36</sup> Table 1 reports the results from the traditional cross sectional, log-linear, structural gravity model that are obtained with the OLS estimator. The model is estimated using non-zero nominal manufacturing data on bilateral international trade flows for 69 countries over the period 1986-2006, and we present the results using multiple years over 4-year intervals, 1986, 1990, 1994, 1998, 2002, and 2006.<sup>37</sup> Following the earlier discussion, the estimates for each year are obtained with *exporter* and *importer* fixed effects, which will control for the multilateral resistances and also absorb any other country-

---

<sup>35</sup>This implies that selection can be written as a deterministic function of the right-hand side controls; that is,  $s_{ij} = S(Z_{ij})$  or that, after controlling for the right-hand side controls, selection is independent of  $v_{ij}$ .

<sup>36</sup>Baldwin and Taglioni (2006) emphasize these points.

<sup>37</sup>For consistency and comparison purposes, throughout the analysis we will rely on the dataset from Baier et al. (2016). Two advantages of these data include a panel dimension and consistently constructed internal trade flows. We refer the reader to Baier et al. (2016) for a description of the data construction and sources. Yotov et al. (2016) offer a detailed discussion of various databases that have been used for gravity estimations.

specific factors that may affect bilateral trade flows on the importer and exporter side. The final column of Table 1 includes pooled cross-sectional estimates of the trade costs parameters.

The estimates from Table 1 are in accordance with our expectations and with corresponding numbers from the literature. Most variables have the expected signs and are significantly different from zero. Bilateral distance has a strong negative impact on bilateral trade; all else equal, bilateral pairs that are farther apart tend to trade less. We also note that the estimates are relatively stable over time. Several studies have documented the surprising result that the (absolute value of the) distance elasticity has remained stable over time or has increased; see for example Disdier and Head (2008) and Combes et al. (2008). The coefficient estimates on distance in gravity estimates are expected to decrease over time due to improvements in communications and technology. The stable (or increasing) measured distance elasticities is frequently referred to as the *distance puzzle*. A possible explanation for the distance puzzle could be that the distance variable in gravity regressions stands as a proxy for many determinants in international trade and as such it may combine forces that neutralize the impact of the improvements in communication and technology. We would expect that a proper measurement of transportation costs, which directly reflects improvements in communication and technology, may lead to a resolution of the “distance puzzle in international trade.”<sup>38</sup>

In most years, bilateral trade tends to be higher for countries that share a common border; however, the effect of the contiguity is imprecisely measured. Countries that have the same common (official) language (*LANG*) tend to trade more and those countries that were in a colonial relationship (*CLNY*) also tend to trade more. The estimated impact of free trade agreements (*FTA*) varies across years. In some years, the estimated impact of trade agreements is imprecisely measured, in other years the effect is negative and significant, and in other years the *FTAs* are shown to boost trade. It is tempting to attribute the fact that the impact of *FTAs* in Table 1 has increased from negative and statistically significant to positive and statistically significant to the trend that *FTAs* have become more comprehensive and deeper over time. However, we believe that a more plausible explanation for the unstable *FTA* estimates is that the *FTA* variable is endogenous.<sup>39</sup> We address the endogeneity issue in the next section. Before that, we note that the results presented in this table are not unique to this data set. Head and Mayer (2014) present results from a wide range of studies that highlight the negative relationship between bilateral trade and distance, and the positive impacts of economic mass, linguistic and

---

<sup>38</sup>The “distance puzzle” has attracted significant attention in the trade literature. Coe et al. (2002) generalize this result to define the “missing globalization puzzle” to reflect the fact that the total estimates of trade costs in gravity equations are stable over time. We refer the reader to Buch et al. (2004), Carrère and Schiff (2005), Brun et al. (2005), Boulhol and de Serres (2010), Yotov (2012), Lin and Sim (2012), Larch et al. (2016), Bergstrand et al. (2015), and Borchert and Yotov (2016) for studies that have attempted to explain the distance and missing globalization puzzles.

<sup>39</sup>See Treffer (1993), Magee (2003), Baier and Bergstrand (2002, 2004, 2007), and Egger et al. (2011) for relevant discussions.



colonial ties, trade agreements, and currency unions on trade flows.<sup>40</sup>

## Log-Linear Structural Gravity Estimation with Endogenous Variables

The estimates from Table 1 reveal that, even after controlling for exporter and importer effects, the effect of a free trade agreement is negative in some years, insignificant in other years, and positive and significant in others. As noted in Baier and Bergstrand (2007), the coefficient on FTA is the least stable of the coefficient estimates in the standard cross-sectional gravity specification. Baier and Bergstrand (2007) also noted that at the time of their paper no consensus had been reached on the impact of trade agreements on bilateral trade flows. For example, Tinbergen (1962) found a weak statistical relationship between preferential agreements and bilateral trade; Aitken (1973), Abrams (1980), Brada and Mendez (1985) found an economically and statistically significant positive effect on bilateral trade flows from membership in the European Community (EC); Bergstrand (1985), Frankel et al. (1995), and Frankel (1997) find negative and significant effects from EC membership.

Baier and Bergstrand (2007) argued that it seems unlikely that many of the covariates that measure trade policies are exogenous; that is, the tariffs and the decision to enter into trade agreements are unlikely to be randomly distributed across country pairs even after conditioning on other bilateral and unilateral controls. In their paper, they addressed several ways to handle violations of **SGE OLS 2**. If the policy variables are correlated with the error term, estimation can still be achieved by using standard *instrumental variable* (IV) techniques.<sup>41</sup> Baier and Bergstrand (2007) argue that finding variables that satisfy the conditions for IV estimation can be a challenge. Assuming that the researcher has access to a panel of bilateral trade flows, a simpler approach that allows for consistent estimation occurs when the policy variable is correlated with an unobserved component that is *fixed* or relatively slow moving over time; that is, Baier and Bergstrand (2007) assume that the error term can be expressed as

$$\ln(1 + \tilde{v}_{ijt}) = c_{ij} + e_{ijt}.$$

Estimates of the trade costs parameters are consistently estimated if the following assumption holds. Reintroducing time subscripts and defining  $s_{ijt}$  to be an indicator variable that equals unity when  $X_{ijt} > 0$  and zero otherwise, we express the gravity equation as

$$X_{ijt} = \exp\left(Z_{ijt}\beta\right) + u_{ijt}$$

---

<sup>40</sup>Table 4 in Head and Mayer report the results from a meta-type analysis covers more than 150 studies upward of 2500 estimates.

<sup>41</sup>See Egger et al. (2011) for a recent paper that employs IV techniques to study the impact of trade liberalization in a gravity setting.

where  $\exp\left(Z_{ijt}\beta\right) = \exp\left(\mathbf{z}_{ijt}^N\beta^N + \mathbf{z}_{ijt}^H\beta^H + \mathbf{w}_{ijt}^P\psi^P + \delta_{it}^X + \delta_{jt}^M\right)$ .

The four conditions that need to be satisfied in order to consistently estimate the log-linear, fixed effects structural gravity equation include:

$$\text{SGE FE 1 : } u_{ijt} = X_{ijt}^{SG}(\tilde{v}_{ijt})$$

$$\text{SGE FE 2 : } E(e_{ijt} | \mathbf{Z}_{ij}, c_{ij}, \mathbf{s}_{ij}) = 0$$

$$\text{SGE FE 3 : } E(s_{ijt} | \mathbf{Z}_{ij}, c_{ij}, u_{ijt}) = E(s_{ijt} | \mathbf{Z}_{ij}, c_{ij})$$

$$\text{SGE FE 4 : } E(e_{ijt}^2 | \mathbf{Z}_{ij}, \mathbf{s}_{ij}, c_{ij}) = \sigma_e^2$$

where  $\mathbf{Z}_{ij} = (Z_{ij1}, Z_{ij2}, \dots, Z_{ij(t-1)}, Z_{ijt})$  and  $\mathbf{s}_{ij} = (s_{ij1}, s_{ij2}, \dots, s_{ij(t-1)}, s_{ijt})$

- Assumption **SGE FE 1** again allows us to write the structural gravity equation in multiplicative form; that is,

$$X_{ijt} = X_{ijt}^{SG}(1 + \tilde{v}_{ijt}),$$

or log-linear as

$$\ln(X_{ijt}) = \mathbf{z}_{ijt}^N\beta^N + \mathbf{z}_{ijt}^H\beta^H + \mathbf{w}_{ijt}^P\beta^P + \delta_{it}^X + \delta_{jt}^M + c_{ij} + e_{ijt}. \quad (68)$$

- Assumption **SGE FE 2** implies that all right-hand side variables are *strictly exogenous*. The *strict exogeneity* assumption means that contemporaneous bilateral trade flows do not impact *any* current or future trade costs or influence the exporter-time/country-time controls; that is, there are no feedback effects.
- Assumption **SGE FE 3** allows bilateral trade flows to take on zero values, but the realization of zero trade flows are correlated with the right-hand side control variables and are uncorrelated with the error term  $e_{ijt}$ . Again this assumption assumes zero bilateral trade flow are *missing at random*.
- Assumption **SGE FE 4** again is the standard assumption that the errors are homoskedastic. As before, in case the assumption is not satisfied, efficiency gains can be obtained by using heteroskedasticity-robust standard errors.

Fixed effect estimation can account for the potential endogeneity.<sup>42</sup> Once pair fixed effects are introduced, all time-invariant variables are swept up by the pair fixed effects. This has two important implications. On a positive note, the pair fixed effects will absorb and control for all

---

<sup>42</sup>Baier and Bergstrand (2007) estimate the structural gravity equation using first differences, and Baier et al. (2014) estimate the structural gravity model assuming a random component. The preferred specification will depend on the error structure of  $e_{ijt}$ 's.

observable and unobservable bilateral trade costs that may affect trade in addition to trade policy. Egger and Nigai (2015) and Agnosteva et al. (2014) argue that the pair fixed effects do indeed carry systematic information about trade costs in addition to the information captured by the standard gravity variables. Thus, the comprehensive treatment of the time-invariant bilateral trade costs with pair fixed effects should improve the estimates of the time-varying bilateral trade policies and transportation improvements that can still be identified in the presence of all fixed effects. A potential challenge with the use of pair fixed effects is the inability to identify the impact of any time invariant determinants of trade flows, which may include important transportation links.

Table 2 reports gravity estimation results that are obtained after controlling for exporter-time, importer-time, and bilateral fixed effects. Given the set of our policy variables, the only time-varying bilateral covariate is the free trade agreement (dummy) variable. Once we control for the fixed effects, the regional trade agreement dummy is associated with higher bilateral trade. Note that the coefficient estimate on *FTA* is positive and significant. These results are consistent with those found in Baier and Bergstrand (2007) and Anderson and Yotov (2016).<sup>43</sup>

Many researchers include lagged effects of the policy variables in gravity regressions. The reason why these are included is because it is often the case that some of the tariff reductions or some aspects of the *provisions* are phased in over time. Furthermore, it is possible that the trade agreements may lead to new or tightening relationships between the trading partners that need time to realize. Columns (2)-(3) of Table 2 report the estimates from a specification with bilateral fixed effects and exporter-time and importer-time fixed effects that allow for phasing-in of the effects of FTAs. The estimates from column (2) include a single lag, while in column (3) we allow for two lags. For brevity, we focus our discussion on the estimates from column (3). Two main findings stand out. First, we note that once the lags are introduced, the current FTA variable is no longer statistically significant. The interpretation of this result is that FTAs may need some time to influence actual trade flows and, therefore, their immediate impact is small (and insignificant). Second, on a related note, we see that the estimates of the lagged FTA variables are both positive and statistically significant. This supports that intuition that FTAs need some time to impact trade. In addition, we see that the FTA estimates are increasing over time. Baier and Bergstrand (2007) and Anderson and Yotov (2016) experiment with longer lags to find that

---

<sup>43</sup>The impact of trade agreements on trade flows is smaller than in Baier and Bergstrand (2007) and Anderson and Yotov (2016). The difference with the estimates from Baier and Bergstrand (2007) may be due to the fact that we only assess the impact of agreements that were implemented after 1986, while the sample from Baier and Bergstrand covers agreements from the 1950s on. If it was the case that the earlier trade agreements created more trade, our coefficient estimates would be lower. See Baier et al. (2015) for an explanation on why the measured *average treatment effect* of *FTA* may be falling over time. The difference between the current estimates and the corresponding indexes from Anderson and Yotov (2016) could be due to the fact that the latter are obtained with the PPML estimator and with a sample that also includes intra-national trade flows. We explore these alternatives and their implications below.

the relationship is inverse u-shaped and that the average FTA effects are exhausted after ten to fifteen years of their initial implementation and entry into force.

It is often argued that trade agreements may be responding to changes in trade flows; that is, trade agreements are signed by countries that are already trading a lot. A simple test of reverse causality involves including a lead or leads of the trade agreement in the trade-flow equation. Column 4 of Table 2 reports the results of including a lead and we conclude that this test finds little support for reverse causality.

An alternative approach to estimating equation (68) is to note that the fixed effects can be modeled as

$$c_{ij} = \bar{z}_{ij}^N \pi^N + \bar{z}_{ij}^H \pi^H + \bar{w}_{ij}^P \pi^P + \bar{\delta}_j^M \pi^M + \bar{\delta}_i^X \pi^X + b_{ij} \quad (69)$$

where the bar denotes averages for each bilateral pair, the  $\bar{\delta}_j^M$  and  $\bar{\delta}_i^X$  are the average of the importer-year and exporter-year dummy variables which need to be included with unbalanced panels, and  $b_{ij}$  is random variable that is uncorrelated with gravity control variables. Then given the assumptions **SGE FE 1-SGE FE 4**, the model can be specified as

$$\ln(X_{ijt}) = \mathbf{z}_{ijt}^N \beta^N + \mathbf{z}_{ijt}^H \beta^H + \mathbf{w}_{ijt}^P \beta^P + \delta_{it}^X + \delta_{jt}^M + \bar{z}_{ij}^N \pi^N + \bar{z}_{ij}^H \pi^H + \bar{w}_{ij}^P \pi^P + \bar{\delta}_j^M \pi^M + \bar{\delta}_i^X \pi^X + b_{ij} + e_{ijt}. \quad (70)$$

Written in this form the model is referred to as a *correlated random effects model*.<sup>44</sup> The correlated random effects model employs unifies the fixed effects and random effects approach to modeling the unobserved heterogeneity. The correlated random effects coefficients are reported in column 5 of Table 2. As expected, the coefficient on the FTA variable is identical to the fixed effects coefficient reported in column 1.<sup>45</sup> The coefficients on the time-invariant gravity controls are also reported in column 5. These coefficients are similar to the cross-sectional and pooled OLS estimates reported in Table 1.

While the inclusion of bilateral fixed effects seems to allow us to more precisely identify the impact of regional trade agreements, bilateral *fixed effects* estimation may not be appropriate when more years and more countries are added to our data. Some authors have attempted to correct for trends in the trade costs and globalization in general. For example, Baier et al. (2014) employ a random growth model that allows for trends in trade costs over time. Incorporating internal trade flows, Bergstrand et al. (2016) adopt the structural gravity equation to allow for trends in globalization. An open challenge to trade and transport economists remains to account for the potential endogeneity of the trade cost variables when an *omitted variable* that is correlated with the endogenous variable is moving over time.

<sup>44</sup>See Cameron and Trivedi (2005) and Wooldridge (2010) for additional details.

<sup>45</sup>Estimating 70 by pooled OLS will yield the same coefficient estimates on the time varying covariates; in this case, the coefficient on FTAs.

## Log-Linear Structural Gravity Model and Sample Selection

When there are missing bilateral trade flow observations, log-linear estimation can yield consistent estimates as long the  $s_{ij}$ 's are uncorrelated with the error term in the structural gravity equation. Helpman et al. (2008) (HMR) address the case of incidental truncation in a model with heterogenous firms and fixed costs associated with exporting to each market. Egger et al. (2011) use a standard model of monopolistic competition with fixed bilateral exporting costs that will also deliver zero bilateral trade flows. These papers introduced *incidental truncation* and they propose a modified Heckman (Heckit) approach to account for firm heterogeneity and selection. We present a simplified version of this model for the ease of exposition. With a cross-section of data, to consistently estimate the parameter coefficients when selection is correlated with the unobserved components (incidental truncation) we need the following set of conditions:

$$\text{SGE SS 1 : } u_{ij} = X_{ij}^{SG}(\tilde{v}_{ij}) \text{ and } \ln(1 + \tilde{v}_{ij}) = v_{ij}$$

$$\text{SGE SS 2 : } s_{ij} = 1 [Z_{ij}\eta + Z^{IV}\eta^{IV} + v_{ij}^{ss} > 0] \text{ where } v_{ij}^{ss} \sim \mathcal{N}(0, 1)$$

$$\text{SGE SS 3 : } E(v_{ij} | \mathbf{Z}_{ij}, \mathbf{Z}_{ij}^{IV}) = 0$$

$$\text{SGE SS 4 : } E(v_{ij} | v_{ij}^{ss}) = \rho_{ss}v_{ij}^{ss}$$

As before **SGE SS 1** allows us to write the model in log-linear form. **SGE SS 2** states that selection depends on the standard gravity controls and  $Z^{IV}$  is an instrument that is correlated with the *fixed* exporting costs, but is not correlated with the error term in the structural gravity equation. As in Wooldridge (2010), **SGE SS 3** assumes that  $v_{ij}$  is uncorrelated with  $Z_{ij}$  and  $Z_{ij}^{ss}$ . **SGE SS 4** allows for the correlation between unobserved components in the selection equation and the error term in the structural gravity equation. Given these assumptions, the probability that country  $i$  exports to country  $j$  is given by

$$Pr(s_{ij} = 1) = \Phi\left(Z_{ij}\eta + Z^{IV}\eta^{IV}\right),$$

In this case, the log-linear HMR version of the gravity equation can be estimated by

$$\ln(X_{ij}) = Z_{ij}\beta_{ij}^{ss} + \rho_{ss}\lambda_{ij}^M$$

where  $\lambda_{ij}^M$  is the inverse Mill's ratio and  $\rho_v$  captures the correlation between the structural gravity error and the unobserved factors related to selection. The HMR specification includes additional terms that account for firm heterogeneity and assumes that  $v_{ij}^{ss} \sim \mathcal{N}(0, \sigma_{v^{ss}})$ . The approach pursued by Helpman et al. (2008) is appealing because it accounts for both selection and firm heterogeneity within a framework with solid theoretical foundations that can be elegantly nested within a standard gravity model. HMR show that firm heterogeneity is relatively more

important to account for and selection is of second-order importance. Baier and Garg (2016) extend the HMR to a panel framework and find that selection effects are more important than in the HMR paper, but firm heterogeneity still accounts for more of the variation captured by these additional terms. Despite its structural nature and intuitive appearance, a potential difficulty with the empirical implementation of the HMR approach is that it is hard to find an instrument that is correlated with fixed costs, but unrelated to bilateral trade flows. In addition, not all zeros are due to *incidental truncation*. Finally, Santos Silva and Tenreyro (2015) question the HMR assumptions of heteroskedasticity and normality and they argue that if heteroskedasticity is present in the selection equation, the coefficient estimates in the trade flow equation may be inconsistent.

## Pseudo-Maximum Likelihood Estimators

In many instances the assumptions needed for consistent estimation using a log-linear structural gravity equations are unlikely to hold. For example, the presence of heteroskedasticity and the frequency in which zero bilateral trade flows are observed in trade data have called for alternative estimation techniques. In order to simultaneously deal with zeros and heteroskedasticity, bilateral trade in exponential form can be consistently estimated using a Pseudo Poisson Maximum Likelihood (PPML) Estimator or a Pseudo Gamma Maximum Likelihood (PGML) Estimator. In each of these cases, as long as the mean function is properly specified, these estimators will provide consistent estimates. In order to assess the performance of these models, the Ramsey RESET test can be used as a specification test and a modified version of the Park test can be used for efficiency. For space considerations, we limit our empirical analysis to the PPML specification, but the RESET test and the Park test also apply to the PGML and the Pseudo Multinomial Maximum Likelihood estimators.<sup>46</sup> Before we present our empirical analysis, we discuss why heteroskedasticity can lead to inconsistent estimates of the semi-elasticities which trade and transport economists are interested in.

Santos Silva and Tenreyro (2006, 2011) demonstrate that, owing to *Jensen's Inequality*, the trade cost elasticities that are obtained from log-linear gravity estimations are likely to be inconsistent in the presence of heteroskedastic errors; that is, the variance of the error term is a function of the right-side variables. To see this in the context of the gravity equation estimation, consider the log-linear structural gravity equation given by

$$\ln(X_{ijt}) = Z_{ijt}\beta^{OLS} + u_{ijt},$$

---

<sup>46</sup>Santos Silva and Tenreyro (2006, 2008), Martínez-Zarzoso (2013), Head and Mayer (2014), and Feenstra (2015) discuss and report some results using these specification tests for PPML and other estimators.

where we assume  $E[u_{ijt}|\mathbf{Z}_{ijt}] = \mathbf{0}$  and  $E[u_{ijt}^2|\mathbf{Z}_{ijt}] = \sigma_v^2 \mathbf{h}(\mathbf{Z}_{ijt})$ . Typically, the coefficients on the trade costs are interpreted as the (semi) trade cost elasticities; that is, for some variable  $Z_{ijt}^k$  (semi-)trade cost elasticity is

$$\frac{\partial E(\ln(X_{ijt}))}{\partial Z_{ijt}^k} = \beta^k.$$

However, given the assumed functional form of the gravity equation

$$E[X_{ijt}|Z_{ijt}] = \exp(Z_{ijt}\beta + \frac{\sigma_v^2}{2}h(Z_{ijt})),$$

the trade elasticity of interest is  $\frac{\partial \ln(E[X_{ijt}|Z_{ijt}])}{\partial Z_{ijt}^k}$ , and we can see that the (semi) trade elasticity of expected bilateral trade with respect to variable  $Z_{ijt}^k$  is given by

$$\frac{\partial \ln(E[X_{ijt}|Z_{ijt}])}{\partial Z_{ijt}^k} = \beta^k + \frac{\sigma_v^2}{2} \frac{\partial (h(Z_{ijt}))}{\partial Z_{ijt}^k}.$$

Clearly,  $\frac{\partial E(\ln(X_{ijt}))}{\partial Z_{ijt}^k} \neq \frac{\partial \ln(E[X_{ijt}|Z_{ijt}])}{\partial Z_{ijt}^k}$ , and the OLS parameter estimates will reflect the trade cost elasticities of interest. It is often argued, however, that the heteroskedasticity assumption is violated and, therefore, the log-linear models yield inconsistent estimates of the trade cost elasticities when there is a mass of observations at zero. However, even in the absence of zeros, heteroskedasticity that is a function of the right-hand side controls can result in inconsistent estimates. Made in Manning and Mullahy (2001), this argument is even more prominently emphasized in the gravity context by Santos Silva and Tenreyro (2006, 2011).

The PPML estimators will deliver consistent estimates as long as the functional form of the mean is correctly specified. A simple functional form test for the conditional mean is the heteroskedasticity-robust RESET test outlined in Santos Silva and Tenreyro (2006). The procedure for the RESET test is:

**PPML RESET 1 :** Estimate the PPML model

**PPML RESET 2 :** Save the fitted value  $FIT_{ij} = Z_{ij}\hat{\beta}$  and construct squared (or higher-order) fitted values  $SQFIT_{ij} = (Z_{ij}\hat{\beta})^2$

**PPML RESET 3 :** Estimate the PPML model with the constructed squared fitted values

**PPML RESET 4 :** Test  $SQFIT_{ij} = 0$

Efficiency will depend on the functional form of the variance. If the structural gravity equation mean is given by  $\exp(Z_{ijt}\beta_{ij})$ , Manning and Mullahy (2001) state that the variance of GLM models is given by

$$(X_{ijt} - X_{ijt}^{SGE})^2 = h * \exp(Z_{ijt}\beta_{ij})^\lambda$$

- If  $\lambda = 0$  then variance is constant and non-linear least squares is the more efficient estimator.
- If  $\lambda = 1$  then the PPML estimator would be more efficient.
- If  $\lambda = 2$  then the PGML estimator would be more efficient.

A modified version of the Park test can be carried out following the guidelines suggested Deb et al. (2013).<sup>47</sup>

**PPML Park 1 :** Estimate the PPML model

**PPML Park 2 :** Save the fitted value  $FIT_{ij} = Z_{ij}\hat{\beta}$  and squared error  $V\hat{A}R_{ij} = (X_{ijt} - \hat{X}_{ijt})^2$

**PPML Park 3 :** Estimate the PGML model with the constructed squared error ( $V\hat{A}R_{ij}$ ) on the fitted values  $Z_{ij}\hat{\beta}$

**PPML Park 4 :** Test  $\lambda = 1$  for Poisson

Table 3 and Table 4 present the results from the PPML specification. In order to more easily compare these estimates with Table 1 and Table 2 the importance of the presence of zeros vs. heteroskedasticity, we present two sets of PPML results for each year in our sample. First, we obtain estimates that are based on positive trade flows only. Then, we report results that are based on data that include zero trade flows. Comparison between the estimates with and without zero trade flows for each year in our sample reveals that they are virtually identical. Thus, it is tempting to conclude that the presence of zeros is not a serious issue. However, we caution the reader that, as can be seen from the bottom of Tables 3 and 4, our sample does not include many zeros. The problem of zeros will be more pronounced in more disaggregated samples, and especially so for services trade.

Comparison between the OLS estimates and the PPML estimates reveals some similarities and also some differences. As with the OLS specification, distance is negatively correlated with trade flows, however, we find that the PPML estimates of the effects of distance are significantly smaller

---

<sup>47</sup>Given estimates of  $X_{ijt}^{SGE} = \exp(Z_{ijt}\beta_{ij})$ , it is possible to construct  $(X_{ijt} - X_{ijt}^{SGE})^2 = h * \exp(Z_{ijt}\beta_{ij})^\lambda$  and estimate the parameters. Santos Silva and Tenreyro (2006) suggest a linearized version of the variance to assess whether a specific empirical model can be used. Taking a first-order Taylor series around  $\lambda = \hat{\lambda}$  yields

$$(X_{ijt} - X_{ijt}^{SGE})^2 = h * (X_{ijt}^{SGE})^{\hat{\lambda}} + \lambda * (\lambda - \hat{\lambda}) * \ln(X_{ijt}^{SGE}) * (X_{ijt}^{SGE})^{\hat{\lambda}}.$$

Because heteroskedasticity may be present, Santos Silva and Tenreyro (2006) suggest using weighted least squares to assess the estimates of the linearized specification. If the estimated coefficient  $(\lambda * (\lambda - \hat{\lambda}))$  on  $\ln(X_{ijt}^{SGE}) * (X_{ijt}^{SGE})^{\hat{\lambda}}$  is not significantly different from zero, then the associated GLM estimator cannot be rejected as an appropriate functional form for the variance of the error term.



as compared to their OLS counterparts. Similarly, we find that the positive impact of language on bilateral trade is estimated to be smaller in the PPML results. Another difference between the two sets of estimates is that, unlike the OLS estimates, contiguity is robustly associated with higher bilateral trade; whereas, colonial ties are not. Finally, we note that, unlike the OLS estimates of the effects of FTAs, which varied from negative and significant to positive and significant over time, the PPML estimates on FTAs are positive and significant in all years in the sample. However, the coefficient estimates tend to vary a lot across years; ranging from as high as 1.02 to as low as 0.411 and the estimates show no particular pattern. This type of variability was one of the reasons why Baier and Bergstrand (2007) proposed the panel fixed effects method, which we also implement below in this chapter.

Tables 3 and 4 also present the results for the Ramsey-RESET test and the modified Park test outlined above. For most years, the RESET test provides support for the exponential functional form. The estimates of  $\lambda$  for the modified Park test typically reject the proportional variance assumption. However, the coefficient estimates are closer to proportional variance than the squared variance, as would be the case with PGML estimator.

## Structural Gravity with Internal Trade Flows

Most of the sales of both goods and, especially, services are domestic rather than international. This empirical fact holds after controlling for size or any other country-specific characteristics, and it is often labeled as the ‘*home bias*’ in international trade. Furthermore, while the structural gravity model explicitly accounts for *intra-national* trade and these flows are crucial for general equilibrium analysis and welfare calculations, cf. Arkolakis et al. (2012), the vast majority of the estimating gravity studies do not take intra-national trade flows into account. Yotov et al. (2016) point to several compelling arguments for the use of intra-national trade flows in gravity estimations. First, as noted above, the inclusion of intra-national trade flows will ensure consistency with the structural gravity model, as consumers choose between domestic and foreign varieties. Second, as demonstrated by Dai et al. (2014) the inclusion of intra-national trade flows enables researchers to allow for the possibility that bilateral trade agreements create trade between member countries while diverting away from domestic sales. Third, as argued by Heid et al. (2015), the presence of observations for intra-national trade in estimating gravity equations allows for identification of the effects of non-discriminatory trade policies, e.g. MFN tariffs or export subsidies, which would be absorbed by the importer and exporter fixed effects, respectively, when the gravity equation is estimated with international trade flows only. Fourth, Yotov (2012) and Bergstrand et al. (2015) use intra-national trade flows and adhere closer to gravity theory to demonstrate that the famous ‘distance puzzle’ (see Disdier and Head (2008)) and ‘missing globalization puzzle’ (see Coe et al. (2002)) are both resolved when the effects of distance and globalization on international trade are measured relative to the corresponding ef-

fects on intra-national trade. Finally, the inclusion of internal trade flows in gravity estimations will allow for a consistent comparison between estimated trade costs and calibrated trade costs as discussed in Footnote 26.

Motivated by these arguments, in this section we estimate the gravity equation with international and intra-national trade flows and we discuss implications. Estimation results are presented in Tables 5 and 6. In order to obtain these estimates, first, we add observations for *intra-national* trade flows to our dependent variable.<sup>48</sup> In order to calculate the effect of a country's own border, we include a dummy variable *INTERNAL* that equals one when bilateral trade is intra-national trade and zero otherwise. Several findings stand out from the estimates in Tables 5 and 6. First, as to be expected, we obtain a large positive and statistically significant estimate of the coefficient on internal trade. This result is consistent with the large 'home bias' in international trade. Second, we note that the inclusion of internal trade flows does not have a very significant impact on the estimates of the standard gravity variables. Two differences that are worth mentioning include the slightly larger estimate on contiguity and the smaller variation in the impact of FTAs over time.<sup>49</sup>

Table 7 studies the implications of adding intra-national trade flows for the PPML estimates of the FTAs in a panel setting with pair fixed effects.<sup>50</sup> For brevity we focus the analysis on specifications that include the zero trade flow observations and, consistent, with our symmetric treatment of the standard gravity variables, we use symmetric pair fixed effects.<sup>51</sup> As before, we allow for phasing-in FTA effects and we present the estimates with and without internal trade in successive columns. The main result from Table 7 is that when intra-national trade flows are added to the sample, the estimates of the FTAs increase significantly in magnitude. This result is consistent with the findings of Dai et al. (2014), who explain it as capturing the fact that FTAs create trade between members at the expense of domestic sales.

---

<sup>48</sup>The intra-national trade flows data used in our analysis was constructed by Tom Zylkin as *apparent consumption* defined as the difference between total value of gross production and the total value of exports for each country in the sample. The original source for gross production data is the United Nations Industrial Development Organization (UNIDO) Industrial Statistics (INDSTAT) database. An alternative method to construct intra-national trade flows is to use the complete matrix of international input-output tables and take into account the value added contribution at each stage and to track the global value chain throughout the production process. We refer the reader to Yotov et al. (2016) who offer further discussion on the construction of internal trade flows as well as a detailed presentation and discussion of various databases that are used in gravity estimations.

<sup>49</sup>One cause for concern is that the RESET test and the modified Park test provide less support for the PPML specification.

<sup>50</sup>It is somewhat surprising that the contemporaneous impact of FTAs is not significant without the inclusion of internal trade flows. This is likely due to the sample of countries and the time period chosen.

<sup>51</sup>Estimation results based on positive trade flows only and with asymmetric pair fixed effects are virtually identical to the results from Table 7 and are available by request.

## Structural Gravity and Trends in Globalization

One of the more surprising results from the gravity literature is the relative constancy (or increasing effect) of distance on bilateral trade. Yotov (2012) and Bergstrand et al. (2015) show how inclusion of intra-national trade and panel data techniques accounting for trends in globalization may help explain the *distance puzzle*. Their departing point from column (1) of Table 8, where we used PPML and exporter-time and importer-time fixed effects in order to identify the effects of FTAs, is to include a set of indicator variables that account for trends in globalization. To this end, we construct a series of indicators for each time period that equal unity if the bilateral trade value is international trade flow and zero if the bilateral trade is internal trade. If there are trends in globalization over time, we would expect the coefficient on these indicators to be increasing over time. That is, we follow Bergstrand et al. (2015) to account for possible globalization effects by including a series of indicator variables ( $BRDR\_T$ ) that take a value of one for international trade and are set to zero for intra-national trade for each year ( $T$ ) in our sample. Due to perfect collinearity, we can identify the effects of all but one of the border dummies. We chose the border dummy for the first year in our sample (1986) as the reference group for the analysis. Accordingly, all other border estimates from column (2) of Table 8 should be interpreted as deviations from the border effects in 1986. With these considerations taken into account, the resulting estimating equation becomes:

$$X_{ij,t} = \exp[\beta_1 FTA_{ij,t} + \delta_{it}^X + \delta_{jt}^M + c_{ij} + \sum_{T=1990}^{2006} \beta_T BRDR\_T_{ij}] + u_{ijt}. \quad (71)$$

Comparison between the estimates from columns (1) and (2) of Table 8 reveals two important results. First, we see that the estimate of the effects of free trade agreements in column (2) are smaller as compared to the corresponding index from column (1). The new FTA estimate of 0.206 (std.err. 0.066) suggests that, on average, the free trade agreements that entered into force during the period of investigation have led to about 23 percent increase ( $100 * [e^{0.206} - 1]$ ) in the volume of bilateral trade during the period of investigation. The intuition for this result is that the specification with intra-national trade flows has captured the fact that some of the trade flows that trade agreements ‘create’ between members are sales that are diverted from domestic trade.

Second, we note that the  $BRDR$  estimates are all positive and increasing over time. (Note that the smallest difference in the border estimates is between  $BRDR_{ij,1998}$  and  $BRDR_{ij,2002}$ . A possible intuitive explanation for this result is the drop in trade in response to the 2001 recession.) As described earlier, all of the identified border estimates should be interpreted relative to the omitted reference border effect for 1986. Thus, our results reveal that globalization has played a significant role in promoting bilateral trade flows. Specifically, the estimate of

0.747 (std.err. 0.035) for  $BRDR_{ij,2006}$  implies that globalization forces have contributed to an increase in international relative to intra-national trade by 111 percent ( $100 * [e^{0.747} - 1]$ ) during the period 1986-2006, i.e. our results suggest that globalization forces have lead to a doubling in international trade flows relative to domestic sales. This finding is in contrast to the argument from Coe et al. (2002) that globalization is everywhere but in gravity models. Importantly, while we do obtain positive and significant effects of globalization on trade, our setting does not allow us to identify the specific sources that cause these effects. Improvements in communication and transportation are natural potential candidates to explain our findings and we view this as an important area for future work. Instead, in the next section, we will use the FTA and globalization effects estimates from column (2) in order to demonstrate how these partial estimates can be translated into general equilibrium effects with the structural gravity model.

## 4 Numerical Exercises and Counterfactuals Analysis

While trade policy changes and improvements in transportation channels are often made unilaterally or on a bilateral basis, the impact of such trade cost changes may have significant ripple effects throughout the world. This is the main reason why most trade policy analyses are performed in a general equilibrium (GE) setting. The objective of this section is to demonstrate how the structural gravity model can be used to quantify the general equilibrium impact of changes in trade costs.<sup>52</sup>

### 4.1 GE Analysis with Structural Gravity

A standard general equilibrium counterfactual experiment in the trade literature is to evaluate the change in total exports, nominal wages, consumer prices, welfare, etc. in response to a hypothetical change in trade costs. The latter could be a reversion to autarky, an introduction or elimination of a trade agreement, a change in tariffs, a change in transportation costs, etc. Since, as demonstrated earlier in this chapter, the different models described above result in similar systems of equations, we will focus on the generic single-sector version of the structural gravity model in order to show how the gravity framework can be used to perform general

---

<sup>52</sup>We refer the interested reader to Head and Mayer (2014), Costinot and Rodríguez-Clare (2014) and Yotov et al. (2016) for complementary and more extensive GE analyses and discussions.

equilibrium exercises:<sup>53</sup>

$$X_{ij} = \left( \frac{t_{ij}}{\Pi_i P_j} \right)^{-\epsilon} (W_i L_i)(W_j L_j) \quad (72)$$

$$\Pi_i = \left[ \sum_{j=1} \left( \frac{t_{ij}}{P_j} \right)^{-\epsilon} W_j L_j \right]^{\frac{-1}{\epsilon}} \quad (73)$$

$$P_j = \left[ \sum_{i=1} \left( \frac{t_{ij}}{\Pi_i} \right)^{-\epsilon} W_i L_i \right]^{\frac{-1}{\epsilon}} \quad (74)$$

$$W_i = B \left( \frac{A_i}{\Pi_i} \right)^{\frac{-\epsilon}{\epsilon+1}}. \quad (75)$$

In order to describe the response of the endogenous variables of interest to changes in trade costs, system (72)-(75) has to be solved twice; once in the baseline scenario, i.e. the scenario with trade costs as observed in the data, and once in the counterfactual scenario, i.e. the scenario with the hypothetical vector of bilateral trade costs which reflects the policy change of interest. The following steps deliver the solution of the model in the baseline scenario. First, with given trade data, equation (72) obtains the initial vector of bilateral trade costs. As mentioned earlier, the trade costs vector can be obtained with calibration or estimation methods. Our main focus throughout the chapter has been on estimation. Accordingly, we will use a combination of data and elasticity estimates from our preferred econometric specification (71) to define (the power transform of) the vector of baseline trade costs as:

$$(\hat{t}_{ij}^b)^{-\epsilon} = \exp[\hat{\mu}_{ij} + \hat{\beta}_1 FTA_{ij,t} + \hat{\beta}_{2006} BRDR_{2006ij}], \quad (76)$$

where,  $\hat{\mu}_{ij}$ ,  $\hat{\beta}_1$ , and  $\hat{\beta}_{2006}$  are the estimates of the time-invariant trade costs, the average FTA effects, and the effect of globalization on international trade, respectively, from column (2) of Table 8, and we have used superscript ‘b’ to denote ‘baseline’. We note that equation (76) delivers estimates of the trade costs for 2006, which is the last year in our sample. This is consistent with the fact that most counterfactual experiments in the trade and trade policy literature are

---

<sup>53</sup>The choice to focus on the simple one-sector version of the model is for expositional simplicity. Our setting corresponds the original GE frameworks of Eaton and Kortum (2002) and Anderson and van Wincoop (2003). Stimulated by these seminal papers, a series of more recent studies have extended system (72)-(75) to incorporate non-homothetic preferences, e.g. Fieler (2011); sectors, e.g. Costinot et al. (2012); dynamic forces, e.g. Olivero and Yotov (2012); intermediate goods, e.g. Caliendo and Parro (2015), the informal sector, e.g. Heid (2015); growth through capital accumulation Anderson et al. (2015b); unemployment, e.g. Heid and Larch (2016); financial markets, e.g. Eaton et al. (2016), etc. We note that, under certain assumptions, each of these extended gravity models collapses to the generic system (72)-(75). Thus, the analysis that we offer in this section can be viewed and should serve as a useful benchmark.

performed at a single baseline year. In principle, one can obtain a vector of trade costs for each year in the sample and perform counterfactuals over the whole sample period. For expositional simplicity, in this section we follow the standard approach and focus on a single year. However, extending the analysis to multiple years is straightforward.

In combination with data on national income/expenditure,  $W_i L_i$ , a value for the trade elasticity parameter,  $\epsilon$ ,<sup>54</sup> and subject to an initial normalization of wages,  $W_i = 1, \forall i$ , the vector of bilateral trade costs,  $(\hat{t}_{ij}^b)^{-\epsilon}$ , can be used to solve system (73)-(75) for the multilateral resistances,  $\Pi_i$  and  $P_j$ , and for the country-specific vector of ‘preference-adjusted’ technology,  $B(A_i)^{\frac{-\epsilon}{\epsilon+1}}$ .<sup>55</sup> As noted by Anderson and Yotov (2010), system (73)-(75) can be solved for the multilateral resistances only up to a scalar “if  $\{\Pi_i^0, P_j^0\}$  is a solution, then so is  $\{\lambda \Pi_i^0, P_j^0 / \lambda\}$ ” (p.2160). Therefore, a normalization is needed. We follow Anderson and Yotov (2010) to set the inward multilateral resistance for a country of choice, in our case Germany, to be equal to one. Thus, all other inward and outward multilateral resistances will be measured relative to the IMR for Germany. The solutions of the multilateral resistances in the baseline system can be used, together with data on trade costs, output, expenditure, etc., to obtain any indexes of interest including total exports,  $\sum_j \hat{X}_{ij}^b, \forall j \neq i$ , and real income,  $(\widehat{W}_i \widehat{L}_i)^b / \hat{P}_i^b$ , for each country in the sample in the baseline scenario.

After the model is solved and the various indexes of interest are obtained in the baseline scenario, system (72)-(75) should be solved in the counterfactual scenario, where the new vector of bilateral trade costs reflects the hypothetical trade cost change of interest:

$$(\hat{t}_{ij}^c)^{-\epsilon} = \exp[\hat{\mu}_{ij} + \hat{\beta}_1 FTA_{ij,t}^c + \hat{\beta}_{2006} BRDR\_2006_{ij}^c]. \quad (77)$$

Here, superscript ‘*c*’ denotes ‘*counterfactual*’ and all other variables and parameters are defined above. Equation (77) captures two different possibilities to define counterfactual trade costs, which we explore empirically below. Specifically, it will enable us to (i) introduce or to remove a regional trade agreement by changing the definition of the FTA dummy variable,  $FTA_{ij,t}^c$ ; or (ii) study the impact of globalization by changing the definition of the border dummy  $BRDR\_2006_{ij}^c$ .<sup>56</sup> Given the new vector of bilateral trade costs  $(\hat{t}_{ij}^c)$ , and using the baseline values

---

<sup>54</sup>As demonstrated in Section 2.4, the value of  $\epsilon$  can be estimated directly from an empirical gravity equation with data on a direct price-shifter such as tariffs. Alternatively, and as is the standard practice in the profession,  $\epsilon$  can be borrowed from the literature. We follow this practice and set  $\epsilon = 6$  throughout the analysis.

<sup>55</sup>As noted in Section 2.3, the structural gravity system can also be expressed and solved directly in changes. This approach has gained popularity due to the work of Dekle et al. (2007, 2008). Anderson et al. (2016) capitalize on a property of the PPML estimator to demonstrate how general equilibrium analysis can be performed directly in standard software packages such as Stata.

<sup>56</sup>In principle, one may also change the time-invariant trade cost component  $\hat{\mu}_{ij}$  and the estimates of the FTA effects and globalization, the  $\beta$ ’s. However, we choose to focus here on the more interesting and policy relevant scenarios.

of all parameters and exogenous variables, system (72)-(75) can be solved for the counterfactual values of the multilateral resistances, nominal wages, total exports, and any combination of those indexes, which might be of interest to the researcher or the policy maker. The percentage changes in the indexes of interest between the baseline and the counterfactual scenario would give the general equilibrium impact of the policy or trade cost change in question.

Next, we will explore each of the two possibilities for trade policy changes that are offered by equation (77) to perform two illustrative counterfactual scenarios by (i) introducing a hypothetical FTA between Great Britain and the United States, and (ii) studying the impact of globalization over the period 1986-2006. Before we perform and discuss the results of these practical policy experiments, however, we use our framework to establish some general and intuitive theoretical relationships: we investigate the general equilibrium effects of a uniform increase in technology across all countries; that is  $\hat{A}_i = \gamma > 1 \forall i$ . The second experiment we consider is a uniform decrease in trade costs across all markets; that is, we assume that  $t_{ij} = \mu$  where  $0 < \mu < 1$  for all  $ij$ .

**Theorem 1.** *In the case of a uniform increase in technology, we conjecture that bilateral trade increases by  $\gamma^{\frac{\epsilon}{\epsilon+1}}$ , outward multilateral resistance does not change, the price index falls by  $\gamma^{\frac{-1}{\epsilon+1}}$  and the real wage increases by  $\gamma$ .*

*Proof.* Substitute for the increase in technology into equation 47 assuming that  $\hat{\Pi}_i^{-\frac{\epsilon}{\epsilon+1}} = 1$ . This implies that  $\hat{W}_i = \gamma^{\frac{\epsilon}{\epsilon+1}}$ . Substituting for the wage increase and assuming that  $\hat{\Pi}_i^{-\frac{\epsilon}{\epsilon+1}} = 1$  into equation 46 implies that  $\hat{P}_j = \gamma^{\frac{-1}{\epsilon+1}}$ . Substituting for the changes in wages and the multilateral resistance terms into equation 44 implies bilateral trade grows by  $\gamma^{\frac{\epsilon}{\epsilon+1}}$ . Substituting for the changes in wage and the price index gives the change in the real wage or  $\hat{W}/\hat{P} = \gamma$  and this complete the sketch of the proof.  $\square$

Intuitively, this makes sense. A uniform increase in technology results in higher wages. However, the wages do not rise one-for-one with the increase in technology because the additional production puts downward pressure on the price of the output and so wages rise by  $\gamma^{\frac{\epsilon}{\epsilon+1}}$ . The uniform increase in productivity means that no one firm will experience a relative change in its market access; as a result, market access for all producers remains unchanged. Since each country is producing more output, the additional production puts downward pressure on all prices. Since all prices fall by  $\gamma^{\frac{-1}{\epsilon+1}}$ , the price index in each country will fall by the amount. The change in welfare is partially due to the increase in the nominal wage and the fall in the price index.

**Theorem 2.** *Now consider a uniform reduction in trade costs given by  $\hat{t}_{ij} = \mu$  for all  $ij$ . In this case, there will be no impact on multilateral resistance and the wage rate will remain constant. As iceberg trade costs fall uniformly the price index will fall as well and the change in inward multilateral resistance is given by  $\hat{P}_j = \mu$  for all  $j$ .*

*Proof.* The proof is again by construction. Assuming no change in the outward multilateral resistance term, we see that the wage rate in all countries will remain constant. As a result, we can substitute for the change in trade costs into the inward multilateral resistance terms given that wages and the outward multilateral resistance terms do not change to see that inward multilateral resistance falls by  $\mu$ . Substituting these values for the change in the wage rate, the inward multilateral resistance, and the change in trade costs into the expression for the outward multilateral resistance terms, we can confirm that outward multilateral resistance does not change. Since trade costs and the price index fall by the same percentage, bilateral trade is unaffected (as in Anderson and van Wincoop). The real wage then increases by  $1/\mu > 1$ .  $\square$

Intuitively, this seems plausible since the decline in trade costs does not make any one worker more productive so it should have no impact on her wage rate. Since the fall in trade costs is uniform across all markets, no single firm will have better market access and so there will be no change in market access. Since less of the iceberg melts with each shipment, the fall in trade costs implies the price index will fall because consumers have access to more goods. As a result, the welfare increase is inversely related to the fall in trade costs.

Rarely is it the case where changes in trade costs or changes in incomes are symmetric across countries. In order to evaluate asymmetric changes, we need to incorporate empirical estimates and embed those into a general equilibrium framework. The next two subsections conduct some simple policy experiments to calculate the general equilibrium effects of changes in trade costs.

## 4.2 On the GE Effects of a FTA between Great Britain and US

In our first experiment we simulate the general equilibrium impact of a bilateral free trade agreement between UK and US. Our choice was stimulated by the recent separation of Great Britain from the European Union and by the fact that a bilateral treaty between the two countries may soon be considered. In order to simulate the effects of this agreement we assume that its initial (partial equilibrium) impact will be equal to the average partial FTA effect ( $\hat{\beta}_1 = 0.206$ ) from column (2) of Table 8.<sup>57</sup> Specifically, we construct the counterfactual vector of bilateral

---

<sup>57</sup>An alternative approach to introducing the initial change in trade costs to the model is to use reductions in observable trade barriers, e.g. tariffs and NTMs. However, since tariffs are very small, on average and especially between developed countries such as UK and US, and since most of the important NTMs, e.g. various sanitary and technical standards, are non-discriminatory in nature, we chose to use the average FTA estimate from our sample. Our choice is consistent with the belief that FTAs offer significant stimuli to trade, in addition to simply eliminating observable barriers to trade. Baier et al. (2016) propose a two-step procedure that better capitalizes on the information contained in the estimated effects of prior agreements in order to construct pair- and direction-specific estimates of the effects of FTAs that have not entered into force yet.



trade costs as follows:

$$(\hat{t}_{ij}^c)^{-\epsilon} = \exp[\hat{\mu}_{ij} + \hat{\beta}_1 FTA_{ij,t} + \hat{\beta}_1 FTA\_UK\_US_{ij,t} + \hat{\beta}_{2006} BRDR\_2006_{ij}]. \quad (78)$$

Estimation results of the GE effects of the UK-US agreement are reported in panel “A. GE Effects of UK-US FTA” of Table 9, where we present changes in total exports, real GDP, nominal wages, and consumer prices. Several findings stand out. First, we note that the gains for the FTA members, i.e. UK and US, are relatively small. The estimated increase in exports for US is a bit more than 1 percent, while Great Britain’s exports will increase by less than 2.5 percent. The real income gains are even smaller. UK will enjoy an increase in real income of only 0.2 percent, while, at 0.03 percent, the gain for US will be hardly noticeable. Finally, we note that UK producers will gain relatively more (0.14 percent increase in nominal income) as compared to US producers, and UK consumers will enjoy a relatively larger (but still small) decrease in consumer prices as compared to US consumers.<sup>58</sup>

There are several explanations for the relatively small gains for the FTA members: (i) The average FTA estimate that we obtained in Table 8 is relatively small. Thus, the small initial change in trade costs between the two countries results in small GE effects. For comparison, Baier and Bergstrand (2007) obtain an average FTA effect of 0.76, which is about three times larger as compared to our FTA estimate from Table 8; (ii) Mainly owing to geographical (as opposed to policy) factors, trade costs between UK and US are relatively large to start with; (iii) UK is already quite integrated with a series of developed trading partners (e.g. Germany and France) that have very similar comparative advantage as compared to US; (iv) Note also that while we have introduced the new FTA between UK and US, we have not explicitly captured the fact that UK has separated from the European Union; Finally, (v) we remind the reader that our sample only covers manufacturing. Thus, it omits some important sectors, e.g. services, that account for much of the economic activity, especially relevant is the role of services as intermediates for manufacturing production, in both US and UK. Furthermore, we expect that there will be a significant heterogeneity in the FTA effects across manufacturing sectors that is not captured by our aggregate manufacturing data. This discussion points to potentially important improvements that have to be taken into account when interpreting and extending our benchmark results.

Finally, we note that the effect of the potential FTA between UK and US on outside countries will be very small. This is encouraging evidence that the negative impact of regional integration on non-member countries is not very significant. The variation in the effects on outsiders is intuitive and trade diversion is a natural explanation for our findings. Thus, for example, we see that Canada and Mexico are among the countries that will suffer the most from the UK-US FTA. Furthermore, we observe that Canada will actually suffer more than Mexico. The intuition for

---

<sup>58</sup>We remind the reader that all effects on wages and consumer prices are, by construction, relative to the effects on consumer prices in Germany.

the first result is that Canada and Mexico are among the closest and most dependent US trade allies and, given the size and dependence of these countries on the US economy, any competition on the US market may lead to significant disruptions in the economies of these small partners. The relatively larger impact on the Canadian economy, as compared to Mexico, can be explained by the fact that Great Britain is much closer in terms of comparative advantage to Canada than it is to Mexico.

Other countries that are in the upper end of the distribution in terms of losses due to the UK-US FTA include Nigeria, Kenya and Ireland. The explanation for the effects on latter is that, on the one hand, Ireland is very closely geographically and economically tied to the UK, while at the same time this country is relatively close to the US in terms of comparative advantage. The explanation for the relatively large impact on Nigeria and Kenya are the close ties of these countries to UK due to colonial relationships. Finally, we remind the reader that the introduction of the UK-US agreement in this experiment is not accompanied by a removal of UK from the EU. This may explain the relatively small impact that we obtain for the current EU members in our sample.

### 4.3 On the GE Impact of Globalization, 1986-2006

In our second experiment, we use the partial equilibrium estimates from column (2) of Table 8 to study the impact of globalization. Specifically, we simulate a scenario in which we eliminate the effects of globalization in 2006, i.e. we ask the question: What would the world have looked like (in terms of real GDP and distributional effects on the consumers and producers in each country) if the globalization forces reflected in our partial equilibrium gravity estimates from Table 8 did not take place? Mechanically, we solve a counterfactual scenario where the underlying hypothetical bilateral trade cost vector for 2006 is constructed without taking into account the estimate of  $BRDR_{2006}$ :

$$(\hat{t}_{ij}^c)^{-\epsilon} = \exp[\hat{\mu}_{ij} + \hat{\beta}_1 FTA_{ij,t}], \quad (79)$$

In addition to quantifying the economic impact of globalization on the countries in our sample, from a methodological perspective, this exercise will demonstrate how a uniform shock to the world trading system may have quite heterogeneous effects across countries depending on their importance and initial conditions in this system.

Our findings are reported in panel “B. GE Effects of Globalization” of Table 9. Similar to our analysis of the effects of the UK-US agreement, we report the GE impact on total exports and real GDP and, in addition, we decompose the effects of globalization on the consumers and on the producers in each country. Several findings stand out. First, the estimates from column  $\% \Delta X_{ij}$  of panel B, which are all negative and sizable in magnitude, reveal that total exports for

each of the countries in our sample would have been significantly lower in the scenario without globalization.

Second, despite the fact that the change in the bilateral trade costs vector is uniform, we obtain very heterogeneous effects on total exports across the countries in our sample. Specifically, we estimate a decrease of between 60% and 70% in the exports of smaller and poorer countries such as Nigeria (67%), Senegal (67%), Iran (66%), Kenya (66%), and Qatar (63%). On the other side of the spectrum, the countries whose exports would have suffered the least (but still significantly) include China (25%), Ireland (27%), Japan (32%), Singapore (32%), and Malaysia (33%).

Turning to the analysis of real GDP, the estimates from column  $\% \Delta RGDP$  suggest that, on average, the countries in our sample would have been about 3.3% poorer without the impact of globalization between 1986 and 2006. However, similar to the impact on exports, the effects of globalization on real GDP vary quite a bit across countries. Specifically, our estimates suggest that the countries that have benefited the most from globalization include some Asian economies (e.g. Hong Kong (9%), Macao (7%), and Singapore (7%)), some smaller European economies (e.g. Holland (7%), Ireland (7%), and Denmark (6%)), as well as Canada (6%) and Mexico (5%).

Finally, the estimates from columns  $\% \Delta W$  and  $\% \Delta IMR$  of Table 9 speak to the distributional effects of globalization on consumers and producers in the world. We remind the reader that, by construction, all indexes that capture effects on consumers and producers are measured relative to the impact on the consumers (via consumer prices, i.e. IMRs) in Germany. Thus, we can only make relative comparisons. Our estimates suggest the following. Based on the results from column  $\% \Delta W$ , we conclude that the producers who have benefited the most from globalization include producers in China, Ireland, Singapore, Japan, Korea, and Malaysia, while the producers who have benefited the least are in Kenya, Senegal, Nigeria, Ecuador, and Iran. We find these results intuitive.

Turning to consumers, the estimates from column  $\% \Delta IMR$  reveal that, with slight variations, the consumers in the countries where producers gain the most from globalization are actually the ones who have ‘suffered’ the highest prices, and vice versa. The intuitive explanation for this result is that once producers face better options to export they also increase domestic prices. Another very important message with significant policy implications from our findings in panel B is that the gains in terms of real GDP for the poorest countries in the world have come not from higher nominal income but rather from lower consumer prices.

The analysis in this section should serve as an informative benchmark but it is also subject to caveats. Some of these caveats, e.g modeling sectors, introducing intermediates, using more comprehensive datasets, etc. have already been addressed by existing studies (see Footnote 53). Other caveats, however, have not been tackled yet and can potentially be addressed through a closer collaboration and better understanding of the impact of transportation improvements on trade costs in a general equilibrium setting. We point to such unexplored areas as possible

fruitful directions for future work in the concluding remarks.

## 5 Conclusion

The main objective of this chapter was threefold. First, we reviewed the main theoretical developments in the gravity literature from its very early, a-theoretical applications to the latest structural contributions. Next, we discussed challenges and implemented methods to estimate empirical gravity equations and we presented a series of estimation results that emphasized the importance of proper account for various data and estimation issues with gravity estimations. Finally, we demonstrated how to perform numerical general equilibrium analysis and simulations with the structural gravity model. Throughout the analysis in this chapter we attempted to emphasize the links and importance of transportation costs for the trade literature and we pointed to avenues where we believe interdisciplinary contributions between the international trade and transportation economics fields will be most valuable.

We see several areas where the gravity trade literature can benefit from better transportation data and better understanding of transportation costs in order to improve gravity estimations and partial equilibrium analysis. First, while the use of pair fixed effects has established itself as a comprehensive and flexible method to account for time-invariant trade costs, the gravity literature lags behind in our treatment of time-varying bilateral costs. Anecdotal evidence points to the fact that improvements in transportation and communication have played significant role in decreasing trade costs and promoting trade. However, mostly due to lack of comprehensive datasets, such improvements have been accounted for mainly in case studies. We expect that proper treatment and account of transportation improvements may point to explanations for the ‘distance’ and ‘missing globalization’ puzzles in international trade. Another open challenge to trade and transport economists is to identify the factors that have contributed to the growth of world trade above and beyond income growth. Finally, very often transportation improvements are country-specific, e.g. the building of a new autobahn or improvements in domestic infrastructure in general. The treatment of the multilateral resistances with exporter and importer fixed effects in structural gravity estimations prevents identification of the partial equilibrium effects of such country-specific transportation and infrastructure improvements, which often are subject to policy debate. We view the development of proper methods to identify the partial equilibrium impact of such improvements on bilateral trade as an important contribution to the academic and policy literature.

We also see at least two opportunities for collaboration on the GE side. First, we note that in almost all international trade policy experiments that are known to us, trade liberalization between two partners, e.g. a FTA between UK and US, only has direct/partial effects on trade between the liberalizing countries. In other words, the initial ‘shock’ to the vector of bilateral

trade costs in response to the formation of the agreement is limited to member countries only. This is a standard and well justified choice when bilateral trade policies are analyzed. However, such an approach may not necessarily be valid when applied to study the impact of improvement in bilateral transportation channels where, for example, a new railway between two countries will also directly affect the shipping costs of many other countries in the region. Proper modeling of these relationships presents a challenge to both trade and transportation economists. Allen and Arkolakis (2014) offer a notable effort in this direction. These authors use the gravity framework to characterize the spatial distribution of economic activity with emphasis on the link to geography. The framework is used to estimate the topography of trade costs, productivities and amenities in the United States and to quantify the welfare effects of the construction of the interstate highway system in US.

Another area that can benefit from collaboration between trade and transportation economists is to disentangle and quantify the effects of improvements in transportation and communication from the effects of targeted trade policies. As discussed earlier in this section, only recently have trade economists been able to capture the effects of globalization with the empirical estimating gravity model. However, the measures of globalization that have been used in the trade literature are remote proxies for the transportation and the communication improvements at best. Borchert and Yotov (2016) offer reduced-form evidence that the country-specific gravity estimates of the effects of globalization from gravity estimations are correlated to the ratio of air-to-rail transportation as a proxy for the shift towards higher value-to-weight goods in a country's export bundle. However, this is only modest preliminary evidence that transportation is indeed an important driver of globalization. We expect that, in combination with the solid GE tools from the trade literature, further efforts to properly measure the impact of transportation improvements within the gravity framework will generate novel insights about the welfare and inequality implications of international trade and trade policy.

## References

- Abrams, R. (1980). International trade flows under flexible exchange rates. *Federal Reserve Bank of St. Louis Economic Review*, 65:3–10.
- Agnosteva, D., Anderson, J., and Yotov, Y. (2014). Intra-national trade costs: Measurement and aggregation. *NBER Working Paper No. 19872*.
- Aitken, N. (1973). The effect of the eec and efta on european trade: A temporal cross-section analysis. *American Economic Review*, 63(5):881–892.
- Ali, S. and Dadush, U. (2011). Trade in intermediates and economic policy. *VoxEU*, 9 February 2011.
- Allen, T. and Arkolakis, C. (2014). Trade and the topography of the spatial economy. *Quarterly Journal of Economics*, 129(3):1085–1140.
- Anderson, J. (1979). A theoretical foundation for the gravity equation. *American Economic Review*, 69(1):106–116.
- Anderson, J. (2011). The gravity model. *Annual Review of Economics*, 3:133–160.
- Anderson, J., Larch, M., and Yotov, Y. (2015a). Growth and trade with frictions: A structural estimation framework. *NBER Working Paper No. 21377*.
- Anderson, J., Larch, M., and Yotov, Y. V. (2015b). *Estimating general equilibrium trade policy effects: Ge ppml*. *CESifo Working Paper No. 5592*.
- Anderson, J. and van Wincoop, E. (2001). Borders, trade, and welfare. *Brookings Trade Forum*, pages 207–230.
- Anderson, J. and van Wincoop, E. (2003). Gravity with gravitas: A solution to the border puzzle. *American Economic Review*, 93(1):170–192.
- Anderson, J. and van Wincoop, E. (2004). Trade costs. *Journal of Economic Literature*, 42(3):691–751.
- Anderson, J., Vesselovsky, M., and Yotov, Y. (2016). Gravity with scale economies. *Journal of International Economics*, 100:174–193.
- Anderson, J. and Yotov, Y. (2010). The changing incidence of geography. *American Economic Review*, 100(5):2157–2186.
- Anderson, J. E. and Yotov, Y. V. (2016). Terms of Trade and Global Efficiency Effects of Free Trade Agreements, 1990-2002. *Journal of International Economics*, 99(C):279–298.
- Arkolakis, C., Costinot, A., and Rodríguez-Clare, A. (2012). New trade models, same old gains? *American Economic Review*, 102(1):94–130.
- Arkolakis, C., Klenow, P., Demidova, S., and Rodriguez-Clare, A. (2008). Endogenous varieties and the gains from trade. *American Economics Review Papers and Proceedings*, 98:444–450.
- Arkolakis, C., Ramondo, N., Rodríguez-Clare, A., and Yeaple, S. (2013). Innovation and Production in the Global Economy. *NBER Working Paper No. 18972*.
- Armington, P. (1969). A theory of demand for products distinguished by place of production. *IMF Staff Papers*, 16:159–176.

- Baier, S. and Bergstrand, J. (2001). The growth of world trade: Tariffs, transport costs, and income similarity. *Journal of International Economics*, 53(1):1–27.
- Baier, S. and Bergstrand, J. (2002). On the endogeneity of international trade flows and free trade agreements. unpublished manuscript.
- Baier, S. and Bergstrand, J. (2004). The economic determinants of free trade agreements. *Journal of International Economics*, 64(1):29–63.
- Baier, S. and Bergstrand, J. (2007). Do free trade agreements actually increase members’ international trade? *Journal of International Economics*, 71(1):72–95.
- Baier, S., Bergstrand, J., and Clance, M. (2015). Heterogeneous economic integration agreements effects. *CESifo Working Paper*, 2015.
- Baier, S., Bergstrand, J., and Feng, M. (2014). Economic integration agreements and the margins of international trade. *Journal of International Economics*, 2:339–350.
- Baier, S. and Garg, M. (2016). Estimating the impact of trade agreements in the presence of sample selection and firm heterogeneity: A panel data approach. *Manuscript*.
- Baier, S., Yotov, Y., and Zylkin, T. (2016). On the widely differing effects of free trade agreements: Lessons from twenty years of trade integration. *School of Economics Working Paper Series*, 2016-15.
- Baldwin, R. and Taglioni, D. (2006). Gravity for dummies and dummies for gravity equations. *NBER Working Paper No. 12516*.
- Behrens, K., Mion, G., Murata, Y., and Südekum, J. (2014). Trade, Wages, And Productivity. *International Economic Review*, 55:1305–1348.
- Benedictis, L. D. and Taglioni, D. (2011). The gravity model in international trade. *The Trade Impact of European Union Preferential Policies*, pages 55–89.
- Bergeijk, P. and Brakman, S. (2009). *The Gravity Model in International Trade*. Cambridge University Press.
- Bergstrand, J. (1985). The gravity equation in international trade: Some microeconomic foundations and empirical evidence. *Review of Economics and Statistics*, 67(3):474–481.
- Bergstrand, J. (1989). The generalized gravity equation, monopolistic competition, and the factor-proportions theory of trade. *Review of Economics and Statistics*, 71(1):143–53.
- Bergstrand, J. and Egger, P. (2011). Gravity equations and economic frictions in the world economy. In Bernhofen, D., Falvey, R., Greenaway, D., and Kreckemeier, U., editors, *Palgrave Handbook of International Trade*. Palgrave Macmillan, New York.
- Bergstrand, J., Egger, P., and Larch, M. (2016). Economic determinants of the *Timing* of preferential trade agreement formations and enlargements. *Economic Inquiry*, 54(1):315–341.
- Bergstrand, J., Larch, M., and Yotov, Y. (2015). Economic integration agreements, border effects, and distance elasticities in the gravity equation. *European Economic Review*, 78:307–327.
- Borchert, I. and Yotov, Y. V. (2016). Globalization and International Trade. *Manuscript*.
- Boulhol, H. and de Serres, A. (2010). Have Developed Countries Escaped the Curse of Distance? *Journal of Economic Geography*, 10(1):113–139.

- Brada, J. and Mendez, J. (1985). Economic integration among developed, developing and centrally planned economies: A comparative analysis. *Review of Economics and Statistics*, 85:549–556.
- Brun, J.-F., Carrère, C., Guillaumont, P., and de Melo, J. (2005). Has Distance Died? Evidence from a Panel Gravity Model. *World Bank Economic Review*, 19(1):99–120.
- Buch, C. M., Kleinert, J., and Toubal, F. (2004). The Distance Puzzle: On the Interpretation of the Distance Coefficient in Gravity Equations. *Economics Letters*, 83(3):293–298.
- Caliendo, L., Feenstra, R., Romalis, J., and Taylor, A. (2015). Tariff reductions, entry, and welfare: Theory and evidence for the last two decades. *NBER Working Paper 21768*.
- Caliendo, L. and Parro, F. (2015). Estimates of the trade and welfare effects of nafta. *Review of Economic Studies*, 82:1–44.
- Cameron, A. and Trivedi, P. (2005). *Microeconometrics - Methods and Applications*. Cambridge University Press, Cambridge, United Kingdom.
- Carrère, C. and Schiff, M. (2005). On the Geography of Trade. Distance is Alive and Well. *Revue économique*, 56(6):1249–1274.
- Coe, D., Subramanian, A., Tamirisa, N., and Bhavnani, R. (2002). The missing globalization puzzle. *IMF Working Paper No. 171*.
- Combes, P., Mayer, T., and Thisse, J. (2008). *Economic Geography*. Princeton University Press.
- Costinot, A., Donaldson, D., and Komunjer, I. (2012). What goods do countries trade? a quantitative exploration of ricardo’s ideas. *Review of Economic Studies*, 79(2):581–608.
- Costinot, A. and Rodríguez-Clare, A. (2014). Trade theory with numbers: Quantifying the consequences of globalization. Chapter 4 in the *Handbook of International Economics Vol. 4*, eds. Gita Gopinath, Elhanan Helpman, and Kenneth S. Rogoff, Elsevier Ltd., Oxford.
- Dai, M., Yotov, Y. V., and Zylkin, T. (2014). On the Trade-diversion Effects of Free Trade Agreements. *Economics Letters*, 122(2):321–325.
- Deardorff, A. (1998). Determinants of Bilateral Trade: Does Gravity Work in a Neoclassical World? In *The Regionalization of the World Economy*, NBER Chapters, pages 7–32. National Bureau of Economic Research, Inc.
- Deb, P., Manning, W., and Norton, E. (2013). Modeling health care costs and counts. *iHEA World Congress in Sydney, Australia*.
- Dekle, R., Eaton, J., and Kortum, S. (2007). Unbalanced Trade. *American Economic Review: Papers and Proceedings*, 97:351–355.
- Dekle, R., Eaton, J., and Kortum, S. (2008). Global Rebalancing with Gravity: Measuring the Burden of Adjustment. *IMF Staff Papers*, 55(3):511–540.
- Disdier, A. and Head, K. (2008). The puzzling persistence of the distance effect on bilateral trade. *Review of Economics and Statistics*, 90(1):37–48.
- Donaldson, D. (2015). The gains from market access. *Annual Review of Economics*, 7:619–647.
- Donaldson, D. (2016). Railroads of the raj: Estimating the impact of transportation infrastructure. *American Economic Review*, forthcoming.



- Dornbusch, R., Fischer, S., and Samuelson, P. (1977). Comparative advantage, trade, and payments in a ricardian model with a continuum of goods. *American Economic Review*, 67(5):823–839.
- Eaton, J. and Kortum, S. (2001). Trade in capital goods. *European Economic Review*, 45(7):1195–1235.
- Eaton, J. and Kortum, S. (2002). Technology, geography and trade. *Econometrica*, 70(5):1741–1779.
- Eaton, J., Kortum, S., Neiman, B., and Romalis, J. (2016). Trade and the global recession. *American Economic Review*, 106(11):3401–38.
- Eaton, J., Kortum, S., and Sotelo, S. (2013). International trade: Linking micro and macro. *in: Acemoglu, D., M. Arellano and E. Dekel (eds.), "Advances in Economics and Econometrics - Tenth World Congress", Volume II, Applied Economics, Cambridge: Cambridge University Press*, pages 329–370.
- Egger, P. and Larch, M. (2011). An assessment of the europe agreements' effects on bilateral trade, gdp, and welfare. *European Economic Review*, 55(2):263–279.
- Egger, P., Larch, M., Staub, K., and Winkelmann, R. (2011). The trade effects of endogenous preferential trade agreements. *American Economic Journal: Economic Policy*, 3(3):113–143.
- Egger, P. and Nigai, S. (2015). Structural gravity with dummies only. *CEPR Discussion Paper No. DP10427*.
- Egger, P. and Staub, K. (2014). Glm estimation of trade gravity models with fixed effects. unpublished manuscript.
- Feenstra, R. (2004). *Advanced International Trade: Theory and Evidence*. Princeton University Press, Princeton, New Jersey.
- Feenstra, R. (2015). *Advanced International Trade: Theory and Evidence*. Princeton University Press, Princeton, New Jersey.
- Felbermayr, G., Heid, B., Larch, M., and Yalcin, E. (2015). Macroeconomic potentials of transatlantic free trade: A high resolution perspective for europe and the world. *Economic Policy*, 83(3):491–537.
- Fieler, A. C. (2011). Nonhomotheticity and Bilateral Trade: Evidence and a Quantitative Explanation. *Econometrica*, 79(4):1069–1101.
- Frankel, J. (1997). *Regional Trading Blocs*. Institute for International Economics, Washington, DC.
- Frankel, J., Stein, E., and Wei, S. (1995). Regional trading blocs in the americas: The natural, the unnatural, and the super-natural. *Journal of Development Economics*, 47:61–95.
- Grossman, G. and Rossi-Hansberg, E. (2012). Task trade between similar countries. *Econometrica*, 80:593–629.
- Head, K. and Mayer, T. (2014). Gravity equations: Workhorse, toolkit, and cookbook. Chapter 3 in the *Handbook of International Economics Vol. 4*, eds. Gita Gopinath, Elhanan Helpman, and Kenneth S. Rogoff, Elsevier Ltd., Oxford.
- Head, K. and Ries, J. (2001). Increasing returns versus national product differentiation as an explanation for the pattern of u.s.-canada trade. *American Economic Review*, 91(4):858–876.
- Heid, B. (2015). Regional trade agreements, unemployment, and the informal sector. *CESifo Area Conference on Global Economy*.

- Heid, B. and Larch, M. (2016). Gravity with unemployment. *Journal of International Economics*, 101:70–85.
- Heid, B., Larch, M., and Yotov, Y. V. (2015). A simple method to estimate the effects of non-discriminatory trade policy within structural gravity models. *Manuscript*.
- Helpman, E. (1981). International trade in the presence of product differentiation, economies of scale and monopolistic competition: A chamberlin-heckscher-ohlin approach. *Journal of International Economics*, 11(3):305–340.
- Helpman, E. and Krugman, P. (1985). *Market Structure and Foreign Trade - Increasing Returns, Imperfect Competition and the International Economy*. The MIT Press, Cambridge, Massachusetts.
- Helpman, E., Melitz, M., and Rubinstein, Y. (2008). Trading partners and trading volumes. *Quarterly Journal of Economics*, 123(2):441–487.
- Hummels, D. (2001). Toward a geography of trade costs. unpublished manuscript.
- Jacks, D., Meissner, C., and Novy, D. (2011). Trade booms, trade busts, and trade costs. *Journal of International Economics*, 83(2):185–201.
- Johnson, R. and Noguera, G. (2012). Accounting for intermediates: Production sharing and trade in value added. *Journal of International Economics*, 86(2):224–236.
- Kraemer, K. and Dedrick, J. (2002). Dell computer: Organization of a global production network. Unpublished manuscript.
- Krugman, P. (1979). Increasing returns, monopolistic competition and international trade. *Journal of International Economics*, 9(4):469–479.
- Krugman, P. (1980). Scale economies, product differentiation, and the pattern of trade. *American Economic Review*, 70(5):950–959.
- Lai, H. and Treffer, D. (2002). The gains from trade with monopolistic competition: Specification, estimation, and mis-specification. *NBER Working Paper No. 9169*.
- Larch, M., Norbäck, P.-J., Sirries, S., and Urban, D. (2016). Heterogeneous firms, globalization and the distance puzzle. *2016, The World Economy*, forthcoming.
- Larch, M. and Wanner, J. (2014). Carbon tariffs: An analysis of the trade, welfare and emission effects. *CESifo Working Paper No. 4598*.
- Larch, M. and Yotov, Y. (2016). General Equilibrium Trade Policy Analysis with Structural Gravity. Wto working paper ersd-2016-08, World Trade Organization.
- Leamer, E. and Levinsohn, J. (1995). International trade theory: The evidence. Chapter 26 in the *Handbook of International Economics Vol. 3*, eds. Gene M. Grossman and Kenneth S. Rogoff, Elsevier Ltd., Oxford, 3:1339–1394.
- Lin, F. and Sim, N. C. (2012). Death of Distance and the Distance Puzzle. *Economics Letters*, 116(2):225–228.
- Magee, C. (2003). Endogenous preferential trade agreements: An empirical analysis. *Contributions to Economic Analysis & Policy*, 2:Articel 15.

- Manning, W. and Mullahy, J. (2001). Estimating log models: To transform or not to transform. *Journal of Health Economics*, 20:461–494.
- Martínez-Zarzoso, I. (2013). The log of gravity revisited. *Applied Economics*, 45:311–327.
- Melitz, M. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica*, 71(6):1695–1725.
- Melitz, M. J. and Redding, S. J. (2015). New Trade Models, New Welfare Implications. *American Economic Review*, 105(3):1105–46.
- Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica*, 46:69–85.
- Novy, D. (2013a). Gravity Redux: Measuring International Trade Costs With Panel Data. *Economic Inquiry*, 51(1):101–121.
- Novy, D. (2013b). International trade without ces: Estimating translog gravity. *Journal of International Economics*, 89(2):271–282.
- Olivero, M. and Yotov, Y. (2012). Dynamic gravity: Endogenous country size and asset accumulation. *Canadian Journal of Economics*, 45(1):64–92.
- Ossa, R. (2011). A ‘new trade’ theory of gatt/wto negotiations. *Journal of Political Economy*, 119(1):122–152.
- Piermartini, R. and Yotov, Y. V. (2016). Estimating trade policy effects with structural gravity. *WTO Working Paper ERSD-2016-10*.
- Ravenstein, E. (1885). The laws of migration: Part 1. *Journal of the Statistical Society of London*, 48(2):167–235.
- Redding, S. (2011). Theories of heterogeneous firms and trade. *Annual Review of Economics*, 3.
- Redding, S. and Venables, A. (2004). Economic geography and international inequality. *Journal of International Economic*, 62(1):53–82.
- Romalis, J. (2007). Nafta’s and custfa’s impact on international trade. *Review of Economics and Statistics*, 89(3):416–435.
- Samuelson, P. (1952). The transfer problem and transport costs: The terms of trade when impediments are absent. *Economic Journal*, 62(246):278–304.
- Santos Silva, J. and Tenreyro, S. (2006). The log of gravity. *Review of Economics and Statistics*, 88(4):641–658.
- Santos Silva, J. and Tenreyro, S. (2008). Comments on “the log of gravity”. *Manuscript*.
- Santos Silva, J. and Tenreyro, S. (2011). Further simulation evidence on the performance of the poisson pseudo-maximum likelihood estimator. *Economics Letters*, 112(2):220–222.
- Santos Silva, J. and Tenreyro, S. (2015). Trading partners and trading volumes: Implementing the helpman-melitz-rubinstein model empirically. *Oxford Bulletin of Economics and Statistics*, 77:93–105.
- Soloaga, I. and Winters, L. (2001). Regionalism in the nineties: What effect on trade? *North American Journal of Economics and Finance*, 12(1):1–29.

- Tinbergen, J. (1962). *Shaping the World Economy: Suggestions for an International Economic Policy*. The Twentieth Century Fund, New York.
- Trefler, D. (1993). Trade liberalization and the theory of endogenous protection: An econometric study of u.s. import policy. *Journal of Political Economy*, 101(1):138–160.
- Varian, H. (2007). An ipod has global value. ask the (many) countries that make it. *The New York Times* (June 28, 2007).
- Wei, S.-J. (1996). Intra-National versus International Trade: How Stubborn are Nations in Global Integration? *NBER Working Papers No 5531*.
- Wolf, H. C. (2000). Intranational home bias in trade. *The Review of Economics and Statistics*, 82(4):555–563.
- Wooldridge, J. (2009a). Correlated random effects models with unbalanced panels. *Manuscript: Michigan State University*.
- Wooldridge, J. (2009b). *Introductory Econometrics: A Modern Approach*. Cengage Learning Services, Florence, KY, 4th edition.
- Wooldridge, J. (2010). *Econometric Analysis of Cross Section and Panel Data*. The MIT Press, Cambridge, Massachusetts, 2nd edition.
- Yotov, Y. V. (2012). A simple solution to the distance puzzle in international trade. *Economics Letters*, 117(3):794–798.
- Yotov, Y. V., Piermartini, R., Monteiro, J.-A., and Larch, M. (2016). *An Advanced Guide to Trade Policy Analysis: The Structural Gravity Model*. UNCTAD and WTO, Geneva.

Table 1: Log-Linear Structural Gravity

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	1986	1990	1994	1998	2002	2006	Pooled OLS
lnDIST	-1.260*** (0.0468)	-1.210*** (0.0424)	-1.286*** (0.0479)	-1.220*** (0.0429)	-1.138*** (0.0453)	-1.121*** (0.0470)	-1.203*** (0.0185)
CNTG	0.208 (0.175)	0.289 (0.160)	0.0578 (0.186)	0.182 (0.171)	0.282 (0.168)	0.239 (0.170)	0.205** (0.0713)
LANG	0.580*** (0.0960)	0.593*** (0.0929)	0.713*** (0.0923)	0.735*** (0.0830)	0.663*** (0.0792)	0.672*** (0.0856)	0.669*** (0.0360)
CLNY	0.895*** (0.146)	0.723*** (0.137)	0.592*** (0.132)	0.556*** (0.126)	0.521*** (0.119)	0.539*** (0.123)	0.657*** (0.0532)
FTA	-0.625*** (0.111)	-0.466*** (0.0973)	-0.453*** (0.0995)	0.0338 (0.0809)	0.326*** (0.0810)	0.428*** (0.0795)	0.0136 (0.0365)
Exporter FE	Yes	Yes	Yes	Yes	Yes	Yes	No
Importer FE	Yes	Yes	Yes	Yes	Yes	Yes	No
Exp-Year FE	No	No	No	Yes	No	No	Yes
Imp-Year FE	No	No	No	Yes	No	No	Yes
N	3782	3996	4178	4326	4406	4448	25136

Robust standard errors in parentheses

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

Table 2: Log-Linear Structural Gravity: Pair Fixed Effects

	(1)	(2)	(3)	(4)	(5)
	No Lag	One Lag	Two Lags	Lags and Lead	CRE
FTA	0.123*** (0.0362)	-0.0191 (0.0396)	-0.0532 (0.0420)	0.0509 (0.0548)	0.123*** (0.0408)
FTA_LAG4		0.294*** (0.0403)	0.168*** (0.0423)	0.242*** (0.0526)	
FTA_LAG8			0.230*** (0.0444)	0.240*** (0.0582)	
FTA_LEAD4				-0.0592 (0.0505)	
lnDIST					-1.159*** (0.0338)
CNTG					0.262* (0.114)
LANG					0.493*** (0.0586)
CLNY					0.730*** (0.111)
$mean(FTA_{ij})$					-0.0157 (0.104)
Pair FE	Yes	Yes	Yes	Yes	No
Exp-Year FE	Yes	Yes	Yes	Yes	Yes
Imp-year FE	Yes	Yes	Yes	Yes	Yes
Mean-Exp-Year FE	No	No	No	No	Yes
Mean-Imp-year FE	No	No	No	No	Yes
r2	0.937	0.945	0.955	0.962	0.937
N	25073	21287	17289	12615	25073

Standard errors in parentheses

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

Table 3: PPML 1986-1994

	(1)	(2)	(3)	(4)	(5)	(6)
	1986	1986	1990	1990	1994	1994
	Trade>0	All Trade	Trade	All Trade	Trade>0	All Trade
lnDIST	-0.729*** (0.0399)	-0.721*** (0.0397)	-0.645*** (0.0370)	-0.645*** (0.0367)	-0.567*** (0.0324)	-0.566*** (0.0324)
CNTG	0.487*** (0.0752)	0.491*** (0.0759)	0.496*** (0.0875)	0.493*** (0.0874)	0.502*** (0.0801)	0.501*** (0.0800)
LANG	0.303*** (0.0680)	0.304*** (0.0684)	0.294*** (0.0679)	0.297*** (0.0680)	0.245*** (0.0658)	0.246*** (0.0658)
CLNY	-0.0374 (0.0877)	-0.0217 (0.0873)	0.00510 (0.0889)	0.00564 (0.0888)	0.0405 (0.0889)	0.0412 (0.0889)
FTA	0.546*** (0.119)	0.556*** (0.118)	0.614*** (0.0944)	0.615*** (0.0938)	0.907*** (0.0834)	0.910*** (0.0834)
Constant	9.432*** (0.470)	9.360*** (0.469)	9.095*** (0.419)	9.090*** (0.417)	9.977*** (0.460)	9.968*** (0.460)
RESET	0.06	0.14	0.66	0.54	0.93	4556
$\lambda$	1.47 (0.05)	1.56 (0.05)	1.45 (0.07)	1.52 (0.07)	1.44 (0.07)	1.49 (0.08)
N	3782	4556	3996	4556	4178	4556
Exporter FE	Yes	Yes	Yes	Yes	Yes	Yes
Importer FE	Yes	Yes	Yes	Yes	Yes	Yes

$\lambda$  is coefficient estimated from the Park test

P-value of the Ramsey-RESET reported

Standard errors in parentheses

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

Table 4: PPML 1998-2006

	(1)	(2)	(3)	(4)	(5)	(6)
	1998	1998	2002	2002	2006	2006
	Trade>0	All Trade	Trade	All Trade	Trade>0	All Trade
lnDIST	-0.562*** (0.0341)	-0.562*** (0.0340)	-0.597*** (0.0349)	-0.597*** (0.0349)	-0.751*** (0.0360)	-0.751*** (0.0360)
CNTG	0.484*** (0.0738)	0.484*** (0.0738)	0.412*** (0.0697)	0.412*** (0.0697)	0.335*** (0.0658)	0.335*** (0.0658)
LANG	0.180** (0.0640)	0.180** (0.0640)	0.189** (0.0661)	0.189** (0.0661)	0.120 (0.0694)	0.120 (0.0694)
CLNY	0.0648 (0.0911)	0.0649 (0.0911)	-0.00443 (0.0921)	-0.00425 (0.0921)	-0.0547 (0.0982)	-0.0547 (0.0982)
FTA	1.022*** (0.0791)	1.023*** (0.0790)	0.871*** (0.0773)	0.872*** (0.0773)	0.411*** (0.0735)	0.411*** (0.0735)
Constant	10.14*** (0.436)	10.14*** (0.436)	9.111*** (0.448)	9.108*** (0.448)	12.07*** (0.503)	12.07*** (0.503)
RESET	0.92	0.97	0.84	0.87	0.90	0.89
$\lambda$	1.39 (0.05)	1.40 (0.05)	1.47 (0.05)	1.49 (0.06)	1.53 (0.04)	1.54 (0.04)
N	4326	4556	4406	4556	4448	4556
Exporter FE	Yes	Yes	Yes	Yes	Yes	Yes
Importer FE	Yes	Yes	Yes	Yes	Yes	Yes

$\lambda$  is coefficient estimated from the Park test

P-value of the Ramsey-RESET reported

Standard errors in parentheses

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$



Table 5: PPML with Internal Trade 1986-1994

	(1)	(2)	(3)	(4)	(5)	(6)
	1986	1986	1990	1990	1994	1994
	Trade>0	All Trade	Trade	All Trade	Trade>0	All Trade
lnDIST	-0.607*** (0.0902)	-0.606*** (0.0905)	-0.448*** (0.0656)	-0.453*** (0.0657)	-0.499*** (0.0676)	-0.500*** (0.0676)
CNTG	0.563*** (0.126)	0.549*** (0.126)	0.665*** (0.130)	0.659*** (0.130)	0.698*** (0.118)	0.695*** (0.118)
LANG	0.363** (0.125)	0.382** (0.128)	0.171 (0.0987)	0.174 (0.0988)	0.321** (0.105)	0.323** (0.105)
CLNY	0.271 (0.157)	0.287 (0.159)	0.405** (0.125)	0.410** (0.125)	0.167 (0.120)	0.168 (0.120)
INTERNAL	3.928*** (0.257)	3.962*** (0.260)	4.093*** (0.204)	4.087*** (0.204)	3.697*** (0.203)	3.696*** (0.203)
FTA	0.662*** (0.166)	0.693*** (0.171)	0.874*** (0.126)	0.876*** (0.126)	0.495*** (0.117)	0.496*** (0.117)
Constant	10.75*** (0.824)	10.72*** (0.828)	9.690*** (0.634)	9.725*** (0.634)	10.65*** (0.642)	10.66*** (0.642)
RESET	0.00	0.00	0.00	0.00	0.00	0.00
$\lambda$	1.54 (0.10)	1.64 (0.11)	1.83 (0.07)	1.90 (0.07)	1.64 (0.08)	1.72 (0.07)
N	3850	4624	4064	4624	4246	4624
Exporter FE	Yes	Yes	Yes	Yes	Yes	Yes
Importer FE	Yes	Yes	Yes	Yes	Yes	Yes

$\lambda$  is coefficient estimated from the Park test

P-value of the Ramsey-RESET reported

Standard errors in parentheses

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

Table 6: PPML with Internal Trade 1998-2006

	(1)	(2)	(3)	(4)	(5)	(6)
	1998	1998	2002	2002	2006	2006
	Trade>0	All Trade	Trade	All Trade	Trade>0	All Trade
lnDIST	-0.556*** (0.0592)	-0.557*** (0.0592)	-0.659*** (0.0548)	-0.660*** (0.0548)	-0.702*** (0.0591)	-0.703*** (0.0591)
CNTG	0.672*** (0.114)	0.671*** (0.114)	0.553*** (0.113)	0.552*** (0.113)	0.498*** (0.107)	0.497*** (0.107)
LANG	0.395*** (0.100)	0.395*** (0.100)	0.372*** (0.0986)	0.372*** (0.0986)	0.287** (0.0898)	0.288** (0.0899)
CLNY	0.118 (0.112)	0.119 (0.112)	0.0916 (0.108)	0.0920 (0.108)	0.0710 (0.0993)	0.0711 (0.0993)
INTERNAL	3.398*** (0.182)	3.397*** (0.182)	3.071*** (0.170)	3.071*** (0.170)	2.839*** (0.181)	2.838*** (0.181)
FTA	0.520*** (0.104)	0.520*** (0.104)	0.403*** (0.0958)	0.403*** (0.0959)	0.362*** (0.0966)	0.362*** (0.0966)
Constant	11.22*** (0.596)	11.23*** (0.596)	10.93*** (0.630)	10.93*** (0.630)	11.72*** (0.576)	11.72*** (0.576)
RESET	0.00	0.00	0.00	0.00	0.00	0.00
$\lambda$	1.53 (0.07)	1.58 (0.08)	1.55 (0.07)	1.58 (0.07)	1.43 (0.06)	1.45 (0.07)
N	4394	4624	4474	4624	4516	4624
Exporter FE	Yes	Yes	Yes	Yes	Yes	Yes
Importer FE	Yes	Yes	Yes	Yes	Yes	Yes

$\lambda$  is coefficient estimated from the Park test

P-value of the Ramsey-RESET reported

Standard errors in parentheses

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

Table 7: Structural Gravity with Intra-national Trade Flows

	(1)	(2)	(3)	(4)	(5)	(6)
	'No Internal'	Internal	'No Internal'	Internal	'No Internal'	Internal
FTA	-0.0475 (0.0667)	0.510*** (0.0755)	-0.0793 (0.0696)	0.169* (0.0696)	-0.0458 (0.0627)	0.0632 (0.0643)
FTA_LAG4			0.106* (0.0434)	0.520*** (0.0476)	0.0581 (0.0478)	0.289*** (0.0558)
FTA_LAG8					-0.0488 (0.0337)	0.264*** (0.0368)
N	27084	27492	22565	22905	18036	18308

Standard errors in parentheses

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ 

Table 8: Gravity &amp; Globalization

	(1)	(2)
FTA	0.505*** (0.0760)	0.206** (0.0662)
BRDR-1990		0.222*** (0.0173)
BRDR-1994		0.352*** (0.0222)
BRDR-1998		0.581*** (0.0273)
BRDR-2002		0.602*** (0.0334)
BRDR-2006		0.747*** (0.0354)
N	27684	27684

Standard errors in parentheses

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

Table 9: General Equilibrium Analysis with Structural Gravity

	A. GE Effects of UK-US RTA				B. GE Effects of Globalization			
	$\% \Delta X_{ij}$	$\% \Delta RGDP$	$\% \Delta W$	$\% \Delta IMR$	$\% \Delta X_{ij}$	$\% \Delta RGDP$	$\% \Delta W$	$\% \Delta IMR$
ARG	-0.02	-0.00	-0.01	-0.01	-43.00	-3.06	-1.87	1.23
AUS	-0.06	-0.00	-0.01	-0.00	-55.06	-1.86	1.94	3.87
AUT	-0.01	-0.00	-0.00	-0.00	-38.24	-5.76	-3.57	2.32
BEL	-0.04	-0.00	-0.00	0.00	-40.43	-3.41	-2.34	1.11
BGR	-0.01	-0.00	-0.00	0.00	-48.47	-2.90	0.95	3.97
BOL	-0.04	-0.00	-0.01	-0.01	-57.61	-2.69	3.93	6.80
BRA	-0.03	-0.00	-0.01	-0.01	-37.73	-1.02	-2.44	-1.43
CAN	-0.07	-0.01	-0.02	-0.01	-38.88	-5.85	-1.70	4.41
CHE	-0.02	-0.00	-0.00	-0.00	-40.24	-3.32	-2.52	0.83
CHL	-0.03	-0.00	-0.01	-0.01	-37.40	-2.42	-3.03	-0.62
CHN	-0.02	-0.00	-0.01	-0.01	-25.38	-1.05	-5.27	-4.27
CMR	-0.02	-0.00	-0.00	-0.00	-52.03	-2.95	2.20	5.31
COL	-0.05	-0.00	-0.01	-0.01	-56.68	-1.49	5.16	6.75
CRI	-0.06	-0.00	-0.01	-0.01	-42.37	-3.45	-0.47	3.08
CYP	-0.10	-0.00	0.01	0.01	-59.61	-1.98	5.79	7.92
DEU	-0.03	-0.00	-0.00	0.00	-34.58	-3.66	-3.66	0.00
DNK	-0.03	-0.01	-0.00	0.00	-38.03	-5.85	-3.09	2.93
ECU	-0.05	-0.00	-0.01	-0.01	-60.06	-1.73	7.18	9.07
EGY	-0.04	-0.00	-0.00	-0.00	-55.66	-2.24	2.95	5.31
ESP	-0.04	-0.00	0.00	0.00	-49.00	-2.65	0.35	3.08
FIN	-0.03	-0.00	-0.00	0.00	-37.06	-3.65	-3.37	0.29
FRA	-0.04	-0.00	-0.00	0.00	-41.58	-3.57	-1.90	1.73
GBR	2.30	0.21	0.14	-0.07	-44.06	-3.76	-1.16	2.70
GRC	-0.04	-0.00	0.00	0.00	-57.70	-2.21	3.78	6.13
HKG	-0.02	-0.00	-0.01	-0.01	-46.92	-9.12	-2.94	6.80
HUN	-0.01	-0.00	-0.00	-0.00	-40.15	-4.23	-2.86	1.43
IDN	-0.02	-0.00	-0.01	-0.01	-39.96	-2.21	-3.15	-0.95
IND	-0.04	-0.00	-0.01	-0.00	-44.61	-1.09	-0.57	0.52
IRL	-0.07	-0.02	-0.01	0.01	-27.38	-6.60	-5.21	1.49
IRN	-0.02	-0.00	-0.00	-0.00	-65.96	-0.73	6.41	7.19
ISL	-0.08	-0.01	0.00	0.01	-56.90	-2.22	3.47	5.82
ISR	-0.05	-0.00	-0.01	-0.00	-39.44	-3.07	-2.00	1.11
ITA	-0.03	-0.00	-0.00	0.00	-39.77	-2.75	-2.20	0.56
JOR	-0.05	-0.00	-0.00	-0.00	-59.93	-1.06	4.69	5.81
JPN	-0.03	-0.00	-0.01	-0.01	-31.92	-1.32	-4.65	-3.37
KEN	-0.10	0.00	0.01	0.01	-65.87	-0.75	8.57	9.39
KOR	-0.02	-0.00	-0.01	-0.01	-34.11	-1.92	-4.46	-2.59
KWT	-0.04	-0.00	-0.01	-0.00	-53.09	-2.70	1.36	4.18
LKA	-0.06	-0.00	-0.01	-0.00	-48.67	-1.93	0.08	2.05
MAC	-0.03	-0.01	-0.01	-0.01	-36.83	-7.45	-3.17	4.63
MAR	-0.02	-0.00	0.00	0.00	-52.02	-3.06	1.55	4.75
MEX	-0.06	-0.01	-0.02	-0.01	-39.78	-5.01	-1.43	3.77
MLT	-0.03	-0.01	-0.00	0.01	-39.64	-7.39	-2.50	5.29
MUS	-0.07	-0.01	-0.00	0.01	-43.18	-6.25	-1.32	5.25
MWI	-0.06	-0.01	0.00	0.01	-51.90	-3.08	2.03	5.27
MYS	-0.02	-0.00	-0.01	-0.01	-33.31	-4.57	-4.45	0.13
NER	-0.01	-0.00	-0.00	0.00	-43.58	-7.82	-0.51	7.93
NGA	-0.12	0.00	0.01	0.01	-67.00	-0.33	7.92	8.28
NLD	-0.03	-0.01	-0.00	0.00	-35.34	-6.67	-3.92	2.94
NOR	-0.04	-0.00	-0.00	0.00	-46.93	-3.78	-0.41	3.50
NPL	-0.01	-0.00	-0.01	-0.01	-53.51	-2.41	1.25	3.75
PAN	-0.02	-0.00	-0.01	-0.01	-60.45	-3.21	4.93	8.41
PHL	-0.03	-0.00	-0.01	-0.01	-34.30	-4.50	-4.44	0.06
POL	-0.02	-0.00	-0.00	0.00	-47.85	-2.38	-0.69	1.73
PRT	-0.03	-0.00	0.00	0.00	-45.54	-4.28	-0.49	3.96
QAT	-0.07	-0.00	-0.00	0.00	-62.67	-1.87	5.52	7.53
ROM	-0.02	-0.00	-0.00	-0.00	-53.36	-2.61	1.69	4.41
SEN	-0.03	-0.00	0.00	0.00	-66.78	-1.28	8.05	9.45
SGP	-0.02	-0.00	-0.01	-0.01	-32.51	-6.91	-4.77	2.31
SWE	-0.03	-0.01	-0.00	0.00	-34.37	-5.04	-3.74	1.36
THA	-0.02	-0.00	-0.01	-0.01	-35.85	-3.37	-4.09	-0.75
TTO	-0.07	-0.00	-0.01	-0.01	-36.30	-2.33	-2.20	0.13
TUN	-0.00	0.00	-0.00	-0.00	-47.59	-3.75	-0.37	3.51
TUR	-0.04	-0.00	-0.00	0.00	-50.46	-1.86	1.04	2.95
TZA	-0.04	-0.01	0.00	0.01	-50.72	-4.03	3.56	7.91
URY	-0.02	-0.00	-0.01	-0.01	-46.11	-3.53	-1.17	2.45
USA	1.05	0.03	0.00	-0.03	-50.88	-1.63	1.12	2.80
ZAF	-0.06	-0.00	-0.00	0.00	-46.38	-1.93	-0.57	1.39

Standard errors in parentheses