Tariffs, Non Tariff Measures and Imports

A comparison of econometric approaches to estimate the gravity equation[1]

Master's Thesis in Economics & Business, specialisation International Economics

Student: Eline van den Bosse

Supervisor: Dr. Koen G. Berden

Second reader: Dr. Julian Emami Namini

Rotterdam, 7 April 2013

Abstract: This thesis studies the effect of Tariffs and Non Tariff Measures on Imports, by using bilateral trade data. The data is specified on 4-digit HS-code, with a total of 1223 different products. Theory describes that using OLS for the estimation of the gravity equation has three major shortcomings: the bias created by the logarithmic transformation of the dependent variable, the violation of the homoskedasticity assumption, and the zero trade values who are removed from the analysis or are manipulated by adding a small number. Theory suggests to use Poisson Models to solve these problems. In the analysis we compare OLS, Poisson Model, Negative Binomial Poisson Model, and Zero Inflated Negative Binomial Poisson Model and apply them to the gravity equation. As control variables, common border and common language are used. The main conclusion is that tariffs have a robust significant negative impact on imports over all the different econometric specifications, and increasing tariffs raise the probability of a zero trade flow. There is evidence that NTMs have a positive impact on imports and the probability to trade. The effect of NTMs on imports is an overall effect. Individual effects of several types of NTMs are unknown and can still be negative. The effect found possibly suffers from several problems, such as the underlying definition and reverse causality. Robustness checks show that it is likely that at least the probability to trade is negative related to having NTMs in place.

Key words: Gravity Model, Tariffs, Non Tariff Measures, International Trade, Imports

JEL classification: F13, F14

[1] Acknowledgements: Special thanks to my supervisor for his suggestions and helpful comments. S.D.G.

Table of contents

Abbreviations

AHS	Effectively Applied Tariffs
AIC	Akaike Information Criterion
AVE	Ad valorem Equivalent
BIC	Bayesian Information Criterion
COMTRADE	Common Format for Transient Data Exchange
CTHA	Chemical Tariff Harmonisation Agreement
EPA	Environmental Protection Agency
EU	European Union
FDA	Food and Drug Administration
GDP	Gross Domestic Product
HM	Heckman Model
HS	Harmonised System of trade data
IMF	International Monetary Fund
Ln	Natural Logarithm
MRT	Multilateral Resistance Term
N	Number of observations
NBPM	Negative Binomial Poisson Model
NTM	Non Tariff Measure
OECD	Organisation for Economic Co-operation and Development
OLS	Ordinary Least Squares
PM	Poisson Model
REACH	Regulation on registration, evaluation, authorisation and restriction of chemicals
SD	Standard Deviation
SPS	Sanitary and Phytosanitary measures
TBT	Technical Barriers to Trade
TRAINS	Trade Analysis and Information System Database
TSCA	Toxic Substances Control Act
UNCTAD	United Nation's Conference on Trade and Development
USD	US Dollars
WITS	World Integrated Trade Solution
WTO	World Trade Organization
ZINBPM	Zero Inflated Negative Binomial Poisson Model

1. Introduction

"We are in a trade world today that is characterised more by Non-Tariff Barriers than by traditional tariffs"

WTO Director-General Pascal Lamy, 2013[2]

For more than half a century, the gravity equation has been used to explain the effects of all kinds of trade costs on bilateral trade flows. Common language, trade agreements, colonies, immigrants, common borders, bilateral distance and political system are examples of these trade costs. Several empirical studies have evaluated the impact of Tariffs and Non Tariff Measures (NTMs) on bilateral trade flows (Disdier and Marette, 2010; Hoekman and Nicita, 2008). The economic theory behind the impact of tariffs is very clear. Tariffs raise the price of an imported good and a higher price lowers trade (Bowen et al, 2012). Tariffs are easily quantifiable, but NTMs are not and they are also more difficult to evaluate because they are multidimensional in nature, not all NTMs have the same effect and their effect is likely to be non-linear. The definition used by the WorldBank of Non Tariff Measures is "any government action with a potential effect on the value, volume, or direction of trade. These are all barriers to international trade other than the tariffs for example, quotas, licensing, voluntary export restraints."[3] NTMs have become more prominent in the regulation of international trade. It becomes increasingly important to get insight in the effects of NTMs on trade.

This paper adds to the literature by providing more comprehensive empirical evidence on the effect of Tariffs and NTMs on imports by breaking down bilateral trade flows into 1223 product categories (4-digit HS-code). Secondly, this paper contributes by constructing a large database with the most recent data on NTMs provided by TRAINS. The analyses of the impact of NTMs on trade are mostly based on data before 2010. Before 2010, less countries were in the dataset and the data was less complete on the different types of NTMs. Nowadays, there is data available from TRAINS on 6-digit HS-code and for more countries. Because of the improved data, a better estimation of the effect of NTMs on trade can be made. The third contribution is providing insight into the econometric issues regarding different approaches to

[2] Contribution of the WTO Director-General Pascal Lamy to the conference "Changing Landscape of International Trade", 14 March 2013
[3] WITS glossary, http://wits.worldbank.org/glossary.html

measure the gravity equation. Looking at the econometric estimation of the gravity equation, this paper tries to build a correct specification by comparing different econometric approaches. Hence, the overall research question of this paper is as follows: *"What is the effect of Tariffs and Non Tariff Measures on Imports as found by using different econometric specifications of the gravity equation?"*

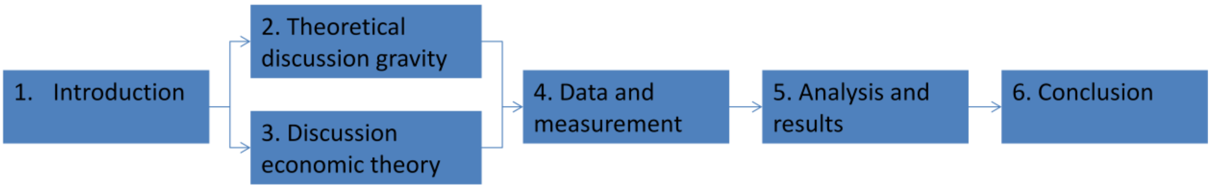
Firstly, the research question is answered by deriving the effects based on theory. Secondly, the research question is empirically tested. The empirical analysis is based on a database constructed for this paper, which contains more than 6 million bilateral trade flows on a 4-digit product category level. The gravity equation is estimated for all models by using OLS, OLS with small added numbers for the dependent variable, Poisson Model, Negative Binomial Poisson Model, and Zero Inflated Negative Binomial Poisson Model. Robustness checks have been carried out by using a different definition for the variables of distance and language, by estimating without fixed effects, removing outlier Egypt from the analysis, and estimating the equation without tariffs so we can use the full database.

The theoretical analysis indicates that an increase in tariffs will reduce trade (Bowen et al, 2012). The effect for NTMs is unknown, based on the theory that NTMs can have negative effects and positive effects (Disdier and Marette, 2010; Fugazza, 2013). When there are several NTMs in place for one product level, the overall impact is related to the relative strength of the different NTMs in place. These theoretical assumptions are tested empirically in this research, including carrying out an extensive sensitivity analysis. In the empirical analysis, we find that tariffs always have a negative effect on the trade value of imports and the probability to trade. NTMs have a positive effect on imports, and a positive effect on the probability to trade. Our results support our theoretical prediction that tariffs have a negative effect on trade. The positive effect that we find for the impact of NTMs on imports suggests that the overall effect of NTMs on trade is positive. Using a dummy-variable in the analysis makes it impossible to split the result into individual effects. The positive effect that we find is probably a biased effect because of the definition underlying the data. Another shortcoming is that a dummy-variable for NTMs does no justice to the variety in NTMs. In the robustness checks, we find that the effect of NTMs on the probability of trade changes in a negative effect.

The thesis proceeds as follows. Chapters 2 and 3 provide an overview of the literature.

Chapter 2 contains a theoretical discussion of the gravity equation. Chapter 3 describes the economic mechanisms behind the relation between NTMs and imports and the measurement of NTMs. Chapter 4 describes the data and estimation framework. Chapter 5 discusses the analysis and results. Lastly, chapter 6 presents the conclusions. The structure is schematically presented in Figure 1.1 below.

Figure 1.1 Schematic overview



2. Theoretical discussion of the gravity equation

2.1 The evolution of the gravity equation

The flow of products, people and information can be predicted by gravity models, who are derived from Newton's law of gravity. The origin of the gravity model of trade lies in the work of Tinbergen (1962). He introduced the gravity model to explain international bilateral trade. The gravity equation predicts that the gravitational force between two countries is proportional to the product of the masses of the two countries and inversely proportional to the distance between them (Burger et al, 2009).

The standard gravity equation for international trade takes the form:

$$T_{ij} = K\frac{M_i^{\beta_1} M_j^{\beta_2}}{D_{ij}^{\beta_3}} \quad (1)$$

where T_{ij} is the bilateral trade flow between countries i and j, M_i (M_j) is the mass of the country of origin (destination), D_{ij} is the bilateral distance between country i and j, K is a constant, β_1 is the potential to generate flows, β_2 is the potential to attract flows, and β_3 is a resistance factor reflecting the distance decay in trade (Burger et al, 2009).

The standard gravity equation can easily be extended with other variables, such as institutions, contiguity, common language, common border, free trade agreements, etc. The gravity equation was popular because of its high explanatory power. The model started with empirical validity, but after a stream of papers, the missing theoretical foundation has been laid. Early theoretical foundations can be found in the work of Anderson (1979) and Bergstrand (1985), and Helpman and Krugman (1985). Both Anderson and Bergstrand tried to solve McCallum's 'border puzzle', by addressing the role of multilateral prices (Baier and Bergstrand, 2009). Anderson and Van Wincoop (2003) show that bilateral trade not only depends on bilateral trade costs, but also on 'multilateral resistance terms' (MRTs). Omitting these MRTs may result in omitted variable bias (Anderson and Van Wincoop, 2003). There are three ways to take MRTs into account (Baier and Bergstrand, 2009). First, through country fixed effects for importers and exporters. Second, Anderson and Van Wincoop (2003) estimate a nonlinear system of trade flows and price equations. Third, Baier and Bergstrand (2009) estimate the model using simple averages.

The gravity equation can be estimated with different econometric methods. In section 2.2, different econometric methods and their pros and cons are discussed. In section 2.3 a summary of the econometric approaches and its characteristics is presented.

2.2 Econometric approaches to measure the gravity equation

2.2.1 Ordinary Least Squares (OLS)

The easiest and most used form of estimation is Ordinary Least Squares. Gravity models can be estimated in terms of natural logarithms (ln). Equation 1 becomes a linear equation when we apply a logarithmic transformation. When we add a random disturbance term (ε), the equation becomes testable (Burger et al, 2009).

$$\mathrm{LnT_{ij}} = \ln K + \beta_1 \ln M_i + \beta_2 \ln M_j - \beta_3 \ln D_{ij} + \varepsilon_{ij} \qquad (2)$$

Due to its log-linear structure, the coefficients of the gravity model can be interpreted as elasticities or ratios of percentage changes (Verbeek, 2012).

To be able to estimate unbiased and efficient coefficients of the gravity equation with OLS, we need to fulfil the so-called Gauss-Markov conditions. It is likely that at least one of these conditions cannot be fulfilled. One of the assumptions is that all the error terms have the same variance; this is called homoskedasticity (Verbeek, 2012). Homoskedasticity means – in case of the gravity equation – that the variances for small countries and big countries in terms of GDP, are the same. It is assumed that the observation of a trade flow with value 1 with an expected flow of 2 is as likely as an observed flow of 100,000 with an expected flow of 200,000 (Burger et al, 2009). With cross-sectional data, it is likely that error terms are correlated for a given country. A solution for this type of violation can be to use country fixed effects. The homoskedasticity assumption then means that the error is constant across countries (Gómez-Herrera, 2012).

The gravity equation always predicts positive trade flows, so the predictions are always bigger than zero. In practice, bilateral trade flows are very often zero. Not all countries produce all goods and there is no demand for all goods, even if countries produce a good it is not always traded. Problematic is that the zero trade flows do not occur randomly, but they can be related to independent variables, such as bilateral distance and economic mass. A shortcoming of log-linear estimation is that zero trade flows are dropped from the estimation. The log-linear gravity model cannot include zero trade flows, because the logarithm of zero is not defined. Removing the zero trade flows when they contain information about the independent variables will bias the results. The bias causes an underestimation of the effects of these variables (Linders and De Groot, 2006).

One of the approaches to address the problem of zero-valued trade flows is to add a small positive number to all trade flows. The advantage of adding these small values is that it prevents omission of observations and OLS can still be applied. Problem is that adding a value is ad hoc and does not reflect the underlying expected value. It gives no guarantee that the resulting estimates are consistent (Linders and De Groot, 2006). There is a lack of justification for the choice of the small positive values and it is proven that you can generate all parameter estimates that you like when you choose the constant (Burger et al, 2009).

A third problem that arises by applying OLS is the bias created by the logarithmic transformation of the dependent variable (Burger et al, 2009). The log-normal model generates estimates of lnI_{ij} but not of $I_{ij.}$ This problem is known as Jensen's inequality, which implies that $E(lnI_{ij}) \neq lnE(I_{ij})$. A log-linear estimation can give biased estimates in the presence of heteroskedasticity (Santos Silva and Tenreyro, 2006).

The three problems by estimating the gravity equation with a log-linear structure are the violation of the homoskedasticity assumption, the way that zero trade flows are dealt with and the logarithmic transformation of the dependent variable. The literature provides other techniques that account for these problems (Burger et al, 2009; Santos Silva and Tenreyro, 2006).

2.2.2 Heckman Model (HM)

The Heckman model (1979) is a sample selection model. Heckman was one of the first who emphasized the importance of modelling sample selection. A sample selection model tries to expose the mechanisms why a portion of the sample has an outcome that is observed and why others have not. For the Heckman model, two equations are estimated. The first equation is the Selection equation. This equation determines the binary decision whether or not bilateral trade is observed. The Selection equation is estimated by using a Probit Maximum Likelihood model:

$$\rho_{ij} = \Pr\left(h_{ij} = 1 \middle| x_1\right) = G(x_1, \beta_1) \tag{3}$$

Where ρ_{ij} is the probability that country i, exports to country j, conditional on the observed independent variables x_1. H_{ij} is a binary variable which indicates whether bilateral trade flows are positive or zero (Demaria et al, 2011).

The second equation is the Trade equation, which determines the potential size of bilateral trade. The Trade equation is estimated by using OLS. Exactly the same variables should be included into both equations, except one. The criterion for this variable is that it influences the absence of trade, but does not influence the size of trade. The Trade equation is given by:

$$E\left\{m_{ij} \middle| h_{ij} = 1\right\} = x_2\beta_2 + \sigma_{12}\lambda_{ij} \tag{4}$$

Where m_{ij} is the logarithmic observed trade flow from country i to country j given that the observed trade flow is positive, where x_2 is all the independent variables. σ_{12} is the covariance of unobserved errors of the Selection equation and the Trade equation and is estimated as a coefficient (Demaria et al, 2011).

One of the advantages of the Heckman model is that it provides a solution for the zero bilateral trade flows, because probit modelling does not face problems with zero trade flows. Another advantage is that the model allows for a different impact of the independent variables on the decision to trade and the decision on the amount of trade. On the other side, there are two disadvantages of the model. The first is that there is no solution for the problem of logarithmic transformation that can give a downward bias in the results, because the second stage of the model still makes use of a logarithmic transformed dependent variable (Demaria et al, 2011). The second disadvantage is that it is very hard to find a variable that influences the decision to trade but is unrelated to the volume of trade (Demaria et al, 2011).

2.2.3 Poisson Model (PM)

Attention has been given to the possible use of Poisson Models, after the increasing resistance against using OLS to estimate the gravity equation (Burger et al, 2009). Where the Heckman model only provides a solution for the problem of zero trade flows, Poisson Models (PMs) have a solution for all three of the problems when using OLS as explained before (Demaria et al, 2011). Poisson Models derive originally from the analysis of count data (Burger et al, 2009). The model tries to explain the expected value of the trade flow m_{ij}, given a vector of trade cost characteristics x (Verbeek, 2012).

$$E\{m_{ij}|x\} = \exp\{x\beta\} \tag{5}$$

Where equation 5 relates the expected outcome to trade cost characteristics, we also need to describe the distribution. Because m_{ij} is non-negative, we have to choose a functional form that produces non-negative conditional expectations. A common assumption is that the observed value of the trade flow has a Poisson distribution with a conditional mean (λ) that is a function of the trade cost characteristics (x) (Verbeek, 2012). This implies that the probability mass function of m_{ij} conditional upon x is given by:

$$P\{m_{ij}\} = \frac{\exp\{-\lambda_{ij}\}\lambda_{ij}^{m_{ij}}}{m_{ij}!}, \quad m_{ij} = 0, 1 \ldots \tag{6}$$

Where $m_{ij}!$ is a factorial[4]. As already mentioned, Poisson Models do not face problems with heteroskedasticity, zero trade flows and the logarithmic transformation of the dependent variable. First, Poisson estimates are still consistent, even in the presence of heteroskedasticity. By using large samples, the estimates are also reasonably efficient. Second, the Poisson function deals with zero trade flows, because of its multiplicative form. Third, the Poisson Models estimate m_{ij} instead of ln m_{ij}. This avoids biased estimates of the variable of interest (Burger et al, 2009).

An important drawback of Poisson distribution is, however, that it assumes that the conditional variance of the dependent variable (m_{ij}) is equal to its conditional mean (λ_{ij}). We refer to this condition as equidispersion (Verbeek, 2012; Burger et al, 2009).

$$V\{m_{ij}|x\} = \exp\{x\beta\} \tag{7}$$

$$E\{m_{ij}|x\} \infty V\{m_{ij}|x\} \tag{8}$$

[4] m_{ij} is a notation for m_{ij} * (m_{ij} -1) * (m_{ij} -2) * … * 2 * 1

The equality of the conditional mean and the conditional variance of the distribution is not always a plausible condition. Also in estimating the gravity equation, this condition has been doubted.

Alternative to a Poisson Model, a Negative Binomial Poisson Model (NBPM) can be employed to get along with the violation of the equidispersion condition (Demaria et al, 2011).

2.2.4 Negative Binomial Poisson Model (NBPM)

Equidispersion assumes that the conditional variance of the dependent variable (m_{ij}) is equal to its conditional mean (λ_{ij}). In practice, the conditional variance is often higher than the conditional mean. The reason for this so-called 'overdispersion' is that unobserved heterogeneity is not taken into account in the Poisson Model. This unobserved heterogeneity has its origin in omitted variables. Ignoring this overdispersion will result in consistent, but inefficient estimates (Burger et al, 2009).

To correct for overdispersion, a Negative Binomial Poisson Model (also known as NegBin II model) can be employed. An NBPM is a generalization of the PM, since it has the same expected value of the trade flow m_{ij}, but an extra parameter to model the over-dispersion. The additional term included is the dispersion parameter α (Verbeek, 2012). The variance becomes a function of both the conditional mean (λ_{ij}) and the dispersion parameter (α), thereby incorporating unobserved heterogeneity into the conditional mean (Burger et al, 2009; Cameron and Trivedi, 1986).

$$V\{m_{ij}|x\} = (1 + \alpha^2 \exp\{x\beta\}) \exp\{x\beta\} \tag{9}$$

The dispersion parameter allows the conditional variance to exceed the conditional mean. The overdispersion is increasing in α. The Wald test and likelihood ratio test are available to test for overdispersion. Rejection of the null-hypothesis that $\alpha^2=0$, is an indication of overdispersion. The alternative hypothesis is one-sided ($\alpha^2>0$). By rejection of the null-hypothesis, the NBPM is preferred over the PM. When α=0, the NBPM is the same as the Poisson Model (Verbeek, 2012).

2.2.5 Zero-Inflated Negative Binomial Poisson Model (ZINBPM)

In contrast to OLS, (Negative Binomial) Poisson Models can technically deal with zeros, but they are not able to deal correctly with excess zeros (Burger et al, 2009). When the number of observed zeros exceeds the number of zeros predicted by the model, this can be explained by a non-Poissonness process (Demaria et al, 2011). Zeros in the data are produced by two different processes. The first process is that not all pairs of countries have the potential to trade. A possible reason for a zero probability of trade is a lack of resources. Excess zeros have a trade probability of zero by definition (Demaria et al, 2011). The second process is that trade volumes are zero, even when the theoretical probability of trade is higher than zero. Possible reasons are trade costs, such as bilateral distances and differences in preferences (Burger et al, 2009).

A ZINBPM separates two kinds of zeros: true zeros and excess zeros. The two processes are estimated in two parts. The first step is an estimation of the probability of no bilateral trade at all (excess zeros/non Poisson zeros). A logit model is used to separate the true zeros and excess zeros. The second step is an NBPM given that the zero trade flow has a non-zero probability (true zeros/Poisson zeros) (Demaria et al, 2011).

The two-step approach estimates the parameter ψ_{ij}, which is the proportion of observations with a strictly zero count ($0 \leq \psi_{ij} \geq 1$). When there are no excess zeros (ψ_{ij}=0), the ZINBPM reduces to the NBPM (Burger et al, 2009). To test for excess zeros, the Vuong test is available (Vuong, 1989). Rejection of the null-hypothesis that ψ_{ij}=0, is an indication of excess zeros. The alternative hypothesis is one-sided (ψ_{ij} >0). By rejection of the null-hypothesis, the zero-inflated model is preferred over the non-zero-inflated model (Burger et al, 2009).

2.3 Summary of econometric approaches and its characteristics

In this section, Table 2.1 presents in summarized form the econometric approaches to gravity, characteristics and a summary of shortcomings and solutions.

Table 2.1 Summary of econometric approaches

Econometric approach	Characteristics	Shortcomings and solutions
Ordinary Least Squares (OLS)	• Log-linear least squares • Coefficients interpreted as elasticities / ratios of percentage changes	• Heteroskedasticity Solution: fixed effects • Zero trade flows dropped because of

Econometric approach	Characteristics	Shortcomings and solutions
	• Logarithmic transformation dependent variable	log-linear estimation Solution: add small number • Bias created by logarithmic transformation dependent variable
Heckman Model (HM)	• Two equations: selection equation and trade equation • Have the same variables except one • Selection equation estimated by Probit Maximum Likelihood • Interpretation: probability that country i exports to country j • Trade equation estimated by OLS • Interpreted as the conditional expected trade flow given that the trade observation is positive	• Bias created by logarithmic transformation dependent variable in trade equation • Difficult to find a variable that influences the decision to trade but is unrelated to volume of trade • Probit has no problem with zero trade flows • Allows different impact of the variables on decision to trade and the amount of trade
Poisson Model (PM)	• Poisson distribution • Assumes that the conditional variance of the dependent variable is equal to its conditional mean (equidispersion) • Coefficient interpreted as relative change in conditional mean	• Poisson Model deals naturally with heteroskedasticity, zero trade flows and logarithmic transformation dependent variable • Equidispersion condition has been doubted
Negative Binomial Poisson Model (NBPM)	• Variance as a function of conditional mean and dispersion parameter • Dispersion parameter incorporates unobserved heterogeneity	• Wald test for overdispersion • By rejection of null-hypothesis (α^2=0), NBPM is preferred over PM
Zero Inflated Negative Binomial Poisson Model (ZINBPM)	• Two stages: probability of no bilateral trade and volume of trade • Both equations may contain the same variables • Separates two kinds of zeros: excess zeros and true zeros. Excess zeros have a theoretical trade probability of zero and true zeros have not. • Probability of trade is estimated by Logit • Volume of trade is estimated by NBPM.	• Vuong test for excess zeros • By rejection of null-hypothesis (ψ_{ij}=0), ZINBPM is preferred over NBPM

3. Discussion of economic theory

There are several forms of trade restricting measures. We distinguish Tariffs and Non Tariff Measures. The economic theory behind the impact of tariffs is very clear. A tariff is a tax levied on imports. Economic theory predicts a negative relation between tariffs and imports, because tariffs increase the price of imported goods. A higher price of the imported good will reduce the volume of imports (Bowen et al, 2012). Tariffs are easily quantifiable, but Non Tariff Measures are not. The relation between NTMs and trade is not always a one to one negative relationship. In this chapter, we will discuss the economic relationship between NTMs and trade (section 3.1), the measurement of NTMs (section 3.2) and an illustration of the distorting effect of NTMs for the cosmetics sector (section 3.3).

3.1 The relation between Non Tariff Measures and Trade

Non Tariff Measures have become a prominent part of the regulation of international trade (Ecorys, 2009; Fugazza, 2013). The reasons for having NTMs are diverse. NTMs are often used as a policy instrument to achieve public policy objectives, such as correcting for market failures and protection of public health (WTO, 2012). NTMs can be used as consumer protection, but also as an instrument to protect domestic producers. The UNCTAD developed in 2009 a new coding system for the classification of NTMs, which distinguishes 16 categories[5]. In terms of incidence, the categories A "Sanitary and Phytosanitary (SPS) measures" and B "Technical Barriers to Trade (TBTs)" are the most used NTMs. The SPS measures are measures to protect food safety and animal and plant health, while TBTs are all other regulations, standards, testing and certification procedures (WTO, 2012). In 2010, TBTs were imposed on 30 percent of products and trade for the average country and SPS measures were on average imposed on 15 percent of products and trade (Fugazza, 2013).

[5] World Trade Report 2012, http://www.wto.org/english/res_e/booksp_e/anrep_e/world_trade_report12_e.pdf p.101
A: Sanitary and phytosanitary measures, B: Technical barriers to trade, C: Pre-shipment inspection and other formalities, D: Price control measures, E: Licences, quotas, prohibitions and other quantity control measures, F: Charges, taxes and other para-tariff measures, G: Finance measures, H: Anti-competitive measures, I: Trade-related investment measures, J: Distribution restrictions, K: Restrictions on post-sales services, L: Subsidies (excluding export subsidies), M: Government procurement restrictions, N: Intellectual property, O: Rules of origin, P: Export related measures

The use of NTMs will often have trade effects. These trade effects are less easy to define than for tariffs. There are NTMs who promote trade but in many other cases, they restrict it. Economic theory predicts that NTMs can have both positive and negative effects on the volume of trade. TBTs can be trade-impeding, because of increasing compliance costs for producers, but can also be demand-enhancing, because of decreasing information costs for consumers (Fugazza, 2013). Trade will increase or fall depending on whether the negative effect on supply is smaller than the positive effect on demand. In order to illustrate the impact of NTMs on trade, we summarize the framework used by Disdier and Marette (2010) and Fugazza (2013).

Disdier, Marette (2010) and Fugazza (2013) use a simplified framework for supply and demand for imports, which is a partial equilibrium framework. The market is assumed to be homogeneous except for a characteristic that is potentially dangerous to consumers. Both domestic and foreign goods can have this characteristic. When products are homogeneous, they are perfect substitutable. Demand and supply are derived from respectively quadratic preferences and a quadratic cost function (Disdier and Marette, 2010; Fugazza, 2013). Dependent on the nature of the NTM, the effect will be on the quantity, the demand-side and/or the supply-side.

The first category that we want to consider is the quantity restricting NTM. In the coding system of the UNCTAD, this is category E "Licences, quotas, prohibitions and other quantity control measures". We will illustrate this category by using the quota as a starting point. A quota is a limitation of the quantity of an imported good (WTO, 2012). A quota limits the level of imports to q_A'. This limitation is binding when the maximum quantity is lower than the equilibrium quantity of imports (see Figure 3.1). It is also possible that the quota is set above the level of free trade imports implying that the quota is not binding (see Figure 3.2). In case of a binding quantity restriction, a limitation of the imports to q_A' will raise the domestic price of imports to p_{AD}' which is above the world price p_A. The world price will fall when the importing country is large. We can paint the new demand curve as a line with a kink at q_A'. Who earns the price wedge is dependent on the way the licences/rights are auctioned. The distribution can influence the welfare, but has no effect on the equilibrium in our framework (Fugazza, 2013). A similar analysis applies to other quantity restricting NTMs.

Figure 3.1 Binding quantity restricting NTM

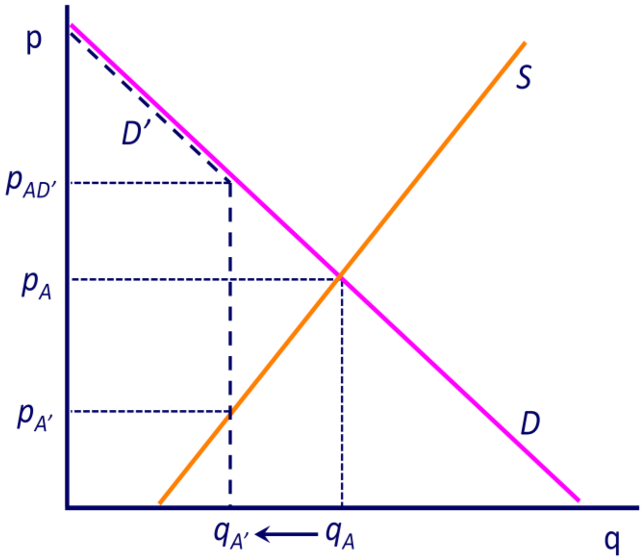
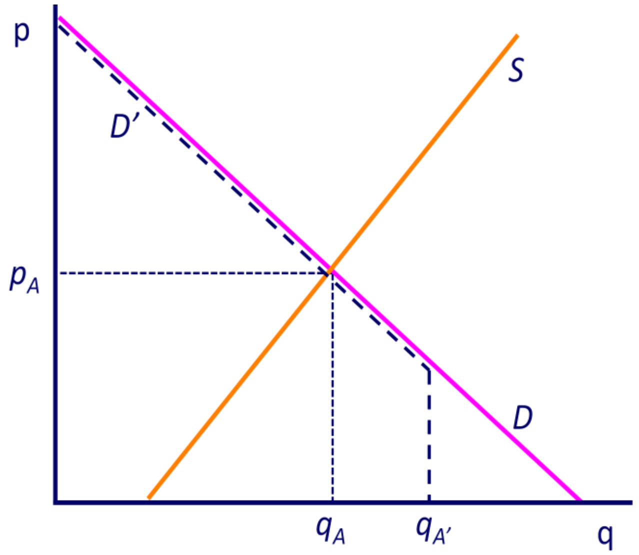
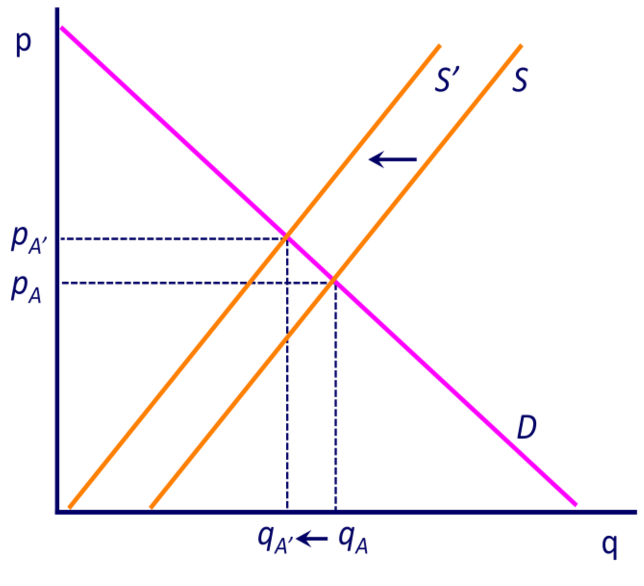


Figure 3.2 Not binding quantity restricting NTM



The second category consists of NTMs that affect the supply-side. The supply-side is the side of production, that can be affected by regulations who specify the production process and/or the product attributes. The specification of the production process can for example be the use of a certain technology. Product attributes are influenced when there is for example a maximum usage for specific substances. The adaptation of the production process and/or the product attributes will shift the supply-curve to the left, and will lower the quantity of imports (see Figure 3.3). This type of regulation is not necessarily protection of the own market, because these regulations can be used to incorporate externalities for products and processes who are hazardous for health and environment. The second category of NTMs can be found in measures such as SPS measures (category A), TBTs (category B), and other NTMs with technical regulations (Fugazza, 2013).

Figure 3.3 Supply-reducing NTM



The third category is the NTM who changes demand. This shift can be demand-enhancing and demand-declining. Information is very important for decision making. Consumers are not always aware of possible damage in consuming a good when there is no information for them available on that characteristic (Frank, 2008). Demand-shifting NTMs are used to correct for certain types of market failures. NTMs can make it compulsory to provide information about product characteristics and possible damage to the consumers. Information about negative characteristics will influence consumer behaviour and therefore reduce demand. When consumers internalize the possible damage of a product, the demand-curve will shift to the left (see Figure 3.4). The obligation to provide information will not always decline demand. The measure can also be informative about positive characteristics and can signal a higher quality. As a response, the consumers may be more willing to pay for the product and that will enhance the demand. In our simplified framework this enhancement will shift the demand-curve to the right (see Figure 3.5) (Fugazza, 2013). Demand shifting NTMs can be found under all technical regulations.

Figure 3.4 Demand-declining NTM

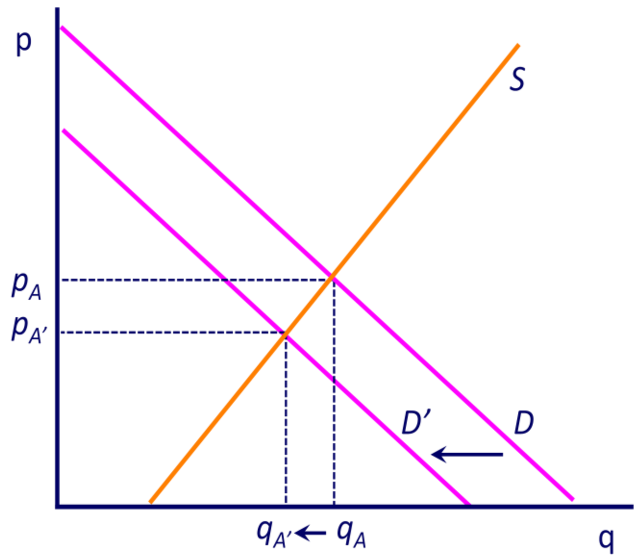
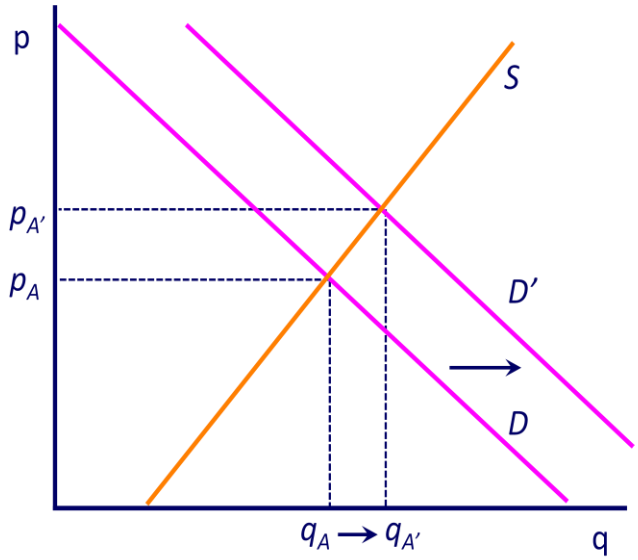


Figure 3.5 Demand-enhancing NTM



The individual effect of an NTM may be difficult to identify when there are more NTMs in place for the same product. Theory suggests that the overall impact is related to the relative strength of the different NTMs in place. That is, there is a dominant NTM in terms of impact. The impact of the other NTMs are covered by the dominant NTM. In general, when one of the NTMs is a quantity restriction (first category), it is likely that multiple NTMs will not add to each other. When all NTMs affect the supply-side (second category), their effects are likely to add to each other. In order to be able to identify the individual effect of an NTM, it is necessary to know to which category the NTM belongs (Fugazza, 2013).

3.2 The measurement of NTMs

In chapter 2, we discussed the econometric approaches to measure the gravity equation. For the estimation, it is important how to capture NTMs in the model. In the literature, we identified five key indicators. The five indicators are: binary variable, count variable, frequency index, coverage ratio and Ad valorem equivalent (AVE) of NTMs. Studies with the residual approach to estimate the effect of NTMs are not included in this indicator list, because a residue is not a variable which is based on NTM data.

The first indicator is the binary variable. The UNCTAD provides data on 6-digit HS-code product level and also for 16 NTM categories. To evaluate the impact of NTMs on product level, a dummy variable is used in many studies (Demaria et al, 2011). The NTM dummy takes the value 1 if there is at least one NTM applicable and 0 if there is no NTM in place. Shortcoming of this method is that it is impossible to identify the individual effect of an NTM. Another shortcoming is that you lose information about the number of NTMs. In some studies, the binary variable is constructed based on the NTM-category. A dummy is created for all 16 categories or for some of the individual categories (mostly TBTs and SPS measures). The advantage of including all categories of NTMs is the ability to isolate the individual impact of each category (Fugazza, 2013).

The second indicator is the count variable. Several studies use the number of NTMs on product level. The advantage of this method is that you can make a distinction within the group of products with an NTM. You will no longer lose information about the number of NTMs, but the disadvantage is that not all NTMs have the same effect, the overall impact is not the sum of the individual impacts and the existence of NTMs do not automatically imply their implementation and enforcement (Fugazza, 2013).

The third indicator is the frequency index. The frequency index is an aggregate indicator of the use and incidence of NTMs and summarises the percentage of traded products to which at least one NTM is applied. The frequency index of NTMs imposed by country j is computed as:

$$F_j = \frac{\sum NTM_i * M_i}{\sum M_i} * 100 \qquad (10)$$

Where NTM_i is a dummy variable reflecting the presence of at least one NTM for product i, M_i is a dummy variable for the existence of imports for product i. Shortcoming of this measure is that the frequency measure only gives information about the general restrictiveness of a country and no restrictiveness of an individual product. The frequency index will give a wrong impression for countries with both highly regulated sectors and sectors without NTMs (Fugazza, 2013).

The fourth indicator is the coverage ratio. The coverage ratio is a measure of the importance of NTMs on overall imports. The coverage ratio is just like the frequency index an aggregate indicator of the use and incidence of NTMs, but the NTMs are weighted by the imports. The coverage ratio is calculated as:

$$C_j = \frac{\sum NTM_i * V_i}{\sum V_i} * 100 \qquad (11)$$

Where NTM_i is a dummy variable reflecting the presence of at least one NTM for product i, V_i is the value of imports for product i. Problem with the coverage ratio is the potential endogeneity of the weights. The weights are the imports and these are probably correlated with the NTMs (that is at least what we want to quantify with our research question). Imports of products who are highly protected are likely to be small and will have a small weight in the coverage ratio, which would underestimate the restrictiveness of those NTMs (Kee et al, 2009). Ideally, this endogeneity is corrected by using trade levels that would arise in an NTM-free world. Because this is impossible, part of the problem can be solved by using trade values of the past (Fugazza, 2013).

The fifth indicator is the ad valorem equivalent (AVE) of NTMs. The AVE tries to summarise all the information about NTMs on tariff level in one measure. There are several methods in the literature to obtain AVEs, but we summarise the method of Kee et al (2009). The method is a two-step approach: the first step is an estimation of the quantity-impact of NTMs on imports and the second step is the transformation of quantity-impacts into price effects by using import demand elasticities. The first step is the estimation of the quantity-impact by using a gravity equation including two NTM related variables. There are two parameters introduced for the variable NTM to separate a product and a country-specific impact. The second step is the transformation into price-equivalent, what we call AVE of NTM. The first step in transforming quantity impacts into price impacts is to differentiate ln m with respect to NTM:

$$\frac{\partial \ln m}{\partial NTM} = \frac{\partial \ln m}{\partial \ln p} \frac{\partial \ln p}{\partial NTM} = \varepsilon \, AVE \quad (12)$$

Where p is the domestic price, m is imports and ε is the price elasticity. The second step is to solve (12) and we obtain:

$$AVE = \frac{1}{\varepsilon} \frac{\partial \ln m}{\partial NTM} = \frac{e^{\beta_{NTM}-1}}{\varepsilon} \quad (13)$$

AVEs are calculated per country at the tariff line level (Kee et al, 2009).

3.3 Illustration NTMs: cosmetics

The overall tariff rates for the chemical sector are very low for the EU and US. Both countries signed the Chemical Tariff Harmonisation Agreement (CTHA). The CTHA includes an agreement for the reduction of tariffs for cosmetics. The applied tariffs for chemicals under the CTHA are 0%, 5.5% and a maximum tariff of 6.5%. Tariffs are already low for chemical products and NTMs have become relatively more important for the chemical sector.

Ecorys surveyed companies for the Ecorys EU-US NTM study about regulatory divergence and restrictions that they face in doing business (Ecorys, 2009). We will summarize the response of one of the companies, which will give a good picture of the effect of NTMs. The response is about the regulatory divergence for lipstick between the EU and the US.

There are all sorts of regulatory divergences who can lead to NTMs, such as differences in domestic regulations. The two main regulations for chemicals are Regulation on registration, evaluation, authorisation and restriction of chemicals (REACH) for the EU and Toxic substances control act (TSCA) for the US. Lipsticks contain all kinds of chemicals. Under REACH is the industry responsible for ensuring that product substances do not adversely affect health or environment. The industry is responsible for the assessment plus its cost. REACH puts reliance on regulation and the TSCA puts more reliance on self-regulation. Other differences exist in the regulatory approach: the substances covered by the two regulatory systems are different. REACH requires more data than TSCA and places a higher burden on the manufactures. Also, REACH applies the principle “No data, no market”, while the US has the Environmental Protection Agency (EPA) who can place a chemical on a blacklist when the EPA identifies a risk (Searles, 2011).

There are some specific concerns for the cosmetics industry. Chemicals that need to be registered, are sometimes restricted and/or need to be authorized. Differences in regulations between the EU and US make it very difficult to produce a product that can be sold in both markets. The first illustration of this problem is the use of sunscreens in cosmetics. The use of some ingredients is prohibited or restricted by Food and Drug Administration (FDA) regulation. Sunscreens in cosmetics are classified as drugs by the FDA and are subject to the rules regarding the import and sale of drugs. Sunscreens are only allowed if they are non-therapeutic and –physiologic (for example when they are used as preservatives). In the EU, the use of sunscreens is allowed after undergoing stringent safety assessments. The second illustration is testing of cosmetics. In the EU, animal testing is prohibited. This is the opposite of the American policy, where animal testing is encouraged and prescribed. It is impossible to produce a product that fulfils both criteria at the same time.

Another sort of regulatory divergence lies in different labelling and packaging requirements. Information about the content of the product, date of durability, precautions for use, and a list of ingredients must be in accordance to the rules, who differ between the EU and US. Cosmetics who are distributed in the US, must comply to the regulations published by the FDA[6]. In the EU, the Cosmetics Directive 76/768/EEC is in force[7]. Labelling mistakes result in more than 22% of all detentions in the US[8]. Different labelling and packaging requirements are very costly, because specialists need to be hired to make compliance guides, warning letters and detailed reports about the labelling.

In section 3.1 we summarized the economic theory about the impact of NTMs on imports. The impact of an NTM is not always negative. The cosmetic producer does not experience the quantitative restrictions as restrictive. Other NTMs such as the animal testing involve high costs to fulfil the requirements. For the cosmetics sector, the total costs of NTMs are estimated to be an additional 10-20% on the product price. NTMs between the EU and US are economically more important than the remaining tariff levels even if intermediate products

[6] Summary of Labeling Requirements, http://www.fda.gov/Cosmetics/Labeling/Regulations/ucm126438.htm
[7] Cosmetics Directive, http://ec.europa.eu/consumers/sectors/cosmetics/documents/directive/index_en.htm#h2-consolidated-version-of-cosmetics-directive-76/768/eec
[8] U.S. FDA Labeling Regulations, http://www.registrarcorp.com/fda-labeling.jsp

and the final products may cross the Atlantic more than once; the tariff levels have a maximum of 6.5%. For producers it is very important that distorting NTMs are removed. A good step is that the US EPA is working on a TSCA reform (U.S. EPA, 2012). The proposed regulation will move in the direction of REACH. The new TSCA will hopefully lower the burden for producers who aim to sell their products in both the EU and US.

4. Data and measurement

In the chapters 2 and 3, the theoretical framework for testing the effect of Tariffs, Non Tariff Measures and Imports was built. Chapter 2 compares the different specifications theoretically. Chapter 3 describes the economic mechanism between trade and NTMs. In this chapter the data and its practical issues are discussed. This chapter will be concluded with a summary of the estimation framework.

4.1 Data

To be able to compare the different specifications to test the gravity equation, a large database has been constructed for this thesis. The sample covers bilateral imports between 159 countries of origin (exporters) and 31 countries of destination (importers) for a total of 1223 products. The list of countries of origin is reported in Appendix A, Table A.1 and the countries of destination are listed in Appendix A, Table A.2.

The standard gravity equation consists of bilateral trade flows, the mass of the country of origin and country of destination, and bilateral distance. The data is distributed over these three categories.

4.1.1 Bilateral trade flows

For the variable bilateral trade flows, we use imports expressed in thousands of US dollars as an indicator. We use imports rather than exports, because the variable NTM is import related. The data for the variable *Volume of imports* is obtained from the WITS database and COMTRADE. Imports are on 4-digit HS-code, which contains 1223 products. These databases contain only positive trade values. For our database, we changed the missing trade values in zero trade values. The assumption is that all missing observations are non-existing bilateral trade flows.

4.1.2 Mass

The variable mass is used as a variable that displays the potential to generate flows (exports) and attract flows (imports). In trade applications, mass is usually reflected by a country's GDP. Other possibilities are population and GDP per capita. GDP is chosen as indicator, because it is a better predictor for potential demand and supply, than population. *GDP exporter (ln)* and *GDP importer (ln)* denote the natural logarithms of the GDPs of the exporting and importing countries in the year 2011. The data for the variables GDP exporter and GDP importer was obtained from the IMF and is in billion current US Dollars.

4.1.3 Distance

To reflect the variable distance we do not only include a variable that reflects the geographical distance. In the dataset, we distinguish three types of distances: geographical distance, cultural distance and economic distance.

Geographical distance is covered by the variables *Geographical distance*, *Geographical distance weighted*, and *Common border dummy*. *Geographical distance* and *Geographical distance weighted* were obtained from the CEPII database. For the variable *geographical distance*, geodesic distances are calculated based on the great circle formula. *Geographical distance weighted* is based on bilateral distances between the biggest cities of those two countries, those inter-city distances being weighted by the share of the city in the overall country's population. A disadvantage of these geographic distance measures is that it is likely to overestimate the distance of trade between two neighbouring countries. Therefore, we also include the variable *Common border dummy*. The dummy is constructed based on information on land boundaries provided by the CIA World Factbook. The dummy takes the value 1 if two countries share a land border. Bilateral distances are not available for the EU. The Institut Géographique National calculated the geographical centre of the EU. For both EU-27 and EU-28, the geographical centre lies in Germany. For the variables *Geographical distance* and *Geographical distance weighted* we used bilateral distances for Germany as a proxy for the bilateral distances between the EU and its trading partners.

Cultural distance between countries is represented by the variables *Common language dummy* and *Common lang9 dummy*. Whether pairs of countries have the same language has been determined based on two different databases. The *Common language dummy* is a dummy based on official languages provided by the CIA World Factbook. The dummy takes the value 1 if both countries have the same official language. The *Common lang9 dummy* is a dummy variable obtained from the CEPII database. The dummy variable takes the value 1 if a language is spoken by at least 9% of the population in both countries. The EU has 24 official languages, but most of them are spoken by a very small part of the inhabitants. When we include all official EU country languages as EU languages, trade with the EU is almost always seen as trade with a country which has the same official language. In practice the inhabitants of a European country do not speak 24 languages, and it is likely that there is no common language. To circumvent this problem as good as possible, official EU languages are only English, French, German and Spanish. All four languages are spoken by at least 9% of the inhabitants of the EU (European Commission, 2012). English, French, German, and Spanish as EU languages are used for both language dummies.

Economic distance is defined as trade distorting measures, represented by the variables *Tariff (weighted average)* and *NTM dummy*. Tariffs and NTMs carry considerable costs to trade. Therefore, we include two variables that proxy these costs. *Tariff (weighted average)* is obtained from the TRAINS database via WITS. Tariffs are defined as effectively applied tariffs (AHS) and we use the weighted average of them. *Tariff (weighted average)* is given in percentage points and is the average of tariffs weighted by their corresponding trade value. In our analysis, the following transformation is done:

$$\ln(tariff_{ij}) \text{, where } tariff_{ij} = (100 + tariff_{ij}\ (weighted\ average))/100 \quad (14)$$

Less than fifteen percent of all (possible) bilateral trade flows has data about tariffs. Most of them (eighty percent) are tariffs applied to positive trade flows. When there is no tariff data available, this is seen as missing data. Including tariffs in the gravity equation will therefore dramatically decrease the zero trade flows with more than five million observations. It is possible that this will bias the results, because zero trade flows can be caused by high tariffs. These tariffs are not included, so they are not in the results.

NTM dummy is derived from NTM data provided by the Trade Analysis and Information System (TRAINS) database (UNCTAD). The NTM database contains data for 35 importing countries. The data for all countries is used, except Afghanistan, Guinea, Lao P.D.R., and Namibia (see Appendix A, Table A.2). All four countries have a political situation who suffers the reliability of the data, and the GDP for these four countries is very small. The UNCTAD uses the following definition of NTMs: “Any government action with a potential effect on the value, volume, or direction of trade. These are all barriers to international trade other than the tariffs for example, quotas, licensing, voluntary export restraints.” Data is available on 6-digit HS-code, but these are transformed to 4-digit HS-codes. NTMs are divided into 16 categories. Category P (export related measures) is removed from the database, so that we only have import related NTMs. The NTM dummy takes the value 1 if there is at least one NTM applicable. The consequence of the UNCTAD NTM definition, is that the presence of an NTM gives no information about the restrictiveness of the NTM. Several countries have NTMs in place for the world (country code 0). In the dataset, we interpreted this as an NTM for all individual countries.

After discussing the individual characteristics of the data and its practical issues, one general comment remains. The data availability is different per country and per variable. For some countries, it was impossible to use the same year for all variables. In Appendix B, Table B.1, a table with the years can be found. The combination of different years is not necessarily a problem, because the data does not change much per year.

4.2 Estimation framework

In chapter 2, we discussed the econometric approaches to estimate the gravity equation. The NBPM and ZINBPM are less restrictive compared to the Heckman model. Poisson models do not rely on stringent normality assumptions and they do not require a variable that influences the decision to trade but is unrelated to volume of trade (Burger et al, 2009). Therefore, the Heckman Model is not included in the analysis. The following five models are included in the analysis:

Table 4.1 Econometric approaches

(1) Ordinary Least Squares (OLS)

(2) OLS with small added number dependent variable

- (2a) Number: +1
- (2b) Number: +0.1
- (2c) Number: +0.01

(3) Poisson Model (PM)

(4) Negative Binomial Poisson Model (NBPM)

(5) Zero Inflated Negative Binomial Poisson Model (ZINBPM)

Section 4.1 discusses the data. Table 4.2 presents some descriptive statistics of the variables included in the gravity equation. See Appendix B for a full description of these variables and their sources.

Table 4.2 Descriptive statistics

	Mean	St.Dev.	Min.	Max.	N
Volume of imports (thousand USD)	2,014,306	4.96e+09	0	1.22e+13	6,069,040
GDP exporter (billion USD) (ln)	3.924	2.137	-1.766	9.651	5,885,623
GDP importer (billion USD) (ln)	4.872	1.995	2.293	9.777	6,069,040
Geographical distance (ln)	8.857	0.739	2.834	9.894	5,923,542
Weighted geographical distance (ln)	8.844	0.785	2.468	9.886	5,923,542
Common border dummy	0.024	0.152	0	1	6,069,040
Common language dummy	0.132	0.339	0	1	6,069,040
Common lang9 dummy	0.145	0.352	0	1	5,923,542
NTM dummy	0.674	0.469	0	1	6,059,703
Tariff (weighted average)	7.755	20.433	0	3,000	715,075

5. Analysis and results

5.1 Empirical results

This section presents the results for the five models. First, we present the results for Ordinary Least Squares (1). We extend this model with adding small numbers to the dependent variable (2). Poisson Models can naturally deal with zero trade flows. Model (3) is a standard Poisson Model. We will extend our analysis by using modified Poisson Models: Negative Binomial Poisson Model (4) to correct for overdispersion and Zero Inflated Negative Binomial Poisson Model (5) to correct for overdispersion and excess zeros. In section 4.1, we discussed the availability of data. Two variables, language and distance, are in two variants in the database. Before we discuss the results, we will pay attention to the choice of the variables in the model.

Language is available as a Common language dummy, which is based on official language, and as Common lang9 dummy, where common language is defined as the same language spoken by at least 9% of the population in both countries. Bilateral distance is available as Geographical distance and Geographical distance weighted. Geographical distance is the bilateral distance between the biggest cities of two countries. Geographical distances weighted are the bilateral distances weighted by the share of the city in the overall country's population. Because of correlation, we cannot use two measures at the same time. Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) and (Pseudo) R^2 are used to test which of the two is best to use in our model. For the language dummies, there is no clear indication that one is preferred over the other (see Table C.2, Appendix C). Therefore, we base the choice on the difference in definition. To capture the real effect of language on imports, spoken language is preferred over official language. Therefore, we use Common lang9 dummy in our model. AIC, BIC and (Pseudo) R^2 give a clear picture about the choice between distance and weighted distance. All AICs and BICs are smaller and the (pseudo) R^2 is higher for weighted distance (see Table C.3, Appendix C). We can conclude that Geographical distance weighted is preferred over Geographical distance.

As we have discussed the choice of the variables, we now go to the results of our base model. Model (1) in Table 5.1 shows the results of the estimation of the log-normal model (OLS), including fixed effects. Because of the nature of this model, the zero trade flows are omitted. In line with the literature, most variables have the correct sign. An increase of 1% in the GDP of the exporter and importer will increase imports by 0.6% in both cases. Distance has a negative effect on imports: a 1% increase in distance will decrease imports by 0.43%. The variables who describe cultural distance both have a significant positive effect: sharing a border increases imports with 66%, while having a common language increases imports with 7.5%. The effects of Tariffs and NTMs are opposite to each other. Tariffs have a negative effect on trade and NTMs have a positive effect on trade. All variables in the gravity equation are highly statistically significant.

In Model (1), zero trade flows are automatically excluded from the analysis, because of the logarithmic transformation of the dependent variable. Model (2) estimates the log-normal model after substituting the zeros by a small positive number. These small positive numbers are 1 (2a), 0.1 (2b), 0.01 (2c). All estimated effects are of the same sign as the variables in Model (1), but the effect sizes differ between the model with and without zeros. The smaller the added numbers, the more negative or positive are the estimates. The estimates for the variables common border and common language vary with more than 50% from the model without added numbers. These outcomes point in the direction that it is indeed possible to generate the parameter estimates whichever you like when you choose the constant. The estimation of Model (1) gives biased results because of omitting the zeros, but the proposed solution to add a small number is ad hoc and there is no justification for the choice of the value. We therefore need to estimate with alternative econometric approaches to avoid these problems.

The three problems by estimating the gravity equation with a log-linear structure are the violation of the homoskedasticity assumption, the way zero trade flows are dealt with and the logarithmic transformation of the dependent variable. As explained in chapter 2, Poisson Models have a solution for all these three problems. The dependent variable for the Models (3) – (5) is no longer the logarithm of imports, but is imports itself. In Model (3) the variable GDP importer is no longer significant and the common language dummy turns out to be negative. Tariffs have a negative effect on trade and NTMs have a positive effect on trade. All other variables have the same sign as in the models estimated by OLS and all variables are highly statistically significant. An assumption of the Poisson Model is equidispersion. In practice is the conditional variance often higher than the conditional mean (overdispersion). To control for overdispersion, we estimate in Model (4) the Negative Binomial Poisson Model. The overdispersion parameter α is estimated to be 2.031, which is highly significant. This means that NBPM is preferred over PM. In Model (3), the estimates for the variables

GDP importer and common language dummy changed compared to the first two models. These two variables change again in Model (4). GDP importer becomes significant negative and the common language dummy is again significant positive. (Negative Binomial) Poisson Models can technically deal with zeros, but they are not able to deal correctly with excess zeros. ZINBPM separates two kinds of zeros: true zeros and excess zeros. The first part of the model is a logit model who estimates a zero trade probability. The second step is an NBPM given that the zero trade flow has a non-zero probability. All variables are significant in the estimation of the value of trade, but distance and common border are not significant in estimating the probability to trade. The higher the tariffs imposed, the higher the probability of zero bilateral trade. An one percent increase of the tariff will increase the odds that the trade value will be a true zero by a factor of exp(1.323) = 3.755. Higher tariffs make it less likely that there is trade. The effect of NTMs is the other way around. Having at least one NTM in place will decrease the probability of zero trade. The odds that the trade value will be a true zero will decrease by a factor of exp(-0.379) = 0.685 if the NTM dummy takes the value 1. The second part is the estimation of the NBPM. The variables geographical distance, common border, common language, NTM and tariff have the same sign as in the Models (1) – (4). The signs for the variables GDP exporter and GDP importer changed compared to NBPM analysis in Model (4): GDP exporter has a significant negative effect on trade and GDP importer has a positive significant effect on trade.

5.2 Robustness checks

We now test the baseline results for robustness. We did already a sensitivity analysis by using different econometric approaches (see Section 5.1). In this paragraph, we consider the sensitivity of the data. First we want to investigate the effect of outlying observations. In our descriptive statistics (Table 4.2), it is striking that the maximum tariff is 3000%. If we look at the data, we see that Egypt is the only country with tariffs higher than 350%, with a maximum of 3000%. In the first robustness check, we remove Egypt from the database to check if tariffs still have a significant negative effect on imports. The results (see Table C.5) show that the effect of tariffs are stronger negative for all econometric approaches compared to the base model. We can conclude that the negative relationship between tariffs and trade is not confined to the high tariffs in Egypt.

Table 5.1 Results base model

	OLS (1) $Ln(T_{ij})$	OLS (2a) $Ln(T_{ij} + 1)$	OLS (2b) $Ln(T_{ij} + 0.1)$	OLS (2c) $Ln(T_{ij} + 0.01)$	PM (3) T_{ij}	NBPM (4) T_{ij}	ZINBPM (5) T_{ij}	
							Logit	Neg. Binomial
GDP exporter (billion current USD)	0.601 (0.003)***	0.590 (0.002)***	0.723 (0.002)***	0.838 (0.003)***	0.544 (0.000)***	0.413 (0.002)***	1.457 (0.101)***	-0.095 (0.013)***
GDP importer (billion current USD)	0.606 (0.007)***	0.582 (0.006)***	0.686 (0.007)***	0.772 (0.008)***	0.046 (3.859)	-0.134 (0.011)***	-0.985 (0.013)***	0.343 (0.002)***
Weighted geographical distance (ln)	-0.427 (0.008)***	-0.266 (0.006)***	-0.290 (0.007)***	-0.303 (0.008)***	-0.160 (0.000)***	-0.239 (0.006)***	-0.004 (0.045)	-0.150 (0.006)***
Common border dummy	0.657 (0.019)***	1.010 (0.016)***	1.253 (0.019)***	1.478 (0.023)***	0.431 (0.000)***	0.847 (0.016)***	0.035 (0.114)	0.938 (0.016)***
Common lang9 dummy	0.075 (0.012)***	0.107 (0.009)***	0.127 (0.011)***	0.140 (0.013)***	-0.093 (0.000)***	0.212 (0.010)***	1.462 (0.057)***	0.272 (0.010)***
NTM dummy	0.578 (0.011)***	0.481 (0.009)***	0.558 (0.011)***	0.612 (0.013)***	0.801 (0.000)***	0.865 (0.009)***	-0.379 (0.045)***	0.810 (0.009)***
Tariff (weighted average) (ln)	-4.391 (0.058)***	-3.024 (0.046)***	-3.517 (0.056)***	-3.773 (0.067)***	-12.978 (0.000)***	-1.885 (0.016)***	1.323 (0.129)***	-1.890 (0.016)***
N	536,613	657,978	657,978	657,978	657,978	657,978	657,978	657,978
Importer fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R^2/Pseudo R^2	0.212	0.212	0.214	0.209	0.295	0.030		
Log likelihood	-1,361,301	-1,605,562	-1,739,937	-1,849,180	-1.89e+10	-3,837,618	3,830,895	
Overdispersion						2.031***	1.991***	
AIC	2,722,675	3,211,199	3,479,948	3,698,434	3.79e+10	7,675,311	7,661,926	
BIC	2,723,089	3,211,620	3,480,370	3,698,856	3.79e+10	7,675,745	7,662,701	

Notes: All regressions include a constant. Standard errors are in parentheses. ***, **, * indicate significance at the 1%, 5%, 10% level.

Next, we analyse whether the results are sensitive to the inclusion of fixed effects. Testing the models without Fixed Effects enables us to carry out the Vuong test. The Vuong test is a test for excess zeros. With a Vuong statistic of 51.89, we have to reject the null-hypothesis that there is no indication of excess zeros. The zero-inflated model is preferred over the non-zero-inflated model. Removing fixed effects from the analysis does not change the results much. Possible reason is that with the inclusion of GDP exporter and GDP importer, already a variant of fixed effects is included. The results can be found in Table C.4.

In the baseline, we used English, French, German and Spanish as EU languages. Spanish is not always treated as a main European language. There are 16 trading partners of the EU who are seen as countries with a common language if we treat Spanish as a main European language. This is potentially important, so we want to test a dummy variable LangEU which is the same as Lang9 except for the EU, where Spanish is no longer treated as a European language. The effect of the dummy variable LangEU is positive for all models and the positive effect of a common language on trade is bigger compared to the Lang9 dummy variable (see Table C.6).

All models find a negative relationship between tariffs and imports, as we had expected to find. In the ZINBPM, we find that tariffs are not only reducing the volume of trade, but are also increasing the chance that there is no trade at all. For NTMs, we find opposite results. The relationship between NTMs and imports is positive and NTMs are decreasing the chance that there is a zero trade flow. In theory, it is possible that we find a positive effect for NTMs on trade. In our discussion of the economic theory (see Section 3.1), we discussed the different NTMs and their impact on trade. NTMs can be demand-enhancing when they provide valuable information to the consumer, but in theory most NTMs reduce trade. Based on the theoretical analysis of the impact of NTMs on imports, we do not expect to find a significant positive relation between these two variables.

There are several reasons to believe that NTMs can still be trade distorting, even by finding this paradoxical result. It is possible that the data is causing this unexpected effect. The first reason lies in the definition used by the UNCTAD. The definition is: “Any government action with a potential effect on the value, volume, or direction of trade. These are all barriers to international trade other than the tariffs for example, quotas, licensing, voluntary export restraints.” The problem with this definition is, that there is no information about the regulatory divergence between countries. When two countries have exactly the same regulations for a specific product, the NTM dummy is for both countries one, but bilateral trade between these two countries is not distorted by the NTMs. These NTMs are only distorting for countries with other regulations. It is very costly to fulfill the requirements for producers in countries with other regulations. The second reason is the use of a dummy. The NTM dummy takes the value 1 if there is at least one NTM applicable and 0 if there is no NTM in place. In section 3.2 we discussed the advantages and disadvantages of the different measurement of NTMs. Shortcoming of the usage of a dummy variable is that it is impossible to identify the individual effect of an NTM when there are more NTMs in place. Another shortcoming is that you lose information about the number of NTMs. A dummy variable cannot make a distinction between products with one NTM and products who are heavily regulated by a lot of NTMs. A dummy variable does not do any justice to the diversity in the number of NTMs and the restrictiveness of NTMs. The third possible reason for finding a positive relation between NTMs and imports, is that the data from the UNCTAD is based on information provided by exporting countries. NTMs are often not reported when there is no trade flow, while these NTMs can be the reason that there is no trade. A fourth reason for the results is that developed countries usually have higher quality standards than developing countries and are trading more. This reverse causality can influence the effect found. A fifth reason is that we lose millions of data points by including tariffs as a control variable.

When we remove tariffs from the analysis, we can use the full dataset with more than 5 million data points. Most of these data points are zero trade flows. When we look to the results (see Table C.7), we see that having an NTM increases the chance to have a zero trade flow. This means that NTMs form trade barriers. The positive effect that we find for the non-zero trade flows can be explained by the logic that when an exporter satisfies the rules, the NTM no longer influences his volume of trade when there is no quantity restriction. Using the full dataset does not change the signs of the remaining variables and all variables are still highly significant.

The last robustness check is not on a scientific basis, but is only done to check the robustness of the positive relation between NTMs and imports. In the above, we explained four possible reasons that we find a positive relationship, even when NTMs can be trade distorting. When the NTM dummy is 1 on a bilateral zero trade flow, the NTMs on this product are probably more distorting than NTMs on a product with a positive trade flow. We want to know what the effect is when we make a distinction in the group with a value 1 for the NTM dummy. Within this group, we change all values into 0.5 for the positive trade flows. This means that we get three groups: dummy=0 when there is no NTM in place, dummy=0.5 if there is at least one NTM and the trade value is positive, dummy=1 if there is at least one NTM and the trade value is zero. When we look to the results, we see that nothing changed for Method 1, because the zero trade flows are not included in the analysis. The effect of the NTM dummy doubled, but the difference disappears when we multiply by 0.5 for the interpretation. For the models with zero trade flows in the analysis, the effect of NTMs turns out to be significantly negative. Tariffs remain significantly negative and all variables have the same sign as in the base model. Even though the changing results for NTMs are not scientific based, it shows that making a distinction in the trade restrictiveness of NTMs can change the results dramatically.

After checking the robustness of the results with different econometric approaches, different definitions of variables and the effect of outlying observations, we can conclude that GDP exporter, GDP importer, having a common border and common language are positive related to imports and that distance and tariffs are negative related to imports. NTMs are positively related to imports and the probability of trade on our base model, but the robustness checks show that this is possibly not the real effect of NTMs on imports.

6. Conclusions and recommendations for further research

Since trade is more and more characterised by Non Tariff Measures instead of Tariffs, it is increasingly important to understand the effects of Tariffs and Non Tariff Measures on trade. The overall research question of this paper is: What is the effect of Tariffs and Non Tariff Measures on Imports as found by using different econometric specifications of the gravity equation? The answer to this question comes from two distinct parts of the paper: a discussion of the economic theory and econometric approaches and an empirical investigation of the econometric approaches.

Economic theory predicts a negative relation between tariffs and imports, because tariffs raise the prices of imported goods. Higher prices of the imported goods reduce the volume of imports. The effect of NTMs is inconclusive, based on the theory that NTMs can have negative effects, zero effects, and positive effects. Positive effects are possible when NTMs are demand-enhancing by giving valuable information about products. When there are several NTMs in place for one product level, the overall impact is related to the relative strength of the different NTMs in place.

Using a comprehensive dataset, we estimated the gravity equation with five econometric approaches. These five approaches are OLS, OLS with small added number for dependent variable, Poisson Model, Negative Binomial Poisson Model, and Zero Inflated Negative Binomial Poisson Model. Estimation with OLS has three problems: violation of the homoskedasticity assumption, the way that zero trade flows are dealt with and the logarithmic transformation of the dependent variable. Poisson Models have a solution for all these three problems, but PMs assume equidispersion and no excess zeros. It is likely that our database suffers from overdispersion and excess zeros. To correct for these shortcomings, we also estimate NBPM and ZINBPM. ZINBPM is estimated in two steps: the first step is the estimation of a logit model for the probability of zero trade. The second step is an NBPM given that the zero trade flow has a non-zero probability. All econometric approaches find a negative relationship between tariffs and imports. Tariffs are not only decreasing trade, but also raise the probability of a zero trade flow. For NTMs we find a positive relation with trade and NTMs are decreasing the probability of no trade.

Using different econometric approaches is a robustness check of the method, but we also did sensitivity analysis of the data. The effect of outlying observations is checked by removing Egypt from the analysis. For the variables common language and bilateral distance we used different definitions in the analysis. All these checks show that the results are robust, because the outcomes barely change. Based on the data, we can conclude that GDP exporter, GDP importer, having a common border and common language are positive related to imports and that distance and tariffs are negative related to imports. Based on the literature, we did not expect to find a significant positive relation between NTMs and imports. It is possible that NTMs are less trade distorting than other studies found. In our study, the effect of NTMs on imports is an overall effect. Individual effects of several types of NTMs can still be negative. A possible reason of the positive effect lies in the data used. The definition of UNCTAD gives no information about the regulatory divergence between countries, the data is based on information provided by exporting countries, we cannot make a distinction by the number and restrictiveness of NTMs when we use a dummy variable, the analysis probably suffers from reverse causality, and we are losing millions of data points by including tariffs as a control variable. In a robustness check, we removed tariffs from the analysis to be able to use the full dataset. Using the full dataset changes the result that NTMs decrease the probability of trade. Now we find that having an NTM increases the probability to find a zero trade flow, which means that NTMs form trade barriers. All other results remain almost the same. The last robustness check is checking the effect of changing the values 1 into 0.5 for the NTM dummy by positive trade values. Tariffs remain significantly negative, but the effect of NTMs turns out to be negative as well. All variables keep the same sign as in the base model. Our findings are important for trade policy. The results suggest to further reduce tariffs, and increase trade facilitating NTMs.

Based on the analysis, we have several recommendations for future research. Our biggest concern is the dummy variable NTM which is based on the TRAINS database. A dummy variable does not do any justice to the diversity in sorts of NTMs, the number of NTMs, and the restrictiveness of NTMs. An area for further research is to break down the overall effect of NTMs into individual effects of the different categories to get more insight in the trade distorting effects per NTM-category.

Another area for further research is related to the database. The analysis probably suffers from combining data from different years. It was impossible to get reliable data for all variables for the year 2012. For some countries, it takes several years to present reliable data. In the future, it must be possible to adjust the database. Because of data limitations, it was also impossible to do panel data analysis. Panel data can give more insight in the relation between Tariffs, NTMs and Imports. When information on the development of NTMs is published, panel data analysis becomes possible.

The results for all variables, except NTMs, are robust. The problems with the NTM dummy are a result of limitations imposed by the dataset and by using a dummy. Hence, future research could include data from the same years, breaking down the dummy into individual effects, and a different measurement of NTMs and probably new data on NTMs. A promising development is the NTM Impact project from the European Commission. Nevertheless, this paper provides new insights in the empirical evidence of the gravity equation on product level, it uses the most recent data available on NTMs, and it provides insight into the econometric issues regarding different approaches to measure the gravity equation. These advancements provide the motivation for future research on the relation between Tariffs, NTMs, and Imports by extending the analysis with new data and another measurement of NTMs.

References

Anderson, J. (1979). *A theoretical foundation for the gravity equation*, American Economic Review, 1979, 69(1), pp. 106-116.

Anderson, J. E. & Van Wincoop, E. (2003). *Gravity with gravitas: a solution to the border puzzle*, American Economic Review, 93, 170-192.

Baier, S. & Bergstrand, J. (2009). *Bonus vetus OLS: A simple method for approximating international trade-cost effects using the gravity equation*, Journal of International Economics, Elsevier, vol. 77(1), February, pp. 77-85.

Bergstrand, J. (1985). *The gravity equation in international trade: some microeconomic foundations and empirical evidence*, Review of Economics and Statistics, 1985, 67:3, August, pp. 474-481.

Bergstrand J., Egger, P. & Larch, M. (2013). *Gravity Redux: Estimation of gravity-equation coefficients, elasticities of substitution, and general equilibrium comparative statics under asymmetric bilateral trade costs*, Journal of International Economics 89 (2013), pp. 110–121.

Bowen, H.P., Hollander, A., & Viaene, J.-M. (2012). *Applied International Trade*, 2nd edition, Palgrave Macmillan.

Burger, M., Van Oort, F. & Linders, G. (2009). *On the Specification of the Gravity Model of Trade: Zeros, Excess Zeros and Zero-inflated Estimation*, Spatial Economic Analysis, 4:2, pp. 167-190.

Cameron, A. & Trivedi, P. (1986). *Econometric Models Based on Count Data: Comparisons and Applications of some Estimators and Tests*, Journal of Applied Econometrics 1, pp. 29-53.

Demaria, F., Rau, M. & Schlueter, S. (2011). *NTMs and gravity-type models: state of the art and analysis of the literature*, EC FP7 project 'NTM impact', NTM impact working paper 11/01.

Disdier, A. & Marette, S. (2010). *The combination of gravity and welfare approaches for evaluating non tariff measures*, American Journal of Agricultural Economics. 92(3), pp. 713–726.

Ecorys (2009). *Non-tariff measures in EU-US trade and investment – an economic analysis,* European Commission, Directorate-General for Trade, Reference: OJ 2007/S 180-219493.

European Commission (2012). *Europeans and their languages*, European Commission, Special Eurobarometer 386.

Frank, R. (2008). *Microeconomics and behavior*, seventh edition, McGraw-Hill.

Fugazza, M. (2013). *The economics behind non-tariff measures: theoretical insights and empirical evidence*, Unctad, Policy issues in international trade and commodities, study series No. 57, pp. 1-26.

Gómez-Herrera, E. (2012). *Comparing alternative methods to estimate gravity models of bilateral trade,* Empir Econ (2013) 44, pp. 1087–1111.

Heckman, J. (1979). *Sample selection bias as a specification error*, Econometrica, 47, pp. 153-161.

Helpman, E. & Krugman, P. (1985). *Market structure and foreign trade*, Cambridge, MA: MIT Press, 1985.

Hoekman, B. & Nicita, A. (2008). *Trade policy, trade costs, and developing country trade,* The World Bank Policy Research Paper 4797.

Kee, H., Nicita, A. & Olarreaga, M. (2009). *Estimating trade restrictiveness indices*, The Economic Journal, 119 (January), pp. 172–199.

Linders, G. & De Groot, H. (2006). *Estimation of the Gravity Equation in the Presence of Zero Flows,* Tinbergen Institute Discussion Paper 2006-072/3.

Reinert, K.A. (2006). *Gravity models.* Princeton Press, pp. 567-570.

Santos Silva, J. & Tenreyro, S. (2006). *The Log of Gravity,* The Review of Economics and Statistics, MIT Press, vol. 88(4), November, pp. 641-658.

Searles, J. (2011). *Chemicals Regulation: A comparison of US and European Approaches*. Steptoe and Johnson LLP.

Siliverstovs, B. & Schumacher, D. (2007). *Estimating Gravity Equations: To Log or not to Log?*, DIW Berlin Discussion paper 739.

Tinbergen, J. (1962). *Shaping the World Economy*, New York, Twentieth Century Fund.

UNESCAP (2008). *Gravity Models: Theoretical Foundations and related estimation issues.*

U.S. Environmental Protection Agency (2012). *Essential Principles for Reform of Chemicals Management Legislation*. EPA

Van Marrewijk, C. (2012). *International Economics: theory, application, and policy*, 2nd edition, Oxford University Press.

Verbeek, M. (2012). *A Guide to Modern Econometrics*, 4th edition, John Wiley & Sons Ltd.

Vuong, Q. (1989). *Likelihood ratio tests for model selection and non-nested hypotheses*, Econometrica, 57, pp. 307-333.

WTO (2012). *World trade report 2012, Trade and public policies: A closer look at non-tariff measures in the 21st century*. WTO Publications.

Appendix A: Countries Included

Table A.1 Countries included as countries of origin

Afghanistan	Ethiopia	Mexico
Albania	Fiji	Moldova
Algeria	Finland	Mongolia
Angola	France	Morocco
Argentina	Gabon	Mozambique
Armenia	Georgia	Myanmar
Australia	Germany	Namibia
Austria	Ghana	Nepal
Azerbaijan	Greece	Netherlands
Bahrain	Guatemala	New Zealand
Bangladesh	Guinea	Nicaragua
Barbados	Guinea-Bissau	Niger
Belarus	Guyana	Nigeria
Belgium	Haiti	Norway
Belize	Honduras	Oman
Bolivia	Hong Kong SAR	Pakistan
Bosnia and Herzegovina	Hungary	Panama
Botswana	Iceland	Papua New Guinea
Brazil	India	Paraguay
Bulgaria	Indonesia	Peru
Burkina Faso	Iraq	Philippines
Burundi	Ireland	Poland
Cambodia	Islamic Republic of Iran	Portugal
Cameroon	Israel	Qatar
Canada	Italy	Romania
Cape Verde	Jamaica	Russia
Central African Republic	Japan	Rwanda
Chad	Jordan	Saudi Arabia
Chile	Kazakhstan	Senegal
China	Kenya	Sierra Leone
Colombia	Kuwait	Singapore
Comoros	Lao P.D.R.	Slovak Republic
Costa Rica	Latvia	Slovenia
Côte d'Ivoire	Lebanon	South Africa
Croatia	Lesotho	Spain
Cyprus	Liberia	Sri Lanka
Czech Republic	Libya	Sudan
Denmark	Lithuania	Suriname
Djibouti	Luxembourg	Swaziland
Dominica	Madagascar	Sweden
Dominican Republic	Malawi	Switzerland
Ecuador	Malaysia	Syria
Egypt	Mali	Tajikistan
El Salvador	Malta	Tanzania
Equatorial Guinea	Marshall Islands	Thailand
Eritrea	Mauritania	The Bahamas

Estonia	Mauritius	The Gambia
Togo	Uganda	Uzbekistan
Tonga	Ukraine	Venezuela
Trinidad and Tobago	United Arab Emirates	Vietnam
Tunisia	United Kingdom	Yemen
Turkey	United States	Zambia
Turkmenistan	Uruguay	Zimbabwe

Table A.2 Countries included as countries of destination

Argentina	European Union	Pakistan
Bolivia	Guatemala	Paraguay
Brazil	India	Peru
Burkina Faso	Japan	Senegal
Chile	Kazakhstan	Sri Lanka
China	Lebanon	Tanzania
Colombia	Madagascar	Tunisia
Costa Rica	Mauritius	Uruguay
Cote d'Ivoire	Mexico	Venezuela
Ecuador	Morocco	
Egypt	Nepal	

Appendix B: Data

Dependent variable:

- *Volume of imports (1000 USD)* was obtained from the WITS and UN COMTRADE databases. For each country pair are two possible observations: Tradevalue_{ij} and Tradevalue_{ji}, where $T_{ij} \neq T_{ji}$. The first subscript is the source country and the second is the destination country. Because of data limitations, the year for which the data is used varies between 2009-2012 (see Table B.1). The WITS database is used for the year 2012 and COMTRADE for the remaining years. COMTRADE provides trade data in USD, so we divided the numbers by 1000 to make the data comparable. The focus is on imports for a set of 159 countries (see Appendix A, table A.1). Imports are on 4-digit HS-code, which contains 1223 products.

Independent variables:

- *GDP exporter (billion current USD)* and *GDP importer (billion current USD)* were obtained from the IMF. We have chosen to use data for the year 2011, because there is not enough (reliable) data available for the year 2012. The IMF estimated the GDP for the year 2011 for the countries: Albania, Botswana, Burkina Faso, Burundi, Cameroon, Cape Verde, Chad, Djibouti, Equatorial Guinea, Eritrea, Gabon, Guinea, Iran, Iraq, Lebanon, Libya, Madagascar, Malawi, Mauritania, Namibia, Niger, Sudan, Swaziland, Syria, Tanzania, Venezuela, and Yemen.
- *Geographical distance* and *Geographical distance weighted* were obtained from the CEPII database dist_cepii.xls, which contains bilateral data for 226 countries/regions. For the variable *geographical distance*, geodesic distances are calculated based on the great circle formula, which uses latitudes and longitudes of the most important cities/agglomerations (in terms of population). *Geographical distance weighted* is based on bilateral distances between the biggest cities of those two countries, those inter-city distances being weighted by the share of the city in the overall country's population. See formula B.1, where pop_k means the population of agglomeration k belonging to country i. Parameter θ measures the sensitivity of trade flows to bilateral distance d_{kl}. Θ is set equal to -1, which corresponds to the usual coefficient in gravity models. Bilateral data is available for 221 countries/regions. [9]

[9] http://mpra.ub.uni-muenchen.de/26469/1/noticedist_en.pdf

$$d_{ij} = \left(\sum_{k\in i}(\frac{pop_k}{pop_i})\sum_{l\in j}(\frac{pop_l}{pop_j})d_{kl}^{\theta}\right)^{1/\theta} \quad \text{(B.1)}$$

- *Common border dummy* is constructed based on information on land boundaries provided by the CIA World Factbook. The dummy takes the value 1 if two countries share a land border.
- *Common language dummy* is a dummy based on official languages provided by the CIA World Factbook. The dummy takes the value 1 if two countries have the same official language.
- *Common lang9 dummy* is a dummy variable obtained from the CEPII database dist_cepii.xls. The dummy variable takes the value 1 if a language is spoken by at least 9% of the population in both countries[10].
- *NTM dummy* is derived from NTM data provided by the Trade Analysis and Information System (TRAINS) database (UNCTAD). The NTM database contains data for 35 importing countries. The data for all countries is used, except Afghanistan, Guinea, Lao P.D.R., and Namibia (see Appendix A, Table A.2). The UNCTAD uses the following definition of NTMs: Any government action with a potential effect on the value, volume, or direction of trade. These are all barriers to international trade other than the tariffs for example, quotas, licensing, voluntary export restraints[11]. Data is available on 6-digit HS-code, but these are transformed to 4-digit HS-codes. NTMs are divided into 16 categories[12]. Category P (export related measures) is removed from the database, so that we only have import related NTMs. The NTM dummy takes the value 1 if there is at least one NTM applicable. Data is only available for one specific year per country (see Table B.1).
- *Tariff (weighted average)* is obtained from the TRAINS database via WITS. Tariffs are defined as effectively applied tariffs (AHS) and we use the weighted average of them. Tariff (weighted average) is given in percentage points and is the average of tariffs weighted by their corresponding trade value[13]. Because of data limitations, the year for which data is used varies between 2007-2012 (see Table B.1).

[10] http://www.cepii.fr/distance/noticedist_en.pdf
[11] http://wits.worldbank.org/wits/glossary.html
[12] See footnote 3
[13] http://wits.worldbank.org/WITS/wits/WITSHELP/Content/Data_Retrieval/P/AQ/C9e.AQ_Tariffs-Result.htm

Table B.1 Year of data per variable

Land	**Import**	**NTMs**	**Tariffs**
Argentina	2012	2012	2011
Bolivia	2012	2012	2011
Brazil	2012	2012	2011
Burkina Faso	2011	2012	2011
Chile	2012	2012	2010
China	2012	2012	2011
Colombia	2012	2012	2011
Costa Rica	2012	2012	2010
Côte d'Ivoire	2011	2012	2011
Ecuador	2011	2012	2011
Egypt	2011	2011	2009
EU	2011	2010	2011
Guatemala	2012	2012	2011
India	2011	2012	2009
Japan	2009	2009	2011
Kazakhstan	2012	2012	2011
Lebanon	2012	2011	2007
Madagascar	2012	2011	2011
Mauritius	2012	2011	2011
Mexico	2012	2012	2010
Morocco	2011	2011	2009
Nepal	2011	2012	2011
Pakistan	2012	2012	2009
Paraguay	2012	2012	2011
Peru	2012	2012	2011
Senegal	2011	2012	2011
Sri Lanka	2011	2012	2011
Tanzania	2011	2011	2011
Tunisia	2011	2011	2008
Uruguay	2012	2012	2011
Venezuela	2011	2012	2011

Note: Because data is used from several databases, some adjustments in the country codes are made to be able to put data together. These adjustments are made for the countries Ethiopia (231), France (250), Germany (276), India (699), Italy (380), Norway (578), Panama (591), Switzerland (756), United States of America (840), Yemen (887).

Appendix C: Results

Table C.1 Results base model

	OLS (1)	OLS (2a)	OLS (2b)	OLS (2c)	PM (3)	NBPM (4)	ZINBPM (5)	
	$Ln(T_{ij})$	$Ln(T_{ij} + 1)$	$Ln(T_{ij} + 0.1)$	$Ln(T_{ij} + 0.01)$	T_{ij}	T_{ij}	T_{ij}	
							Logit	Neg. Binomial
GDP exporter (billion current USD)	0.601	0.590	0.723	0.838	0.544	0.413	1.457	-0.095
	(0.003)***	(0.002)***	(0.002)***	(0.003)***	(0.000)***	(0.002)***	(0.101)***	(0.013)***
GDP importer (billion current USD)	0.606	0.582	0.686	0.772	0.046	-0.134	-0.985	0.343
	(0.007)***	(0.006)***	(0.007)***	(0.008)***	(3.859)	(0.011)***	(0.013)***	(0.002)***
Weighted geographical distance (ln)	-0.427	-0.266	-0.290	-0.303	-0.160	-0.239	-0.004	-0.150
	(0.008)***	(0.006)***	(0.007)***	(0.008)***	(0.000)***	(0.006)***	(0.045)	(0.006)***
Common border dummy	0.657	1.010	1.253	1.478	0.431	0.847	0.035	0.938
	(0.019)***	(0.016)***	(0.019)***	(0.023)***	(0.000)***	(0.016)***	(0.114)	(0.016)***
Common lang9 dummy	0.075	0.107	0.127	0.140	-0.093	0.212	1.462	0.272
	(0.012)***	(0.009)***	(0.011)***	(0.013)***	(0.000)***	(0.010)***	(0.057)***	(0.010)***
NTM dummy	0.578	0.481	0.558	0.612	0.801	0.865	-0.379	0.810
	(0.011)***	(0.009)***	(0.011)***	(0.013)***	(0.000)***	(0.009)***	(0.045)***	(0.009)***
Tariff (weighted average) (ln)	-4.391	-3.024	-3.517	-3.773	-12.978	-1.885	1.323	-1.890
	(0.058)***	(0.046)***	(0.056)***	(0.067)***	(0.000)***	(0.016)***	(0.129)***	(0.016)***
N	536,613	657,978	657,978	657,978	657,978	657,978	657,978	657,978
Importer fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R^2/Pseudo R^2	0.212	0.212	0.214	0.209	0.295	0.030		
Log likelihood	-1,361,301	-1,605,562	-1,739,937	-1,849,180	-1.89e+10	-3,837,618	3,830,895	
Overdispersion						2.031***	1.991***	
AIC	2,722,675	3,211,199	3,479,948	3,698,434	3.79e+10	7,675,311	7,661,926	
BIC	2,723,089	3,211,620	3,480,370	3,698,856	3.79e+10	7,675,745	7,662,701	

Notes: All regressions include a constant. Standard errors are in parentheses. ***, **, * indicate significance at the 1%, 5%, 10% level.

Table C.2 Comparison variables Lang and Lang9

	AIC		BIC		(Pseudo) R^2	
	Lang	Lang9	Lang	Lang9	Lang	Lang9
OLS	2,722,605	2,722,675	2,723,019	2,723,089	0.212	0.212
PM	3.79e+10	3.79e+10	3.79e+10	3.79e+10	0.295	0.295
NBPM	7,675,279	7,675,311	7,675,713	7,675,745	0.030	0.030
ZINBPM	7,674,384	7,661,926	7,675,159	7,662,701	x	x

Table C.3 Comparison variables Dist and Distw

	AIC		BIC		(Pseudo) R^2	
	Dist	Distw	Dist	Distw	Dist	Distw
OLS	2,722,991	2,722,675	2,723,405	2,723,089	0.212	0.212
PM	3.78e+10	3.79e+10	3.78e+10	3.79e+10	0.296	0.295
NBPM	7,675,603	7,675,311	7,676,036	7,675,745	0.030	0.030
ZINBPM	7,663,480	7,661,926	7,664,255	7,662,701	x	x

Table C.4 Results without using Fixed Effects

	OLS (1) $Ln(T_{ij})$	OLS (2a) $Ln(T_{ij} + 1)$	OLS (2b) $Ln(T_{ij} + 0.1)$	OLS (2c) $Ln(T_{ij} + 0.01)$	PM (3) T_{ij}	NBPM (4) T_{ij}	ZINBPM (5) T_{ij}	
							Logit	Neg. Binomial
GDP exporter (billion current USD)	0.637	0.612	0.751	0.869	0.553	0.434	-0.150	0.670
	(0.003)***	(0.002)***	(0.002)***	(0.003)***	(0.000)***	(0.002)***	(0.031)***	(0.002)***
GDP importer (billion current USD)	0.488	0.414	0.471	0.509	0.644	0.700	-1.111	0.429
	(0.002)***	(0.002)***	(0.002)***	(0.002)***	(0.000)***	(0.002)***	(0.025)***	(0.002)***
Weighted geographical distance (ln)	-0.528	-0.353	-0.402	-0.435	-0.200	-0.503	0.941	-0.498
	(0.007)***	(0.005)***	(0.006)***	(0.008)***	(0.000)***	(0.005)***	(0.116)***	(0.005)***
Common border dummy	0.404	0.825	1.021	1.213	0.279	0.288	-19.694	0.290
	(0.018)***	(0.015)***	(0.019)***	(0.022)***	(0.000)***	(0.015)***	(4484.566)	(0.015)***
Common lang9 dummy	-0.005	0.035	0.044	0.047	-0.287	0.006	-0.142	0.008
	(0.011)	(0.009)***	(0.011)***	(0.013)***	(0.000)***	(0.009)	(0.174)	(0.009)
NTM dummy	0.329	0.226	0.249	0.253	0.724	0.697	0.065	0.696
	(0.010)***	(0.008)***	(0.010)***	(0.011)***	(0.000)***	(0.008)***	(0.219)	(0.008)***
Tariff (weighted average) (ln)	-4.182	-2.458	-2.758	-2.796	-11.094	-1.691	-13.791	-1.692
	(0.053)***	(0.043)***	(0.053)***	(0.063)***	(0.000)***	(0.014)***	(2.037)***	(0.014)***
N	536,613	657,978	657,978	657,978	657,978	657,978	657,978	657,978
Importer fixed effects	No	No	No	No	No	No	No	No
Adjusted R^2/Pseudo R^2	0.193	0.190	0.178	0.163	0.288	0.028		
Log likelihood	-1,367,768	-1,616,028	-1,754,364	-1,867,903	-1.91e+10	-3,842,319	-3,842,123	
Overdispersion						2.042***	2.036***	
Vuong							51.89***	
AIC	2,735,551	3,232,073	3,508,744	3,735,822	3.83e+10	7,684,656	7,684,280	
BIC	2,735,641	3,232,164	3,508,835	3,735,913	3.83e+10	7,684,758	7,684,474	

Notes: All regressions include a constant. Standard errors are in parentheses. ***, **, * indicate significance at the 1%, 5%, 10% level.

Table C.5 Results without using Egypt

	OLS (1) $Ln(T_{ij})$	OLS (2a) $Ln(T_{ij} + 1)$	OLS (2b) $Ln(T_{ij} + 0.1)$	OLS (2c) $Ln(T_{ij} + 0.01)$	PM (3) T_{ij}	NBPM (4) T_{ij}	ZINBPM (5) T_{ij}	
							Logit	Neg. Binomial
GDP exporter (billion current USD)	0.602 (0.003)***	0.587 (0.002)***	0.718 (0.002)***	0.831 (0.003)***	0.546 (0.000)***	0.440 (0.002)***	0.213 (0.071)***	-0.057 (0.014)***
GDP importer (billion current USD)	0.598 (0.007)***	0.576 (0.006)***	0.679 (0.007)***	0.764 (0.008)***	-0.023 (0.000)***	-0.167 (0.010)***	-1.003 (0.012)***	0.236 (0.002)***
Weighted geographical distance (ln)	-0.421 (0.008)***	-0.245 (0.006)***	-0.261 (0.007)***	-0.265 (0.009)***	-0.159 (0.000)***	-0.307 (0.006)***	-0.038 (0.040)	-0.085 (0.006)***
Common border dummy	0.640 (0.020)***	1.027 (0.016)***	1.281 (0.020)***	1.518 (0.023)***	0.430 (0.000)***	0.679 (0.016)***	-0.412 (0.121)***	0.930 (0.016)***
Common lang9 dummy	0.098 (0.012)***	0.136 (0.010)***	0.164 (0.012)***	0.182 (0.014)***	-0.092 (0.000)***	0.155 (0.010)***	1.397 (0.053)***	0.185 (0.010)***
NTM dummy	0.595 (0.012)***	0.515 (0.010)***	0.602 (0.012)***	0.665 (0.014)***	0.810 (0.000)***	0.916 (0.009)***	-0.496 (0.043)***	0.696 (0.009)***
Tariff (weighted average) (ln)	-5.086 (0.063)***	-3.623 (0.051)***	-4.233 (0.062)***	-4.561 (0.074)***	-13.163 (0.000)***	-4.091 (0.030)***	1.831 (0.202)***	-4.236 (0.030)***
N	521,636	637,359	637,359	637,359	637,359	637,359	637,359	637,359
Importer fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R^2/Pseudo R^2	0.216	0.221	0.219	0.214	0.295	0.031		
Log likelihood	-1,323,491	-1,553,112	-1,682,494	-1,787,534	-1.88e+10	-3,713,962	-3,712,911	
Overdispersion						2.017***	1.989***	
AIC	2,647,054	3,106,296	3,365,060	3,575,140	3.76e+10	7,427,997	7,425,954	
BIC	2,647,456	3,106,705	3,365,469	3,575,549	3.76e+10	7,428,418	7,426,704	

Notes: All regressions include a constant. Standard errors are in parentheses. ***, **, * indicate significance at the 1%, 5%, 10% level.

Table C.6 Results using other language EU

	OLS (1) Ln(T_{ij})	OLS (2a) Ln(T_{ij} + 1)	OLS (2b) Ln(T_{ij} + 0.1)	OLS (2c) Ln(T_{ij} + 0.01)	PM (3) T_{ij}	NBPM (4) T_{ij}	ZINBPM (5) T_{ij}	
							Logit	Neg. Binomial
GDP exporter (billion current USD)	0.600 (0.003)***	0.590 (0.002)***	0.723 (0.002)***	0.838 (0.003)***	0.544 (0.000)***	0.413 (0.002)***	-0.384 (0.097)***	-0.079 (0.014)***
GDP importer (billion current USD)	0.608 (0.007)***	0.584 (0.006)***	0.688 (0.007)***	0.773 (0.008)***	-0.019 (4.419)	-0.134 (0.011)***	-1.017 (0.014)***	0.303 (0.002)***
Weighted geographical distance (ln)	-0.413 (0.008)***	-0.253 (0.006)***	-0.276 (0.007)***	-0.288 (0.008)***	-0.154 (0.000)***	-0.237 (0.006)***	-0.040 (0.044)	-0.123 (0.006)***
Common border dummy	0.668 (0.019)***	1.019 (0.016)***	1.262 (0.019)***	1.487 (0.023)***	0.451 (0.000)***	0.845 (0.016)***	-0.095 (0.120)	0.960 (0.016)***
Common LangEU dummy	0.129 (0.012)***	0.163 (0.009)***	0.187 (0.012)***	0.202 (0.014)***	0.010 (0.000)***	0.220 (0.010)***	1.328 (0.056)***	0.276 (0.010)***
NTM dummy	0.577 (0.011)***	0.480 (0.009)***	0.557 (0.011)***	0.611 (0.013)***	0.799 (0.000)***	0.863 (0.009)***	-0.423 (0.045)***	0.766 (0.090)***
Tariff (weighted average) (ln)	-4.384 (0.058)***	-3.017 (0.046)***	-3.510 (0.056)***	-3.766 (0.067)***	-12.897 (0.000)***	-1.885 (0.016)***	1.267 (0.132)***	-1.882 (0.016)***
N	536,613	657,978	657,978	657,978	657,978	657,978	657,978	657,978
Importer fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R^2/Pseudo R^2	0.212	0.216	0.214	0.209	0.295	0.030		
Log likelihood	-1,361,261	-1,605,480	-1,739,868	-1,849,124	-1.90e+10	-3,837,616	-3,832,120	
Overdispersion						2.031***	1.992***	
AIC	2,722,596	3,211,033	3,479,810	3,698,322	3.79e+10	7,675,309	7,664,378	
BIC	2,723,010	3,211,455	3,480,232	3,698,743	3.79e+10	7,675,742	7,665,165	

Notes: All regressions include a constant. Standard errors are in parentheses. ***, **, * indicate significance at the 1%, 5%, 10% level.

Table C.7 Results using full dataset and without tariff

	OLS (1) $Ln(T_{ij})$	OLS (2a) $Ln(T_{ij} + 1)$	OLS (2b) $Ln(T_{ij} + 0.1)$	OLS (2c) $Ln(T_{ij} + 0.01)$	PM (3) T_{ij}	NBPM (4) T_{ij}	ZINBPM (5) T_{ij}	
							Logit	Neg. Binomial
GDP exporter (billion current USD)	0.649	0.257	0.379	0.509	0.760	1.063	0.059	0.371
	(0.002)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.002)***	(0.012)***	(0.016)***
GDP importer (billion current USD)	0.689	0.207	0.292	0.383	0.452	0.176	-1.086	0.495
	(0.006)***	(0.001)***	(0.001)***	(0.001)***	(0.000)***	(0.013)***	(0.002)***	(0.002)***
Weighted geographical distance (ln)	-0.525	-0.093	-0.136	-0.183	-0.146	-1.024	0.685	-0.447
	(0.007)***	(0.001)***	(0.001)***	(0.002)***	(0.000)***	(0.005)***	(0.007)***	(0.006)***
Common border dummy	0.837	0.937	1.276	1.624	0.922	0.825	-1.296	0.748
	(0.018)***	(0.004)***	(0.006)***	(0.007)***	(0.000)***	(0.023)***	(0.022)***	(0.018)***
Common lang9 dummy	0.130	0.170	0.270	0.382	0.100	0.922	-0.680	0.386
	(0.011)***	(0.002)***	(0.003)***	(0.003)***	(0.000)***	(0.011)***	(0.009)***	(0.010)***
NTM dummy	0.472	0.032	0.030	0.023	0.667	1.153	0.163	0.914
	(0.011)***	(0.002)***	(0.002)***	(0.003)***	(0.000)***	(0.008)***	(0.007)***	(0.009)***
N	680,674	5,876,544	5,876,544	5,876,544	5,876,544	5,876,544	5,876,544	5,876,544
Importer fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R^2/Pseudo R^2	0.193	0.180	0.199	0.210	0.404	0.051		
Log likelihood	-1,753,293	-1.04e+07	-1.22e+07	-1.37e+07	-2.42e+10	-5,830,248	-5,494,948	
Overdispersion						4.046***	2.902***	
AIC	3,506,659	2.08e+07	2.45e+e07	2.75e+07	4.83e+10	1.17e+07	1.10e+07	
BIC	3,507,070	2.08e+07	2.45e+e07	2.75e+07	4.83e+10	1.17e+07	1.10e+07	

Notes: All regressions include a constant. Standard errors are in parentheses. ***, **, * indicate significance at the 1%, 5%, 10% level.

Table C.8 Results using other definition NTM

	OLS (1)	OLS (2a)	OLS (2b)	OLS (2c)	PM (3)	NBPM (4)	ZINBPM (5)	
	$Ln(T_{ij})$	$Ln(T_{ij} + 1)$	$Ln(T_{ij} + 0.1)$	$Ln(T_{ij} + 0.01)$	T_{ij}	T_{ij}	T_{ij}	
							Logit	Neg. Binomial
GDP exporter (billion current	0.601	0.549	0.659	0.748	0.534	0.400	-1.591	-0.076
USD)	(0.003)***	(0.002)***	(0.002)***	(0.003)***	(0.000)***	(0.002)***	(0.619)**	(0.014)***
GDP importer (billion current	0.606	0.418	0.441	0.440	0.182	-0.093	-0.269	0.231
USD)	(0.007)***	(0.006)***	(0.007)***	(0.008)***	(0.000)***	(0.011)***	(0.198)	(0.002)***
Weighted geographical distance	-0.427	-0.276	-0.305	-0.323	-0.323	-0.221	-0.962	-0.089
(ln)	(0.008)***	(0.006)***	(0.007)***	(0.008)***	(0.000)***	(0.006)***	(0.496)*	(0.006)***
Common border dummy	0.657	0.964	1.175	1.364	0.203	0.851	-1.831	0.955
	(0.019)***	(0.016)***	(0.019)***	(0.022)***	(0.000)***	(0.016)***	(1.372)	(0.015)***
Common lang9 dummy	0.075	0.134	0.167	0.192	-0.116	0.224	-0.923	0.200
	(0.012)***	(0.009)***	(0.011)***	(0.013)***	(0.000)***	(0.010)***	(0.800)	(0.009)***
NTM new	1.156	-1.948	-3.119	-4.418	-0.185	-0.080	42.399	1.848
	(0.023)***	(0.013)***	(0.015)***	(0.018)***	(0.000)***	(0.020)***	(1.665)***	(0.017)***
Tariff (weighted average) (ln)	-4.391	-2.120	-2.176	-1.963	-12.119	-1.743	-1.119	-1.899
	(0.058)***	(0.045)***	(0.055)***	(0.064)***	(0.000)***	(0.016)***	(1.877)	(0.016)***
N	536,613	657,978	657,978	657,978	657,978	657,978	657,978	657,978
Importer fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R^2/Pseudo R^2	0.212	0.240	0.258	0.275	0.287	0.028		
Log likelihood	-1,361,301	-1,595,184	-1,720,897	-1,820,666	-1.92e+10	-3,842,438	-3,730,278	
Overdispersion						2.042***	1.816***	
AIC	2,722,675	3,190,442	3,441,868	3,641,407	3.83e+10	7,684,952	7,460,695	
BIC	2,723,089	3,190,863	3,442,290	3,641,829	3.83e+10	7,685,385	7,461,493	

Notes: All regressions include a constant. Standard errors are in parentheses. ***, **, * indicate significance at the 1%, 5%, 10% level.