# MASTERARBEIT / MASTER'S THESIS

Title der Masterarbeit / Title of the Master's Thesis

"Estimating the Effect of the Euro on intra-European Trade using a new Distance Metric"


Verfasst von / submitted by

Boris Knapp, B.Sc.


Angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (M.Sc.)

Wien, 2017 / Vienna 2017

| | |
|---|---|
| Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet: | A 066 913 |
| Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet: | Masterstudium Volkswirtschaftslehre |
| Betreut von / Supervisor: | Univ.-Prof. Dipl.-Ing. Dr. Robert M. Kunst |

# 1. The Euro as part of the European project (Introduction)

An exhaustive discussion about the positive and negative effects of the Euro and the prospects of the EMU is obviously far beyond the scope of this paper. It is, however, important to keep in mind the bigger picture whenever examining a specific problem.

The introduction of the Euro was not based on economic considerations alone but was mostly politically motivated. The European Coal and Steel Community (ECSC), the predecessor of the EU, was founded in the 1950s as a way to prevent further war among European countries and the peacekeeping aspect has been central to the idea of European integration ever since. The process of European integration is an iterative one and establishing a common currency has only been one of the steps towards a supranational European state. This is not an argument to mitigate the importance of the Euro's economic performance – on the contrary. The very fact that the European project started out as a purely economic partnership shows the importance that is attributed to prosperity for maintaining peace and cooperation. The future of the EU is undoubtedly tied to the economic welfare that its institutions – of which the single currency is the one discussed here– are able to create.

In theory the economic stimulus of a monetary union is straight forward. Price stability benefits both importers and exporters and the elimination of multiple national currencies makes doing business internationally generally less complicated. The resulting increase in trade stimulates economic growth and ultimately leads to higher welfare. This is contrasted with potential drawbacks that come with sharing not only a single currency but also being subject to a single monetary policy. The main problem arises from the fact that, while monetary policy is determined on a supranational level by the ECB, fiscal policymaking remains the responsibility of national parliaments. This inability to balance monetary policy with different fiscal policies and economic conditions of different countries is often referred to as the issue of *one-size-fits-all monetary policy*.

De Grauwe (2013) argues that a single interest rate imposed on all member countries is too low for the booming countries and too high for the countries in recession, thus exacerbating booms and busts at the national level.

There are two ways around this problem. The first one is fiscal policymaking on a supranational level. This would require a substantial transfer of competences from a national to a European level which would mean a massive loss of sovereignty for national states. The treaties of the European Union would not only have to be adopted but also approved by all

EMU members. Democratic legitimation of a structural change of that extent is out of reach at the moment and probably in the near future as well. The second way for this issue to be resolved is through economic convergence. By meeting the Maastricht criteria and due to their interconnectedness, the member states are supposed to converge with respect to their cyclical behavior (short run) and trending development (long run).

Evidence on whether economies do converge is mixed and opinions about the construction of the Euro as a supranational currency differ. In a pre-crisis paper Lane (2006) attributes the Euro macroeconomic divergent forces, due to the fact that the very entry into the Euro was a much bigger structural shock for "peripheral" member countries than for "core" member countries. In particular he argues that the relative decline in interest rates has been much greater for the countries on the periphery. It has improved the long-term credit environment but also generating rapid growth in lending and local housing booms in the favored countries and ultimately build up inflationary pressures. However, Lane continues by arguing that the monetary policy under the ECB as response to the significant common shocks experienced by the member countries to has been superior to what would have prevailed under non-coordinated monetary policies.

More recently, Lane (2012) identifies three phases in the relationship between the Euro and the European sovereign crisis: Increased fiscal risk during the pre-crisis period, amplified fiscal impact once the crisis occurred and prolonged duration of the post-crisis recovery period. Many agree that the institutional framework of the EMU lacks stability during periods of crisis. Dabrowski (2010) attributes the tensions that have appeared since the late summer of 2008 to the incomplete nature of the EU project, its various asymmetries, and particularly to the fact that economic integration has advanced much faster than political integration. De Grauwe (2013) identifies as one of the main design flaws the discontinuation of the central bank's role of lender of last resort for government bonds. Although efforts have been made to introduce alternative stabilizing mechanisms, such as *The European Stability Mechanism* (ESM) and the "Outright Money Transactions" program, they lack the stabilizing power of a central bank at a national level.

For Stockhammer (2011) Germany's policy of aggressive wage restraint is at the root of the crisis of the Euro system and an active coordination of wage policy has to be pursued. While Lapavitsas et al (2010) and Höpner (2013) agree that divergence in competitiveness is a central issue with the EMU, Jones (2009) argues that while labour costs diverged rather drastically from 2000 to 2008, considering a longer timespan provides a more complete description of changes in relative economic competitiveness. By choosing 1990 instead of 2000 as the base year, he shows that, when taking into account the developments prior to the introduction of the Euro, the discrepancies turn out to be far less severe. Jones concludes that the EMU is much better than the system of fixed but adjustable rates it replaced and that instead of engendering conflicts, the single currency has diffused them.

With complex constructs as the EMU issues like the ones mentioned above cannot be avoided. It is therefore important to be able to put them into perspective with regard to their economic impact. Increased trade is a major aspect of monetary unions in general and the Euro area in particular and wealth generating effects can potentially offset negative consequences. The model of economic gravity provides a powerful toolkit for analyzing international trade and has been adopted and extended over the years to address specific questions. In this master's thesis I am going to estimate the gravity equation on a panel of European countries in order to determine the impact of the Euro on intra-European trade. The distinguishing features of my study are the introduction of a new distance metric and the fact that the panel is fully balanced. I am therefore able to determine the bias due to unbalanced data sets by artificially generating missing values.

This paper aims to fulfill three purposes: Firstly, it adds to the literature about the trade stimulating effect of the Euro. Many studies have already suggested that the effect of the Euro is rather small and therefore differs vastly from those of other currency unions. Further replication of similar empirical findings will add to the credibility of those studies.

Secondly, and this is the main purpose of the paper, it introduces a new metric for measuring distances between countries. Attempts to improve on the distance between capitals as a suitable metric have already taken place in the literature. Most of these improved measures rely on the distribution of population, while taking economic distribution into account is less common.[1] A major reason for that might poorer data availability of spatial distribution of economic activity compared to that of population. The distance metric defined in this thesis paper will utilize the *Classification of Territorial Units for Statistics* as defined by the Eurostat and directly take into account the spatial distribution of economic activity.

Thirdly, because I use a high quality, balanced panel data set, I am able to determine the bias due to missing observations and mistreatment thereof. Previous studies have suggested that it is severe and that careful handling of unbalanced panels is required.

---

[1] One paper that utilized spatial economic distribution to define a distance metric is Head and Mayer (2002), which will be mentioned in section 3.

The rest of the paper is structured as follows:
Section 2 is devoted to the gravity equation. The theoretical foundation as well as some background information is provided. Section 3 describes the data set used here. Special attention is given to the new distance metric. Section 4 discusses different econometric specifications of the gravity equation. Some issues arise from the theory itself while others are data related. In section 5 I present my findings, interpret and discuss the results and perform robustness analyses. Section 6 concludes.

## 2 The Gravity Model

Gravity models of trade have been around for quite some time now and continue to play an important role in empirical trade literature. This is due to their high explanatory power and the fact that much effort has been put into improving them, providing a sound theoretical foundation and successfully addressing methodological and data-related issues.

Tinbergen (1962) is regarded to be the first to interpret the concept of gravity in an economic context and apply it to empirically study international trade patterns. Although reasonable, his approach was rather ad-hoc and not derived from any particular economic model. This lack of theoretical foundation was one of the main criticisms in the early days of gravity literature and prevented it from general acceptance in mainstream economics as textbooks until today pay little attention to economic gravity even though it is widely regarded as the workhorse of empirical trade literature.[2]

Early attempts at building a theoretical basis include Leamer and Stern (1970), Anderson (1979), Bergstrand (1985) and Deardoff (1998), who, assuming CES preferences and/or production functions in a general equilibrium setting, derived equations similar to those used in earlier trade literature, but including additional terms.

Anderson and Van Wincoop (2001) showed that problems arise not because of flawed theoretical foundations, but rather due to missing rigor in applying the theory. In particular they pointed out the importance of so-called *multilateral resistance* terms (MTR) already included in Anderson (1979) and provided a tractable formulation of the gravity equation including MTR. Improving on the theoretical basis as well as bring to attention issues of misspecification, their work can be seen as the beginning of the most recent wave of gravity literature. Some

[2] The first textbook explicitly covering economic gravity is Feenstra (2004).

questionable estimates have successfully been re-estimated following their methodological recommendations.[3]

Since the theoretic foundations are solid and widely accepted among trade economists, recent contributions to the literature focus mainly on methodological improvements. Subsection 2.3 provides further insight in this respect.

## 2.1 Basic Concept

In its most basic form the gravity model of trade resembles the original – Newton's Law of Gravity. In Newtonian physics two objects i and j attract each other with a gravitational force $F_{ij}$ that is equal to the product of their masses $m_i$, $m_j$ and the gravitational constant G, divided by the square of the distance between the two bodies $d_{ij}$.

$$F_{ij} = G\frac{m_i m_j}{d_{ij}^2} \qquad (1)$$

Translated into the context of international trade, attraction is represented by trade ($X_{ij}$), the mass variables become economic masses, i.e. GDPs ($M_{i(j)}$) and the distance stays distance in the physical sense ($D_{ij}$). By introducing $\alpha_1$, $\alpha_2$ and $\alpha_3$, we allow the elasticities to differ from their physical pendants.

$$X_{ij} = A\frac{M_i^{\alpha_1} M_j^{\alpha_2}}{D_{ij}^{\alpha_3}} \qquad (2)$$

The intuition behind economic gravity is that trade flows from i to j depend positively on i's supply of (tradable) goods and j's demand for (tradable) goods – both approximated by GDP – and negatively on trade costs and barriers – approximated by distance. The multiplicative form allows for straightforward log-linearization and application of simple estimation techniques like OLS.

$$\ln(X_{ij}) = \ln(A) + \alpha_1 \ln(M_i) + \alpha_2 \ln(M_j) + \alpha_3 \ln(D_{ij}) \qquad (3)$$

Eq. (3) is in fact the equation that was estimated by Tinbergen (1962) yielding a convincingly good fit and laying the cornerstone for decades worth of literature. Tinbergen himself already

---

[3] : Anderson and van Wincoop (2001) re-estimate the border effect, Rose and van Wincoop (2001) re-estimate the currency union effect

augmented this basic equation with political and semi-economic factors like dummies for neighboring countries and preferential trade agreements, which continue to be included until today and are also part of the equations estimated in this thesis.

## 2.2 Theoretic foundation

One of the most widely accepted theoretic foundations of the gravity equation is provided in Anderson and Van Wincoop (2001). Baldwin and Taglioni (2006)'s extension thereof accounts for the time dimension of most panel data.[4] A structural gravity equation like in Anderson and Van Wincoop (2001) or Baldwin and Taglioni (2006) can be derived from a surprisingly simple setup assuming only CES preferences represented by a Dixit-Stiglitz aggregator function and international prices that differ from domestic prices:

$$P_{ij} = P_i \tau_{ij} \qquad (4)$$

$P_{ij}$ is the price of country i's commodity in country j, $P_i$ is country i's domestic price for its own commodity and $\tau_{ij}$ is the mark-up for country i's commodity shipped to country j. $\tau_{ij}$ is supposed to capture all trade frictions and barriers and is assumed to be a function of cultural (common language or colonial history), political (taxes, quota, trade agreements and currency unions) and geographical (adjacency and bilateral distance) ties and barriers. Taking this into account, the simple optimization problem[5]

$$\max \quad Q_j = \left[\sum_{i=1}^{N} Q_{ij}^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}} \quad (5) \qquad s.t. \sum_{i=1}^{N} P_{ij} Q_{ij} = E_j \quad (6)$$

yields a structural gravity equation equivalent to that of Baldwin and Taglioni.[6]

$$X_{ij} = \frac{\tau_{ij}^{1-\sigma} E_i E_j}{\Omega_i P_j^{1-\sigma}} \qquad (7)$$

[4] Even cross-sectional data has a panel structure as it is double indexed, except when instead of unilateral trade flows (imports or exports) average trade flows (average of imports and exports) are used.
[5] A more detailed derivation can be found in the Appendix.
[6] Baldwin and Taglioni (2006) allow for multiple varieties produced by each nation only to make a simplifying symmetry assumption later. Assuming only one variety per country to begin with yields the same exact equation.

$\Omega_i$ and $P_j^{1-\sigma}$ correspond to Anderson and Van Wincoop (2001)'s *multilateral resistance* terms, while Baldwin and Taglioni (2006) summarize them into a term they call *gravitational un-constant*.

At first sight, this appears to deviate from the analogy with the physical law, but on closer inspection it is an even more accurate resemblance of it. To really understand this, one must be aware that Newton's formula in section 2.1 considers the case of two bodies in otherwise empty space. When no additional forces apply, F as stated in Eq. (1) describes the force of attraction fully – the so called *two-body problem*. With the presence of bodies other than i and j, interactions with all bodies have to be taken into account. The resulting force is then the sum of all forces – the so called *n-body problem*.

The economic reasoning behind MTR is rather straight forward. Although trade friction between two trading partners matters, it is also important how strong it is compared to trade friction with all other potential trading partners because the respective trading partner is only one possible origin or destination of the trade flow. The following example will provide a better insight:

Countries, seeking diversified consumption, will engage in trade preferring trading partners with whom trading costs are small. Two pairs of countries might have similar costs of trading as approximated by distance. Trading costs between New Zealand and Australia might be similar to those between Portugal and Sweden. Those trading costs might be substantial, but in the case of Australia and New Zealand they are small in comparison to trading costs with all other potential trading partners. Sweden and Portugal on the other hand each have many closer, more attractive trading partners and therefore their mutual trading costs are high in comparison. The main point here is that trade between two countries is determined not only by trading costs between them, but also by trading costs between them and all of their potential trading partners, hence, *multilateral trade resistance*.

If we assume a functional from of bilateral trade resistance as in Eq. (8)[7]

$$\tau_{ij} = D_{ij}^{\overline{\alpha}_3} e^{(EU^{\overline{\alpha}_4} + CU^{\overline{\alpha}_5} + ADJ^{\overline{\alpha}_6} + COML^{\overline{\alpha}_7})} \qquad (8)$$

[7] $\tau_{ij}$ can in principle include any geographical, cultural and political variables of interest. In fact, any variable that is assumed to have an influence on bilateral trade resistance. This specification corresponds to the one estimated in section 5 of this paper.

and further include a time index to account for the time dimension of our panel, a stochastic version of Eq. (7) is:

$$x_{ij,t} = \alpha_0 + \alpha_1 y_{i,t} + \alpha_2 e_{j,t} + \alpha_3 d_{ij} + \alpha_4 eu + \alpha_5 cu + \alpha_6 adj + \alpha_7 coml - \omega_{i,t} + (\sigma - 1)p_{j,t} + \varepsilon_{ij,t} \qquad (9)$$

Small letters represent natural logarithms of the corresponding variables in capital letters, $\alpha_3$ to $\alpha_7$ reflect elasticities of substitution according to $\alpha_* = (1-\sigma)\bar{\alpha}_*$ and $\varepsilon_{ij,t}$ is the usual, normally distributed idiosyncratic error term. $\omega_{i,t}$ and $p_{j,t}$ are the *multilateral resistance* terms and are usually unobserved.

There are multiple possibilities to control for MTR. Baier & Bergstrand (2002) and Carrere (2006) interpreted MTR as *remoteness*, which they define as output-weighted average of bilateral distances.

$$R_i = \left[ \sum_{k=1,k\neq i}^{N} Y_k (D_{ik})^{1-\sigma} \right]^{1/1-\sigma} \qquad (10)$$

$R_{i(j)}$ is the remoteness of country i(j), Yi(j) is output of country i(j), $D_{i(j)k}$ is bilateral distance between countries i(j) and k, while *sigma* is a measure of elasticity.

Besides modelling MTR explicitly, they can be controlled for by the inclusion of dummy variables. Depending on the assumptions imposed on MTR, the objective of the study and the (time) structure of the data set, the exact specification can take different forms as explained in the following section.

## 2.3 Correct Specification

Early gravity literature used cross-section data exclusively and relied on basic estimation techniques like OLS. The time dimension was utilized to average observations over time in order to minimize problems due to missing observations and measurement errors. Researchers soon started to exploit the multidimensionality of the data using panel data estimation techniques.

While the improvement of theoretical foundations and econometric methods greatly increased the general acceptance of the gravity model and the credibility of the estimates, it also made apparent the limitations of empirical applicability. As Baldwin and Taglioni (2006) pointed out, most studies suffer from biased estimates due to omitted variables or misspecification. They call these biases the bronze, silver and gold medal errors and stress that the most prominent of these errors, the gold one, stems from omission of or insufficient control for *multilateral resistance*. Eventually they make the point that in order to fully eliminate all three errors a gravity equation including either time-varying country dummies and time-invariant pair dummies[8] (specification 1) or time-varying pair dummies (specification 2) has to be estimated.[9]

Both specifications come with the drawback of making some effects impossible to identify. The inclusion of time-varying pair effects makes the identification of most policy effects impossible. It is therefore rarely useful. By including time-invariant pair dummies, bilateral trade cost determinants that are constant over time cannot be identified. These include distance, adjacency as well as cultural and historical ties. Therefore, this specification is not useful either for studies which aim to estimate such effects. To estimate these, the so-called main effects model including exporter, importer and time effects (specification 3) can be used.

One issue that specification 1 fails to take into account is the significance of trade imbalances. Baldwin and Taglioni (2006) recognize the issue and also note that these imbalances are not randomly distributed but rather persistent. However, they do not account for it in any of their proposed specifications. Figures 1 and 2 suggest that trade imbalances are important in the context of intra-European trade as they show differences between import and export shares. We can see that Germany, besides being the most important importer and exporter, has a considerably larger export share than import share. For the Netherlands, Belgium-Luxembourg and Great Britain such discrepancies are also apparent, while for most countries they are not easy to see due to their small shares. However, some small countries' discrepancies are large enough to be evident despite the sizes of their shares. Ireland clearly exports more than it imports as does Norway, while Greece's import share exceeds its export share greatly.

[8] They obtain dubious estimates which they ascribe to lack of variation in the EMU variable. By estimating the same equation on a larger data set with more variation their estimates become more plausible. There is significant variation in my data set.
[9] Baldwin and Taglioni (2006) p.23

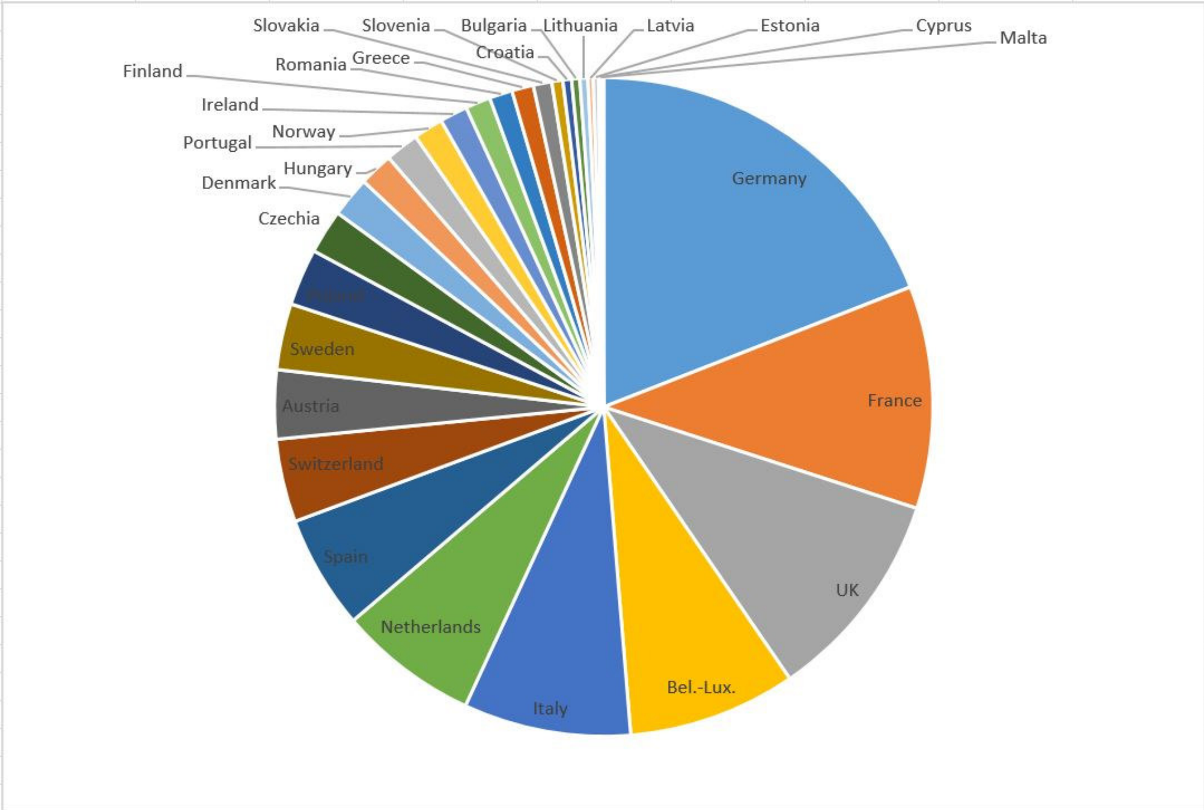


Figure 1: shares of intra-European imports by country (averages 1996-2015)

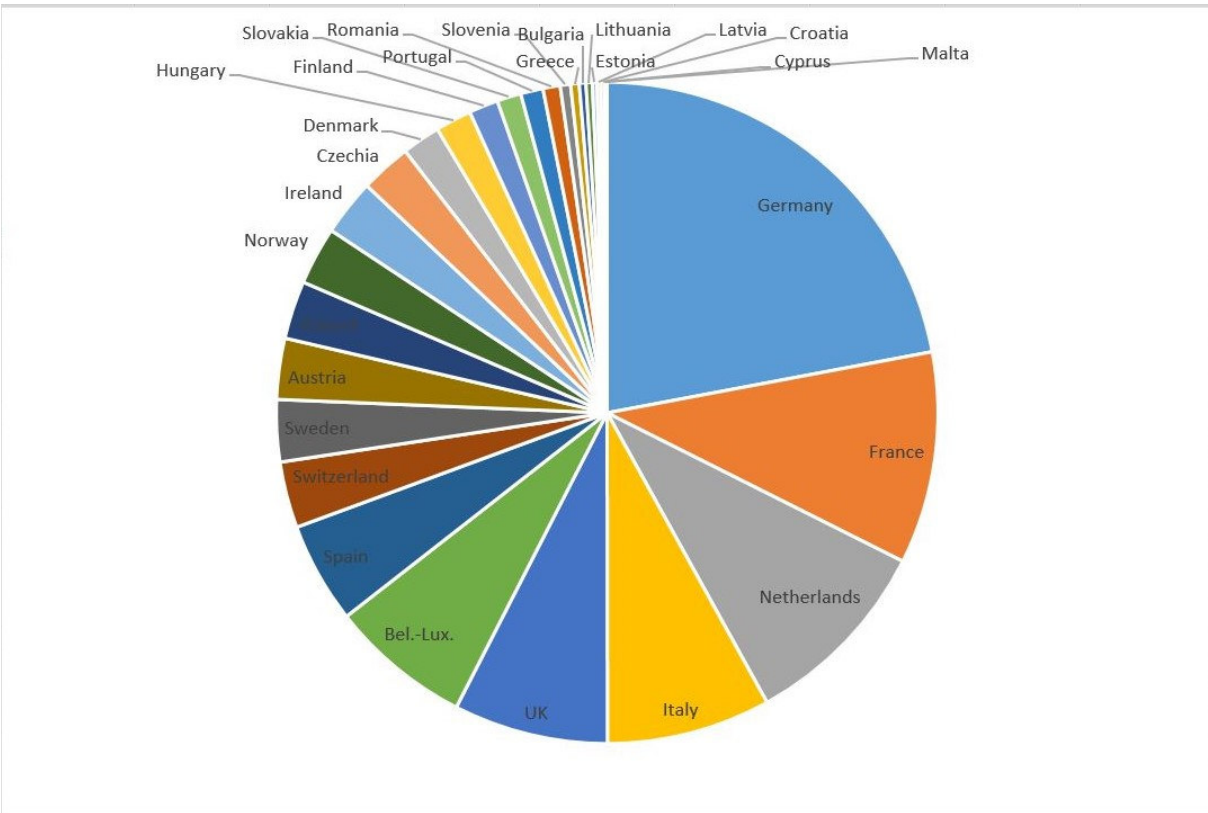


Figure 2: shares of intra-European exports by country (averages 1996-2015)

To account for these tendencies of import or export oriented countries, specification 1 can be adjusted in two different ways. One possibility is to combine it with the main effects model by including importer-specific and exporter-specific effects. Under the assumption of constant tendencies over time, such a specification properly accounts for these. If however, these tendencies are assumed to change over time, specification 1 needs to be adjusted in a different way, namely by replacing the time-varying country dummies by role specific ones, i.e. time-varying importer and exporter dummies (specification 4).

In this paper specifications 3 and 4 will be used. To analyze the dependency on choice of the distance metric, specification 3 will be utilized. Even though this specification suffers from some omitted variable bias it is often used to estimate effects which other specifications are unable to identify. It is therefore important to avoid additional bias due to improper measurement of distance. After that, the gravity equation will be re-estimated following specification 4 and particular attention will be paid to the EU and CU coefficients. Specification 4 should remove most of the misspecification issues that bias these coefficients and we should obtain adequate estimates. Furthermore, the issue of missing trade observations will be studied using specification 4.

# 3 The Data Set

Historically, the huge amount of data that is commonly used for gravity estimations was not easy to come by. 50 years ago neither data availability nor quality was comparable to today's standards. This was especially the case when the estimations were performed on a global level. Some countries' data – mostly developing countries' – are still of poor quality and availability even today. The analysis in this paper benefits from two circumstances:

1. The high quality and good availability of data today in general

2. The high standards of reporting statistical data in Europe in particular

30 countries are included in the data set. These countries are all current member states (EU-28) plus Norway and Switzerland. Because national trade statistics for Belgium and Luxembourg are not available separately before 2000, I treated them as a single country adding up their imports from and exports to other countries. This yields an effective total of 29 countries. (Table 1)

The panel covers a period of 20 years, 1996 - 2015. This should amount to a total of 29x28x20 = 16240 observations. With only one value missing, namely Malta's imports from Slovenia in 2002, the data set comes remarkably close to being fully balanced.

| Country | ISO code | Country | ISO code | Country | ISO code |
|---|---|---|---|---|---|
| Switzerland | CHE | France | FRA | Slovenia | SVN |
| Malta | MLT | Croatia | HRV | Slovakia | SVK |
| Cyprus | CYP | Italy | ITA | Sweden | SWE |
| Bulgaria | BGR | Hungary | HUN | UK | GBR |
| Czech Rep. | CZE | Finland | FIN | Norway | NOR |
| Denmark | DNK | Netherlands | NLD | Estonia | EST |
| Germany | DEU | Austria | AUT | Latvia | LVA |
| Ireland | IRL | Poland | POL | Lithuania | LTU |
| Greece | GRC | Portugal | PRT | Bel-Lux | BEL |
| Spain | ESP | Romania | ROU | | |

Table 1: List of all countries in the data set

Due to the logarithmic transformation missing trade values are a potential problem in gravity model estimations. It is important to distinguish between true zeros, i.e. either no trade or such a small amount that it falls below a certain threshold and missing values for which trade did take place but was not reported. Malta's imports from Slovenia span from 880,419 USD in 2001 to 8,304,817 USD in 2007 and were around 3.8 Million USD on average. The smallest reported trade value in the data set is that of Latvia to Malta in 1997 amounting to only 1234 USD from which we can conclude that the threshold value for reporting imports is smaller than that. The chance that Slovenian exports to Malta were so low that year seems unreasonably small and not having exported anything seems even more unlikely. I therefore looked at the Slovenian export statistics and observed that Slovenia did report exports to Malta in the extent of 1,456,823 USD, which seems to be in line with the rest of the data and I therefore set Malta's imports to that value.

With that I have a fully balanced panel which to my knowledge has not been used before in gravity model estimations. This is of particular importance because of the logarithmic specification. Nitsch (2002) examines the effect of missing values on gravity estimation results and concludes that in Rose (2000) the effect on the currency union coefficient was striking (Nitsch, 2002, p.469). In section 5 I determine the bias due to missing values by generating them in my balanced data set and comparing the results to those of the base case with the full data set.

The three basic variables are trade flows, GDP and distance. Dummy variables for sharing a land border, speaking the same language, both countries being EU members and both countries using the Euro currency are used to control for the respective characteristics, while additional dummies are used to control for *multilateral resistance*.

## 3.1 GDP and Trade Data

I consider unilateral trade flows for the dependent variable. Theoretically, using import data should be equivalent to using export data since by definition country A's imports from country B equal country B's exports to country A. In reality imports and exports often do not match. Some authors argue[10] that import data is more reliable than export data because of the fact that imports are usually taxed and therefore incentives to report them correctly are higher. Even though this argument is somewhat mitigated since intra-EU trade is not taxed, the data set goes back to 1996 when not all included countries where EU members.

The United Nations provide trade data that can be accessed through their web interface.[11] The database offers the choice between goods and services as well as between imports and exports reported at different frequencies (monthly or annual). Furthermore, the data can be categorized by either SITC, HS or BEC. For the data set in this paper annual imports of goods were downloaded. The trade values are denominated in current US dollars.

GDP data was obtained from Eurostat through their web interface[12], where it is available on a national level as well as NUTS level. The latter becomes relevant in the following section, where a new distance measure is introduced. Eurostat's GDP data is denominated in current EUR.

Not many studies address the issue of correct denomination specifically as it – depending on the specification of the gravity equation – may not be of importance. If time-varying country(-role) dummies are included in order to control for *multilateral resistance*, the proper GDP deflators are captured by these dummies as well. This is the case in my specification 4 of the gravity equation. If, however, as in specification 3 for example, such dummies are not included, proper deflation is required. The fact that trade volume is, as in my

[10] See for example Baldwin and Taglioni (2006), p.13
[11] https://comtrade.un.org/data/
[12] http://ec.europa.eu/eurostat/de/data/database

data set, usually expressed in USD might wrongfully lead to the conclusion that a US price index should be used.[13] According to Baldwin and Taglioni this might create biases via spurious correlation.[14] To properly account for inflation, the reporting country's price trend needs to be taken into account. In order to obtain GDP deflators for each country in the data set, real time series were downloaded alongside the nominal GDP time series. Using these deflators, trade volumes as well as GDP data were deflated.

## 3.2 Distance measure

Section 2 of this paper illustrated how drawing on the analogy between Newton's Law of Gravity and its economic counterpart requires adequate interpretation of the variables. Distance is the only variable that translates naturally into the economic context. This led to the early literature focusing primarily on trade, mass and other variables while using a rather rudimentary distance metric. Tinbergen (1962) introduced a simple proxy using great circle distances between countries' economically most important cities.[15] To this day, many economists use this or similar distance metrics and while this may suffice for capturing the basic idea, there are better, more sophisticated distance measures available in the literature today. The distance measure is supposed to reflect how far goods have to be shipped. Since they are not shipped from capital to capital distance between them might not be representative. The majority of trade takes place between the economically most active areas. While population has also been shown to be a determinant of trade, economic activity is estimated to have a significant effect more consistently. Therefore, a proper distance measure has to take into consideration the distribution of economic activity.

Considering the example of Germany and Italy makes this very apparent. The distance between Berlin and Rome is around 1185 km. However, since most of Germany's economic activity takes place in the south and west of the country and Italy's economic activity is concentrated in the north, these countries are actually closer together from a trade related perspective. (Fig. 3)

[13] Bussière, Fidrmuc, and Schnatz (2008) deflate USD-denominated trade data by the US industrial price index. Furthermore they deflate GDP data by the US CPI.

[14] Baldwin and Taglioni (2006), p.7

[15] Tinbergen (1962) referred to them as *economic centers*. To avoid confusion about the concept of *economic centers* introduced in this paper, I avoid using this expression at this point.

Figure 3: Distribution of Germany's and Italy's economic activity

An interesting, recent contribution is a paper by Rauch (2016) who proposes to measure distance as a harmonic mean of pairwise distances and bases this approach on geometrical considerations. In fact, he argues that assuming trade as a stationary process spreading out randomly from a point of origin into two dimensional space, a gravity equation including the size of both countries and distance in the form he proposes emerges from geometric considerations alone. Without the inclusion of an economic model Rauch's approach lacks credibility as a real contender to explain the gravity equation, however, his rationalization of the persistence of the size of distance coefficients is convincing. He argues that the inverse-square law in three-dimensional space should reduce to an inverse-proportional law in two dimensions.

Mayer and Zignano (2005) aim to find consistent measures for internal (within a country) and external (between two countries) distance. Using weighted distance measures based on bilateral distances between the largest cities of a country or two countries, respectively. They argue that this kind of consistency in measuring distance is crucial for estimating the boarder effect properly. Antweiler (2007) agrees and expands on that idea by proposing a more rigorous way of measuring distance. He argues that considering only a fixed number of cities does not take into account the complete distribution of population. Using the *Gridded Population of the World* version 3 (GPWv3) database, he calculates population-based weights and instead of the biggest cities he takes the center points of the squares formed by the grid. There are two problems with this approach. One is of technical nature as the resolution of the grid is limited and squares around country borders potentially overlap with multiple countries, which makes the attribution of the population within that square imprecise. The other problem is of conceptual nature as basing the distance measure on population distribution does not account for distribution of economic activity directly, but only approximately.

Head and Mayer (2002) overcome these problems by using a somewhat similar approach to mine. They also rely on regional GDP statistics to account for distribution of economic activity but use a different concept to calculate distance. While they use a concept like Rauch (2016), Mayer and Zignano (2005), Antweiler (2007) and Head and Mayer (2002), i.e. calculate harmonic means of pairwise distances, the metric presented in this thesis is based on a concept closely related to that of center of mass in physics. In physics, when calculating the gravitational pull between two objects that is caused by their masses, the distribution of these masses matters. One way to take into account mass distribution is to represent each body by its center of mass. In three-dimensional Euclidean space this is a straight-forward procedure. If the masses of all the parts and their positions (by choosing a sufficiently fine resolution, parts can accurately be represented as points) are known, the center of mass can be calculated as a weighted average.

A similar concept can be applied for the distribution of economic activity, although the spherical geometry requires some additional considerations. The first one is whether a centroid on a sphere always exists and if so, whether it is always unique. Importantly, we are interested in a centroid <u>on</u> a sphere, not <u>of</u> a sphere. We do not look at the sphere as an object in three-dimensional Euclidean space, but rather at its surface as a curved, two-dimensional space. Galperin (1993) analyzed centers of mass in curved spaces and concluded that on a sphere of arbitrary dimensions a center of mass can be correctly defined. He proved its existence and its uniqueness conditional on all points being on one hemisphere. This holds for every country in our data set. The second consideration is how to define and calculate a center point. Russ and Filmore (2001) suggest two algorithms for iteratively calculating such points.

Their suggestions, however, go beyond the scope of this paper in terms of complexity and I therefore use a simpler method which still takes into account the distribution of economic activity (=mass) and the spherical shape of the earth. I call the centroid *economic center* and the distance between two countries' economic centers *economic distance* as opposed to *capital distance*. Economic centers are defined as follows:

1. Every country is divided in its NUTS-1 regions. If not at least 3 NUTS-1 regions are defined for a country, NUTS-2 is taken instead. If not at least 3 NUTS-2 regions are defined, NUTS-3 is taken instead.
2. The center of each NUTS region is defined as the largest city by population in that respective region or the capital of that region.
3. Eurostat provides GDP data on NUTS level. With that the economic weight of each region can be calculated by dividing the region's GDP through the country's GDP. For example, in 2010 Germany had a GDP of around 2.9 trillion EUR while Bavaria's regional GDP amounted to around 500 billion EUR. Bavaria's economic weight is therefore $500/2900 \approx 0.17$.
4. The coordinates, latitude and longitude, of each regional center are converted into Cartesian coordinates.
5. The weighted sum of all regions is calculated for each country. This is done by calculating weighted averages for all three coordinates separately.
6. The resulting coordinates represent the economic center. Lastly, they are converted back into latitudes and longitudes.

This procedure is best explained by an example, hence, the *economic center* of Germany will be calculated step by step. Germany is divided into 16 NUTS 1 regions which correspond to Germany's constituent states. An overview of all the relevant data is provided in table 2.

| NUTS 1 region | Center | Regional GDP | Weight | Coordinates |
|---|---|---|---|---|
| Baden-Würtemberg | Stuttgart | 440 057m € | 0.151 | 48.776N, 9.183E |
| Bayern | München | 524 064m € | 0.180 | 48.137N, 11.576E |
| Berlin | Berlin | 117 750m € | 0.040 | 52.519N, 13.408E |
| Brandenburg | Potsdam | 62 150m € | 0.021 | 52.396N, 13.061E |
| Bremen | Bremen | 30 360m € | 0.010 | 53.076N, 8.807E |
| Hamburg | Hamburg | 103 566m € | 0.036 | 53.551N, 9.993E |
| Hessen | Wiesbaden | 251 517m € | 0.086 | 50.082N, 8.241E |
| Mecklenburg-Vorpommern | Schwerin | 38 634m € | 0.013 | 53.633N, 11.417E |
| Niedersachsen | Hannover | 254 659m € | 0.087 | 52.374N, 9.739E |
| Nordrhein-Westfalen | Düsseldorf | 627 219m € | 0.215 | 51.226N, 6.783E |
| Rheinland-Pfalz | Mainz | 128 135m € | 0.044 | 50N, 8.271E |
| Saarland | Saarbrücken | 33 685m € | 0.012 | 49.233N, 7E |
| Sachsen | Dresden | 109 097m € | 0.037 | 51.049N, 13.738E |
| Sachsen-Anhalt | Magdeburg | 55 845m € | 0.019 | 52.133N, 11.617E |
| Schleswig-Holstein | Kiel | 84 364m € | 0.029 | 54.325N, 10.141E |
| Thüringen | Erfurt | 54 549m € | 0.019 | 50.978N, 11.029E |

Table 2: Germany's NUTS 1 regions. Their economic (column 3 and 4) and geographic (column 5) data are used to compute Germany's *economic center.*

The latitudes and longitudes are converted into Cartesian coordinates assuming a spherical earth and according to the following formulas:

$$x = \cos(\varphi) * \cos(\lambda)$$
$$y = \cos(\varphi) * \sin(\lambda)$$
$$z = \sin(\varphi) \qquad (11)$$

Where $\varphi = \frac{lat*\pi}{180}$ and $\lambda = \frac{lon*\pi}{180}$ are latitude and longitude converted from degrees to radians.

Berlin's geodetic coordinates 52.519N, 13.408W for instance translate to $\begin{pmatrix} 0.592 \\ 0.141 \\ 0.794 \end{pmatrix}$.

The *economic center*'s Cartesian coordinates are calculated as the weighted average:

$$\begin{pmatrix} x_{ec} \\ y_{ec} \\ z_{ec} \end{pmatrix} = \sum_{i=1}^{n} \delta_i \begin{pmatrix} x_i \\ y_i \\ z_i \end{pmatrix} \qquad (12)$$

Where n is the number of regions, $\delta_i$ is region i's weight and $\begin{pmatrix} x_i \\ y_i \\ z_i \end{pmatrix}$ are the Cartesian coordinates of region i's center. The conversion back into geodetic coordinates again assumes earth to be a sphere:

$$\lambda_{ec} = atan2(y_{ec}, x_{ec})^{16}$$
$$\varphi_{ec} = atan2\left(z_{ec}, \sqrt{x_{ec}^2 + y_{ec}^2}\right) \qquad (13)$$

Berlin's *economic* center in Cartesian coordinates is $\begin{pmatrix} 0.626 \\ 0.106 \\ 0.772 \end{pmatrix}$, which translates to 50.551N, 9.591E.

Germany's *economic center* is close to the city Fulda about 343 km southwest of Berlin. That is the most extreme example in the data set. Nevertheless, more than half of the countries' *economic centers* are distanced more than 100 km from their capitals. On average the distance between capital and *economic center* is around 98 kilometers. Figure 4 shows capitals (red dots), while *economic centers* are depicted in figure 5 (green dots).

[16] The function atan2 is in principle the arctan function but takes two arguments instead of only one. Therefore it can take into account the signs of both arguments and output the appropriate quadrant of the calculated angle. A definition of the atan2 function can be found in the appendix.

Figure 4: red dots represent capitals

Figure 5: green dots represent *economic centers*

Going back to our initial example of Germany and Italy, their *economic distance* amounts to about 830 kilometers, which represents a reduction of almost 30% compared to capital distance. Surprisingly, this is not at all an extreme example as the new measure changes distances quite drastically. On average the distance changes more than 10% with the most drastic case being the distance between Austria and Slovakia. Their capitals lie only 58 km apart while their *economic distance* is 281 km – an increase of 387%! Table 3 shows the biggest differences between economic and capital distance in both absolute and relative terms.

| country 1 | country 2 | economic dist. - capital dist. (relative change) | country 1 | country 2 | relative change (absolute value) |
|---|---|---|---|---|---|
| Germany | Portugal | -490.47 km (-21.19%) | Austria | Slovakia | +387.25% (+222.99 km) |
| Germany | Spain | -451.83 km (-24.17%) | Finland | Estonia | +219.56% (+184.24 km) |
| Germany | Finland | +405.02 km (+36.66%) | France | UK | +72.29% (+281.38 km) |
| Germany | Italy | -354.63 km (-29.94%) | Germany | Denmark | +71.16% (+252.75 km) |
| Germany | Estonia | +348.96 km (+33.54%) | Hungary | Austria | +68.28% (+146.59 km) |

Table 3: most extreme changes between capital distance and *economic distance* between

### 3.3 Other Determinants of Trade Friction

The dummy of main interest is the CU dummy. It takes the value 1 if both countries are using the Euro at that point in time and is supposed to reflect the effect the Euro has on inter-European trade. Furthermore, several dummy variables that are widely regarded as economically and statistically relevant were included in order to avoid omitted variable bias.

## 4 Endogeneity

Before we turn to our estimations some known issues that are not addressed in this thesis should not go unmentioned. There are three potential sources of endogeneity bias.

- Endogeneity due to omitted variables
- Endogeneity of GDP
- Endogeneity of the EU and CU dummies

The first type of endogeneity is a question of proper specification. It is substantial if Anderson and Van Wincoop (2006)'s *multilateral resistance* is not accounted for properly. I believe that with the specifications to be estimated in the following section this source of endogeneity is kept to a minimum.

The second type of endogeneity is difficult to avoid since GDP depends on (net) imports by definition and can only be avoided by using instruments for GDP. Frankel (1997) accounted for endogeneity of income using labor force and the stock of physical and human capital as instruments for GDP and concluded that "the endogeneity of income makes little difference". (p. 135)

The third type of endogeneity is directly associated with the variables of interest, the EU and CU dummies. Countries don't enter free trade agreements or unions randomly but rather for reasons that might be related to the level of trade.[17] For unions where trade is the predominant aspect, this argument is convincing. However, there is reason to believe that the EU is a special case and endogeneity might be less important. As already mentioned in the introductory section of this thesis, the EUs agenda is not purely economic but largely driven by political factors. The reasons for entering might therefore be less related to the level of trade and endogeneity less of an issue. Still endogeneity of EU and CU is likely to prevail and the effects might therefore be *underestimated*.[18]

# 5 Statistical Analysis

We are now turning to the estimations of the gravity equation according to specifications 3 and 4 that were explained in section 2.3. This section consists of three parts. First, following specification 3 the new distance metric will be compared to the concept of capital distance that is still used frequently in gravity literature. Second, specification 4 will be estimated and differences between the estimates of specification 3 and 4 analyzed. Thirdly, missing trade observations will be generated randomly and different ways to deal with missing data compared. Moreover, robustness of the obtained results with respect to sample reduction will be tested at the end of each part.

## 5.1 Economic Distance versus Capital Distance

Section 3 made apparent the discrepancies between capital distance and *economic distance*. The differences are substantial both in absolute and in relative terms. The question remains, however, what impact these differences make with respect to the estimates of the gravity equation. Using specification 3

$$x_{ij,t} = \alpha_0 + \alpha_1 y_{i,t} + \alpha_2 e_{j,t} + \alpha_3 d_{ij} + \alpha_4 eu + \alpha_5 cu + \alpha_6 adj + \alpha_7 coml - \omega_{i,t} + (\sigma - 1)p_{j,t} + \vec{\delta}^T \vec{d} + \varepsilon_{ij,t} \quad (14)$$

[17] Baier and Bergstrand (2002), p.1

[18] Baier and Bergstrand (2002), p.24

I will try to answer this question. Note that here $\vec{d}$ is a vector containing the dummy variables according to specification 3, i.e. importer-specific, exporter-specific and year-specific dummy variables. Table 4 displays the estimation results.

| **lnTRADE** | **Regression 1 (capital distance)** | **Regression 2 (economic distance)** |
|---|---|---|
| ln1GDP | 0.830*** (0.077) | 0.890*** (0.074) |
| ln2GDP | 1.734*** (0.077) | 1.796*** (0.074) |
| lnDIST | -1.385*** (0.014) | -1.727*** (0.015) |
| EU | 0.198*** (0.021) | 0.134*** (0.020) |
| CU | -0.046** (0.021) | -0.057*** (0.020) |
| BORDER | 0.426*** (0.026) | 0.035 (0.026) |
| COMLANG | -0.383*** (0.046) | -0.501*** (0.044) |
| CONSTANT | -2.770* (1.509) | -2.162 (1.437) |
| Adj. $R^2$ | 0.915 | 0.923 |
| Num. of obs. | 16240 | 16240 |
| Coefficients of importer, exporter and time dummies not shown<br>*: significant at 10%, **: significant at 5%, ***: significant at 1%<br>Standard errors in paranthesis | | |

Table 4: regressions following specification 3

The choice of distance measure clearly has an effect on the estimates. First, the coefficient on distance itself has increased (in absolute terms) by several standard errors from -1.385 to -1.727. With the new distance measure countries generally move closer together. This is especially the case for some of the economically most important countries that trade the most like Germany, France, Spain and Italy. The reduction of distance is balanced by a higher coefficient. Moreover, the increased distance effect seems to be robust to different samples as the respective coefficients in table 5 are consistently and significantly higher than that of Regression 1.

More importantly, the choice of the distance measure affects the coefficients of the main policy variables EU and CU. The trade stimulating effect of the European Union is mitigated from around 22% ($e^{0.198}\approx1.219$) to around 14% ($e^{1.134}\approx1.143$), while the effect of the Euro becomes even more negative. It increases (in absolute terms) from -4.5% ($e^{-0.046}\approx0.955$) to

-5.5% ($e^{-0.057}\approx0.945$). The change due to the new distance metric aside, the negative effect of the Euro is quite surprising. While many studies have estimated the effect of the Euro to be small or even insignificant, a significantly negative effect has not yet been estimated to my knowledge. I will hold off from discussing this result until the next part of this section, where I expect to obtain a more accurate estimate following specification 4.

The effects of adjacency and speaking the same language are affected by the new distance metric as well. The effect of the BORDER dummy loses its significance while the COMLANG dummy's effect becomes even more negative. The negative effect of speaking the same language is somewhat surprising. Although the effect has been shown to differ for different languages, it is generally positive for the languages that are relevant in this data set, i.e. German, English and French.

| **lnTRADE** | **Regression 3** | **Regression 4** | **Regression 5** | **Regression 6** | **Regression 7** |
|---|---|---|---|---|---|
| Ln1GDP | 0.838***<br>(0.125) | 0.875***<br>(0.079) | 0.862***<br>(0.080) | 0.670***<br>(0.075) | 0.660***<br>(0.081) |
| ln2GDP | 2.152***<br>(0.125) | 1.823***<br>(0.079) | 1.753***<br>(0.080) | 1.735***<br>(0.075) | 1.701***<br>(0.081) |
| lnDIST | -2.211***<br>(0.026) | -1.806***<br>(0.017) | -1.763***<br>(0.016) | -1.560***<br>(0.015) | -1.597***<br>(0.017) |
| EU | 0.190***<br>(0.034) | 0.133***<br>(0.022) | 0.178***<br>(0.022) | 0.253***<br>(0.018) | 0.288***<br>(0.019) |
| CU | 0.095*<br>(0.055) | -0.023<br>(0.025) | -0.086***<br>(0.022) | -0.086***<br>(0.018) | -0.140***<br>(0.021) |
| BORDER | 0.082*<br>(0.045) | 0.152***<br>(0.030) | 0.290***<br>(0.031) | 0.031<br>(0.022) | 0.256***<br>(0.027) |
| COMLANG | - | -0.680***<br>(0.056) | -1.031***<br>(0.056) | -0.262***<br>(0.034) | -0.741***<br>(0.043) |
| CONSTANT | -2.680<br>(2.421) | -1.783<br>(1.549) | -0.841<br>(1.548) | 0.396<br>(1.451) | 1.360<br>(1.566) |
| Adj. $R^2$ | 0.897 | 0.913 | 0.911 | 0.928 | 0.911 |
| Num. of obs. | 6840 | 14440 | 14040 | 11040 | 9240 |
| omitted | If either importer or exporter are among Eurozone founders | If both importer and exporter are among Eurozone founders | Germany and France | Cyprus, Malta, Estonia, Lithuania and Latvia | Germany, France, Cyprus, Malta, Estonia, Lithuania and Latvia |
| Coefficients of importer, exporter and time dummies not shown<br>*: significant at 10%, **: significant at 5%, ***: significant at 1%<br>Standard errors in paranthesis | | | | | |

Table 5: regressions on restricted data samples following specification 3 and using *economic distance*

## 5.2 Specification 3 versus Specification 4

Because time-invariant importer and exporter fixed effects along with time dummies, as used in specification 3, do not fully control for *multilateral resistance*, the gravity equation is now estimated according to specification 4 as described in section 2.3 of this thesis paper.

$$x_{ij,t} = \alpha_0 + \alpha_1 y_{i,t} + \alpha_2 e_{j,t} + \alpha_3 eu + \alpha_4 cu - \omega_{i,t} + (\sigma - 1) p_{j,t} + \vec{\gamma}^T \vec{g} + \varepsilon_{ij,t} \qquad (15)$$

Here, $\vec{g}$ is a vector consisting of importer-time, exporter-time and country-pair fixed effects while $\vec{\gamma}$ contains the respective coefficients. The estimation results are shown in table 6.

| **lnTRADE** | **Reg. 8** | **Reg. 9** | **Reg. 10** | **Reg. 11** | **Reg. 12** | **Reg. 13** |
|---|---|---|---|---|---|---|
| ln1GDP | 1.046***<br>(0.026) | 0.931***<br>(0.042) | 0.724***<br>(0.027) | 0.836***<br>(0.029) | 0.724***<br>(0.021) | 0.522***<br>(0.028) |
| ln2GDP | 1.079***<br>(0.030) | 1.136***<br>(0.037) | 0.982***<br>(0.039) | 0.702***<br>(0.046) | 0.750***<br>(0.021) | 0.888***<br>(0.032) |
| EU | 0.018<br>(0.029) | 0.137**<br>(0.065) | 0.059*<br>(0.035) | 0.048<br>(0.033) | 0.185***<br>(0.024) | 0.209***<br>(0.027) |
| CU | -0.021<br>(0.027) | 0.197***<br>(0.064) | -0.002<br>(0.033) | -0.015<br>(0.030) | 0.026<br>(0.176) | 0.037<br>(0.028) |
| CONST. | -6.076***<br>(0.652) | -5.470***<br>(0.723) | -0.306<br>(0.620) | 1.398*<br>(0.808) | 2.779***<br>(0.403) | 3.083***<br>(0.572) |
| Adj. $R^2$ | 0.961 | 0.936 | 0.953 | 0.951 | 0.972 | 0.962 |
| obs. | 16240 | 6840 | 14440 | 14040 | 11040 | 9240 |
| omitted | none | If <u>either</u> importer <u>or</u> exporter are among Eurozone founders | If <u>both</u> importer <u>and</u> exporter are among Eurozone founders | Germany and France | Cyprus, Malta, Estonia, Lithuania and Latvia | Germany, France, Cyprus, Malta, Estonia, Lithuania and Latvia |
| Coefficients of importer-time, exporter-time and country-pair dummies not shown<br>*: significant at 10%, **: significant at 5%, ***: significant at 1%<br>Standard errors in paranthesis | | | | | | |

Table 6: regressions following specification 4

Right away it can be seen that the point estimates differ heavily. The GDP coefficients drop to about unity and are more in line with the literature, in particular with Rauch (2016). They drop even further in the robustness analysis when estimating the equation on subsamples but always stay highly significant.

The policy variables turn insignificant (Reg. 8), however, this is not robust to using different subsamples. While the CU coefficient turns significant (and positive, in contrast to specification 3) in only one robustness regression (Reg. 9), the effect of EU gains significance in four out of five (all but Reg. 11). CU turns significant in the smallest subsample when all observation where at least one of the trading partners is a Eurozone founding member are dropped. Eurozone founding members consist mostly of so-called core states, which leads to the implication that positive Euro effects are mostly driven by periphery trade.

## 5.3 Missing Trade Observations and Zero Trade Flows

A distinctive feature of the data set used in this thesis paper is the fact that it is fully balanced. Missing values are a common issue in gravity literature and improper treatment thereof has been shown to alter estimation results. Another problem that goes hand in hand with missing trade observations are zero trade flows. Not all countries trade with each other. For studies performed at a global level this is a common situation. The problem that arises is due to the multiplicative nature of the gravity equation. It can predict arbitrarily small trade flows but not zero trade if the continuous variables (here: the mass variables and distance) are restricted to be positive and finite. Furthermore, it may be difficult to distinguish between actual zeros and trade observation that are existent but missing, because data bases often do not provide additional information to why the data is reported as missing or zero.

| **lnTrade** | **Reg. 8** | **Reg. 14** | **Reg. 15** | **Reg. 16** | **Reg. 17** |
|---|---|---|---|---|---|
| ln1GDP | 1.046***<br>(0.026) | 0.899***<br>(0.029) | 1.055***<br>(0.102) | 1.057***<br>(0.040) | 1.042***<br>(0.307) |
| ln2GDP | 1.079***<br>(0.030) | 1.083***<br>(0.029) | 1.198***<br>(0.119) | 1.217***<br>(0.022) | 1.088***<br>(0.355) |
| EU | 0.018<br>(0.029) | 0.019<br>(0.029) | 0.117<br>(0.114) | 0.032<br>(0.031) | 0.113<br>(0.342) |
| CU | -0.021<br>(0.027) | -0.025<br>(0.027) | 0.112<br>(0.106) | -0.025<br>(0.029) | 0.118<br>(0.317) |
| CONST. | -6.076***<br>(0.652) | -3.709***<br>(0.358) | -7.789***<br>(2.544) | -8.232***<br>(0.622) | -8.948<br>(7.616) |
| Adj. $R^2$ | 0.961 | 0.961 | 0.6142 | 0.965 | 0.134 |
| obs. | 16240 | 16086 | 16240 | 14629 | 16240 |
| omitted | Full data set | ca. 1% of observations are randomly missing | ca. 1% of observations are randomly set to zero (=arbitrary small amount) | ca. 10% of observatio ns are randomly missing | ca. 10% of observations are randomly set to zero (=arbitrary small amount) |
| Coefficients of importer-time, exporter-time and country-pair dummies not shown<br>*: significant at 10%, **: significant at 5%, ***: significant at 1%<br>Standard errors in paranthesis | | | | | |

Table 7: regression using the full, balanced data set (column 2) is compared to regressions on data sets with randomly missing trade observations (columns 3-6)

Tables 7-9 show estimation results when missing observations or zeros are generated in my balanced panel. The second column always shows Reg. 8 from table 6 as the baseline case.

Table 7 analyzes the situation of randomly missing values. In columns 3 and 4 ca. 1% of the trade observations are missing, while in columns 5 and 6 around 10% missing values were generated. Surprisingly, the point estimates and their significance hardly change. It is less surprising for Reg. 14 and Reg. 16 because the missing observations were simply omitted from the data set. It is however surprising for Reg. 15 and Reg. 17, where missing trade observations were set to be around 3USD.[19] What does change drastically is the explanatory power of the gravity equation as the adjusted $R^2$ drops to around 61% in Reg. 15 and 13% in Reg. 17.

| **lnTrade** | **Reg. 8** | **Reg. 18** | **Reg. 19** | **Reg. 20** | **Reg. 21** |
|---|---|---|---|---|---|
| ln1GDP | 1.046***<br>(0.026) | 0.934***<br>(0.032) | 1.615***<br>(0.073) | 0.993***<br>(0.036) | 2.301***<br>(0.193) |
| ln2GDP | 1.079***<br>(0.030) | 1.041***<br>(0.024) | 1.605***<br>(0.085) | 1.052***<br>(0.023) | 1.963***<br>(0.223) |
| EU | 0.018<br>(0.029) | 0.085***<br>(0.028) | -0.311***<br>(0.081) | 0.192***<br>(0.026) | -0.720***<br>(0.215) |
| CU | -0.021<br>(0.027) | -0.009<br>(0.026) | -0.090<br>(0.075) | 0.044*<br>(0.023) | -0.305<br>(0.199) |
| CONST. | -6.076***<br>(0.652) | -4.091***<br>(0.466) | -21.835***<br>(1.814) | -4.695***<br>(0.541) | -36.862***<br>(4.781) |
| Adj. $R^2$ | 0.961 | 0.962 | 0.800 | 0.964 | 0.660 |
| obs. | 16240 | 16078 | 16240 | 14616 | 16240 |
| omitted | Full data set | ca. 1% of observations are randomly missing, smaller trade volumes are more likely to be missing | ca. 1% of observations are randomly set to zero (=arbitrary small amount), smaller trade volumes are more likely to be zero | ca. 10% of observations are randomly missing, smaller trade volumes are more likely to be missing | ca. 10% of observations are randomly set to zero (=arbitrary small amount), smaller trade volumes are more likely to be zero |
| Coefficients of importer-time, exporter-time and country-pair dummies not shown<br>*: significant at 10%, **: significant at 5%, ***: significant at 1%<br>Standard errors in paranthesis | | | | | |

[19] The goal was to choose an arbitrary small amount. For simplicity the ln was set to be 1, which amounts to a trade volume of e≈2.72.

Table 8: regression using the full, balanced data set (column 2) is compared to regressions on data sets with randomly missing trade observations (columns 3-6), where the probability for an observation to be missing decreases as the value increases.

The changes are much more severe in Table 8, where missing observations were generated randomly again but with one important difference. Instead of every observation having the same chance to be missing, the probability decreases as the trade volume increases. This procedure is supposed to account for the fact that small trade values can fall below reporting thresholds and are more likely to not be reported than the important, big trade flows. I believe that this is a more realistic scenario than the previous one. Not only do the mass coefficients vary a lot but the effect on the EU coefficient is drastic. It is significant in all four regressions and changes signs depending on the way of treating missing values. When they were omitted from the regression, the European Union's trade stimulating effect was estimated to be positive. On the other hand, when missing observations were set to be zero (arbitrarily small), the effect was estimated to be negative. The currency union effect was not affected much as it stayed small and insignificant in Reg. 18, 19 and 21 and was estimated to be positive at a 10% level in Reg. 20.

| **lnTrade** | **Reg. 8** | **Reg. 22** | **Reg. 23** | **Reg. 24** | **Reg. 25** |
|---|---|---|---|---|---|
| ln1GDP | 1.046***<br>(0.026) | 0.959***<br>(0.034) | 1.430***<br>(0.041) | 0.908***<br>(0.028) | 2.057***<br>(0.084) |
| ln2GDP | 1.079***<br>(0.030) | 1.035***<br>(0.024) | 1.381***<br>(0.048) | 0.843***<br>(0.042) | 1.869***<br>(0.098) |
| EU | 0.018<br>(0.029) | 0.067**<br>(0.029) | -0.120***<br>(0.046) | 0.161***<br>(0.027) | -0.965***<br>(0.094) |
| CU | -0.021<br>(0.027) | -0.020<br>(0.026) | -0.047<br>(0.042) | 0.005<br>(0.025) | -0.092<br>(0.087) |
| CONST. | -6.076***<br>(0.652) | -4.608***<br>(0.412) | -15.875***<br>(1.020) | -1.000**<br>(0.461) | -31.408***<br>(2.093) |
| Adj. $R^2$ | 0.961 | 0.962 | 0.916 | 0.963 | 0.807 |
| obs. | 16240 | 16190 | 16240 | 15894 | 16240 |
| omitted | Full data set | Trade values below 200000$ are missing | Trade values below 200000$ are set to zero (=arbitrary small amount) | Trade values below 2000000$ are missing | Trade values below 2000000$ are set to zero (=arbitrary small amount) |
| Coefficients of importer-time, exporter-time and country-pair dummies not shown<br>*: significant at 10%, **: significant at 5%, ***: significant at 1%<br>Standard errors in paranthesis | | | | | |

Table 9: In columns 3-6 trade values below a certain threshold were treated as missing.

A similar situation can be observed in table 9. The mass coefficients vary a lot between Reg. 22 to 25, EU's coefficient becomes significant and turns negative in Reg. 23 and 25, while the coefficient of CU is not affected much.

## 6 Conclusion

The trade stimulating effect of the European Union and the Euro has been the subject of many studies that utilize the gravity model of trade. This thesis paper adds to this literature by introducing a new metric to measure distance between countries. Using this new metric instead of capital distance that is used in most other studies changes the estimated effects of the policy variables EU and CU. This thesis paper shows that the choice of distance metric does influence the estimates, however, the specification of the gravity equation seems to play a more important role. When using dummy variables to properly account for both the cross-sectional and the temporal dimension of *multilateral resistance*, the results change much more drastically.

Results are generally not robust to restriction of the data sample. It is therefore not surprising that missing observations have a big influence on estimates as well. The way of dealing with missing values also seems to play an important role. Interpreting missing trade data as missing trade potentially biases results severely, while simply omitting them from the data set seems to have less of an impact. In case of doubt observations should therefore be eliminated from the data sample before interpreting missing values as missing trade.

The point estimates of the trade stimulating effect of the Euro vary between -5.5% to no significant effect at all. Including the estimations on the restricted data samples, the estimates vary even more. It seems to be difficult to obtain reliable estimates, however, this thesis paper provides support for the numerous studies that ascribe only a marginal effect on intra-European trade to the Euro.

## References


Anderson, James E. "A theoretical foundation for the gravity equation." *The American Economic Review* 69.1 (1979): 106-116.

Anderson, James E. "The gravity model." *Annual Review of Economics* 3.1 (2011): 133-160.

Anderson, James E., and Eric Van Wincoop. "Gravity with gravitas: a solution to the border puzzle." *The American Economic Review* 93.1 (2003): 170-192.

Antweiler, Werner "Internal and external distance: Gravity depends on it." *Vancouver, Canada: University of British Columbia* (2007). Retrieved from http://old-hha.asb.dk/nat/workshop/2007/wa0711.pdf

Baier, Scott L., and Jeffrey H. Bergstrand. "On the endogeneity of international trade flows and free trade agreements." Manuscript, August 2002, retrieved from http://www.nd.edu/~jbergstr/Working_Papers/EndogeneityAug2002.pdf

Baldwin, Richard, and Daria Taglioni. "Gravity for dummies and dummies for gravity equations." *NBER Working Paper Series*, September 2006, w12516, retrieved from http://www.nber.org/papers/w12516.pdf

Baltagi, Badi H., Peter H. Egger, and Katharina Erhardt. "The Estimation of Gravity Models in International Trade." In: Matyas L. (eds) *The Econometrics of Multi-dimensional Panels. Advanced Studies in Theoretical and Applied Econometrics* (2017), 323-348. Springer, Cham

Baltagi, Badi H., Peter H. Egger, and Michael Pfaffermayr. "Panel data gravity models of international trade." (2014). CESifo Working Paper Series, February 2014, No. 4616, retrieved from https://poseidon01.ssrn.com/delivery.php?ID=394013103113029026117021109123069109050063050068079069007112001009071031107020119009034102098026110059062006093116087002015067045007060077040004020105086125117110024020033120010084010008024007088003081093103125074066002026012065027122026070004020067&EXT=pdf

Bergstrand, Jeffrey H. "The gravity equation in international trade: some microeconomic foundations and empirical evidence." *The Review of Economics and Statistics* 67.3 (1985): 474-481.

Bongardt, Annette, Francisco Torres, Carsten Hefeker, Pierre Wunsch, and Christoph Hermann. "Convergence in the EU." *Intereconomics* 48.2 (2013): 72-92.

Braun, Daniela, and Markus Tausendpfund. "The impact of the Euro crisis on citizens' support for the European Union." *Journal of European Integration* 36.3 (2014): 231-245.

Brun, Jean-Francois, Celine Carrere, Patrick Guillaumont, and Jaime de Melo. "Has distance died? Evidence from a panel gravity model." *The World Bank Economic Review* 19.1 (2005): 99-120.

Buss, Samuel R., and Jay P. Fillmore. "Spherical averages and applications to spherical splines and interpolation." ACM *Transactions on Graphics (TOG)* 20.2 (2001): 95-126.

Bussière, Matthieu, Jarko Fidrmuc, and Bernd Schnatz. "EU enlargement and trade integration: Lessons from a gravity model." *Review of Development Economics* 12.3 (2008): 562-576.

Carrere, Céline. "Revisiting the effects of regional trade agreements on trade flows with proper specification of the gravity model." *European Economic Review* 50.2 (2006): 223-247.

Dabrowski, Marek. "The global financial crisis: Lessons for European integration." *Economic Systems* 34.1 (2010): 38-54.

De Grauwe, Paul. "Design Failures in the Eurozone: Can they be fixed?." (2013). LEQS Paper No. 57. Retrieved from https://poseidon01.ssrn.com/delivery.php?ID=682025021087125104024007078124074121029022041052070026086028018005003001120001081121054013022125049112043066117016127077097024007071082078061115009121119075072111125089001060110011120126089109084100125024066095031019008003028000125079081002069021104002&EXT=pdf

De Grauwe, Paul, and Frauke Skudelny. "The impact of EMU on trade flows." Weltwirtschaftliches Archiv 136.3 (2000): 381-402. Retrieved from https://doi.org/10.1007/BF02707286

Egger, Peter. "A note on the proper econometric specification of the gravity equation." *Economics Letters* 66.1 (2000): 25-31.

Egger, Peter, and Michael Pfaffermayr. "The proper panel econometric specification of the gravity equation: A three-way model with bilateral interaction effects." *Empirical Economics* 28.3 (2003): 571-580.

Feenstra, Robert. *Advanced International Trade: Theory and Evidence.* NJ: Princeton University Press, 2004, Princeton.

Frankel, Jeffrey A.. *Regional trading blocs in the world economic system*. Peterson Institute, 1997, Washington, D.C..

Frankel, Jeffrey, and Andrew Rose. "An estimate of the effect of common currencies on trade and income." *The Quarterly Journal of Economics* 117.2 (2002): 437-466.

Galperin, G. A. "A concept of the mass center of a system of material points in the constant curvature spaces." *Communications in Mathematical Physics* 154.1 (1993): 63-84.

Head, Keith, and Thierry Mayer. "Illusory border effects: Distance mismeasurement inflates estimates of home bias in trade." Vol. 1. Paris: CEPII, 2002.

Jones, Erik. "The Euro and the financial crisis." *Survival* 51.2 (2009): 41-54.

Konya, Laszlo, Laszlo Matyas, and Mark Harris. "GATT/WTO membership does promote international trade after all–Some new empirical evidence." (2011). MPRA Paper No. 34978, retrieved from https://mpra.ub.uni-muenchen.de/34978/1/MPRA_paper_34978.pdf

Lane, Philip R. "The real effects of European monetary union." *The Journal of Economic Perspectives* 20.4 (2006): 47-66.

Lane, Philip R. "The European sovereign debt crisis." *The Journal of Economic Perspectives* 26.3 (2012): 49-67.

Lapavitsas, Costas, Annina Kaltenbrunner, Duncan Lindo, J. Michell, Juan Pablo Painceira, Eugenia Pires, Jeff Powell, Alexis Stenfors, and Nuno Torres. "Eurozone crisis: beggar thyself and thy neighbour." *Journal of Balkan and Near Eastern Studies* 12.4 (2010): 321-373.

Linders, Gert-Jan, and Henri LF De Groot. "Estimation of the gravity equation in the presence of zero flows." (2006). Tinbergen Institute Discussion Paper No. 06-072/3. Retrieved from https://poseidon01.ssrn.com/delivery.php?ID=331005097121004077074000079110097090117046093006002049011115025091078012124068016002000030020015104116044116071017102005115108021022011041126123093025088006120006039052003095090078030121072115068091019095077113089013001127093006121095067117020112&EXT=pdf

Mátyás, László. "Proper econometric specification of the gravity model." *The World Economy* 20.3 (1997): 363-368.

Mátyás, László, and László Balázsi. “The estimation of multi-dimensional fixed effects panel data models.” Department of Economics No. 2012_2, retrieved from http://www.personal.ceu.hu/staff/repec/pdf/2012_2.pdf

Nitsch, Volker. "Honey, I shrunk the currency union effect on trade." *The World Economy* 25.4 (2002): 457-474.

Nuroğlu, Elif, and Robert M. Kunst. "Competing specifications of the gravity equation: a three-way model, bilateral interaction effects, or a dynamic gravity model with time-varying country effects?." *Empirical Economics* 46.2 (2014): 733-741.

Pfaffermayr, Michael, and Peter H. Egger. "The pure effects of European Integration on intra-EU core and Periphery trade." (2002). University of Innsbruck Economics Working Paper No. 02/01. Retrieved from https://poseidon01.ssrn.com/delivery.php?ID=099009117110031075065118101024103105007007077053031036010084003084106000011122012058059039038022091116022064091047005029046043086098085127000079012125099076002059011013125095018098028102089102070071090065098121073027066102127092068099123068117&EXT=pdf

Rauch, Ferdinand. "The geometry of the distance coefficient in gravity equations in international trade." *Review of International Economics* 24.5 (2016): 1167-1177.

Rose, Andrew K. "One money, one market: the effect of common currencies on trade." *Economic Policy* 15.30 (2000): 08-45.

Rose, Andrew K., and Eric Van Wincoop. "National money as a barrier to international trade: The real case for currency union." *The American Economic Review* 91.2 (2001): 386-390.

Stockhammer, Engelbert. "Peripheral Europe's debt and German wages: the role of wage policy in the Euro area." *International Journal of Public Policy* 7.1-3 (2011): 83-96.

Tinbergen, Jan. "Shaping the world economy; suggestions for an international economic policy." Twentieth Century Fund, 1962, New York.

Van Rompuy, Herman. "Towards a genuine economic and monetary union." *European Council. Brüssel* 5 (2012). Retrieved from http://welingelichtunlimited.net/wu/wp-content/uploads/2012/06/Brief-van-Rompuy-juni-20121.pdf

Wu, Haonan. "Re-visiting the Distance Coefficient in Gravity Model." *arXiv preprint arXiv:1503.05283* (2015). Retrieved from https://arxiv.org/ftp/arxiv/papers/1503/1503.05283.pdf

# Appendix

## A.1 Derivation of the structural gravity equation from section 2.2

$$\max \quad Q_j = \left[\sum_{i=1}^{N} Q_{ij}^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}} \qquad s.t. \quad \sum_{i=1}^{N} E_{ij} = E_j \qquad w.r.t. \quad Q_{kj}$$

where $Q_j$ is a Dixit-Stiglitz aggregator function, $Q_{ij}$ are goods shipped from country i to country j and $\sigma$ represents the elasticity of substitution between different goods. $E_{ij} = P_{ij}Q_{ij}$ is country j's expenditure on goods from country i and $Q_{kj}$ is one representative bilateral trade flow. The optimization problem can be solved via a Lagrange function yielding the following N F.O.C.s.:

$$\left[\sum_{i=1}^{N} Q_{ij}^{\frac{\sigma-1}{\sigma}}\right]^{\frac{1}{\sigma-1}} Q_{kj}^{\frac{-1}{\sigma}} - \lambda P_{kj} = 0 \qquad \forall\, k = 1, \dots, N \qquad (1a)$$

Multiplying by $Q_{kj}$ and rearranging yields

$$P_{kj}Q_{kj} = \frac{1}{\lambda}\left[\sum_{i=1}^{N} Q_{ij}^{\frac{\sigma-1}{\sigma}}\right]^{\frac{1}{\sigma-1}} Q_{kj}^{\frac{\sigma-1}{\sigma}} \qquad \forall k = 1, \dots, N \qquad (2a)$$

which still holds true for every k=1,..,N. We can therefore sum over of these N F.O.C.s to get

$$\sum_{k=1}^{N} P_{kj}Q_{kj} = \frac{1}{\lambda}\left[\sum_{i=1}^{N} Q_{ij}^{\frac{\sigma-1}{\sigma}}\right]^{\frac{1}{\sigma-1}} \sum_{k=1}^{N} Q_{kj}^{\frac{\sigma-1}{\sigma}} \qquad (3a)$$

The expression on the left hand side is the expenditure of country j on tradable goods $E_j$. Recognizing that the two sums on the right hand side are equal, we can simplify the right hand side to $\frac{1}{\lambda}Q_j$, where $Q_j$ is the Dixit-Stiglitz aggregator function as defined above. Hence, we can rewrite the equation in terms of aggregate variables

$$Q_j = \lambda E_j \qquad (4a)$$

Using $(4a)$ to substitute for $\lambda$ and raising both sides of the equation to the power of $1-\sigma$, we can rewrite the N F.O.C.s in $(1a)$ as

$$P_{kj}^{1-\sigma} = \left[\frac{E_j}{Q_j}\right]^{1-\sigma} \left[\sum_{i=1}^{N} Q_{ij}^{\frac{\sigma-1}{\sigma}}\right]^{-1} Q_{kj}^{\frac{\sigma-1}{\sigma}} \qquad \forall k = 1, \dots, N \qquad (5a)$$

which can again be summed over all k=1,...,N. If we define the Dixit-Stiglitz aggregate price $P_j := \left[\sum_{k=1}^{N} P_{kj}^{1-\sigma}\right]^{\frac{1}{1-\sigma}}$, which can be interpreted as multilateral trade resistance as it depends on all bilateral resistance terms via the bilateral prices $P_{kj}$, after some rearrangement we obtain an expenditure equation in terms of Dixit-Stiglitz aggregates

$$E_j = P_j Q_j \qquad (6a)$$

$(6a)$ relates $P_j$ to the Lagrange multiplier as $P_j = \frac{1}{\lambda}$.

$(1a)$ can be stated as $P_{kj} = \frac{1}{\lambda} Q_j^{1/\sigma} Q_{kj}^{-1/\sigma}$ and $\frac{1}{\lambda}$ substituted by $P_j$ to obtain

$$P_{kj} = P_j Q_j^{\frac{1}{\sigma}} Q_{kj}^{\frac{-1}{\sigma}} \qquad \forall k = 1, \dots, N \qquad (7a)$$

If we raise $(7a)$ to the power of $-\sigma$ and substitute $Q_j$ according to $(6a)$ and rearrange, we obtain

$$Q_{kj} = \frac{P_{kj}^{-\sigma} E_j}{P_j^{1-\sigma}} \qquad \forall k = 1, \dots, N \qquad (8a)$$

Multiplication by $P_{kj}$ yields

$$E_{kj} = \left[\frac{P_{kj}}{P_j}\right]^{1-\sigma} E_j \qquad \forall k = 1, \dots, N \qquad (9a)$$

which states that expenditure of country j on goods from country i depends on country j's total expenditure on (tradable) goods as well as on the relative price. Using the equilibrium condition $E_k = \sum_{j=1}^{N} E_{kj}$ together with $P_{kj} = P_k \tau_{kj}$ and $(9a)$ we can obtain an expression for domestic prices

$$P_k^{1-\sigma} = \frac{E_k}{\sum_{j=1}^{N} \frac{\tau_{kj}^{1-\sigma} E_j}{P_j^{1-\sigma}}} \qquad (10a)$$

which we can use to eliminate domestic prices in $(9a)$ and finally obtain

$$E_{kj} = \frac{\tau_{kj}^{1-\sigma} E_k E_j}{\sum_{j=1}^{N} \left[\frac{\tau_{kj}^{1-\sigma} E_j}{P_j^{1-\sigma}}\right] P_j^{1-\sigma}} \qquad (11a)$$

We already interpreted $P_j^{1-\sigma}$ as multilateral resistance. The other term in the denominator can also be interpreted as multilateral resistance. The fact that these two terms differ means that we have to distinguish between *inward resistance* (multilateral resistance of the importing country) and *outward resistance* (multilateral resistance of the exporting country). Defining *outward resistance* of country k as $\Omega_k := \sum_{j=1}^{N} \frac{\tau_{kj}^{1-\sigma} E_j}{P_j^{1-\sigma}}$ , we obtain a compact representation of the structural gravity equation, which corresponds to Baldwin and Taglioni (2006)'s equation (7).

$$E_{kj} = \frac{\tau_{kj}^{1-\sigma} E_k E_j}{\Omega_k P_j^{1-\sigma}} \qquad (12a)$$

## A.2 Definition of the atan2 function

atan2 can be defined in terms of the standard arctan function as follows:

$$atan2(y,x) = \begin{cases} \arctan\left(\frac{y}{x}\right) & if\ x > 0 \\ \arctan\left(\frac{y}{x}\right) + \pi & if\ x < 0\ and\ y \geq 0 \\ \arctan\left(\frac{x}{y}\right) - \pi & if\ x > 0\ and\ y > 0 \\ \frac{\pi}{2} & if\ x = 0\ and\ y > 0 \\ -\frac{\pi}{2} & if\ x = 0\ and\ y < 0 \\ undefined & if\ x = 0\ and\ y = 0 \end{cases}$$

## A.3 Data

All data available from the author upon request.

## A.4 Abstract


### Abstract (English)

The European Union in general and the euro in particular have faced increasing criticism since the last financial crisis. The gravity model of trade, the workhorse of international trade literature, has been successfully utilized to estimate negligible trade stimulating effects of the euro. It is important to quantify this effect properly in order to be able to put it into perspective with other positive and negative stimuli associated with the euro. This paper aims to fulfill three purposes: First, it adds to the literature about the trade stimulating effect of the euro. Second, and this is the main purpose of the paper, it introduces a new metric for measuring distances between countries. Third, the use of a balanced panel data set makes it possible to determine the bias due to missing observations and mistreatment thereof.

### Abstract (Deutsch)

Seit der letzten Finanzkrise stehen die Europäische Union und insbesondere der Euro in zunehmender Kritik. Empirische Studien unter Verwendung von Gravitationsmodellen legten bisher geringe Auswirkungen der Gemeinschaftswährung auf inter-Europäischen Handel nahe. Das Verstehen und korrekte Quantifizieren dieser Auswirkungen sind deshalb von Bedeutung, weil positive und negative Effekte des Euro ganzheitlich betrachtet und gegeneinander abgewogen werden sollten. Diese Arbeit erfüllt drei Zwecke: Sie trägt zur Literatur über und somit zum Verständnis von handelsstimulierenden Effekten des Euro bei. Weiters, und das ist der wichtigste Beitrag der Arbeit, tut sie dies unter der Verwendung einer neuartigen Distanzmetrik zur Messung von Entfernungen zwischen Staaten. Zuletzt ist es aufgrund der hohen Datenqualität möglich, die Verzerrung zu bestimmen, welche durch fehlende Datenpunkte auftreten kann.