Charles University in Prague

Faculty of Social Sciences
Institute of Economic Studies

MASTER THESIS

Gravity model estimation using panel data - is logarithmic transformation advisable?

Author: **Bc. Božena Bobková**

Supervisor: **Doc. Ing. Vladimír Benáček, Csc.**

Academic Year: **2011/2012**

Prohlášení

Tímto prohlašuji, že uvedenou práci jsem zpracovala samostatně a použila jsem jen uvedené prameny a literaturu.
Dále prohlašuji, že tato práce nebyla použita k získání jiného titulu.
Také souhlasím s tím, aby práce byla zpřístupněna pro studijní a výzkumné účely.

V Praze, dne 16.1. 2012,

Božena Bobková

Poděkování

Tímto bych ráda poděkovala panu docentu Vladimíru Benáčkovi za vedení této práce a za cenné rady, připomínky a podporu, které mi při psaní věnoval. Dále bych ráda poděkovala panu doktoru Vilému Semerákovi za cenné rady a náměty. V neposlední řadě bych ráda poděkovala své matce, která mi při psaní práce byla morální oporou.

Abstract

This thesis investigates the question if the estimation of gravity model of international trade based on the logarithmic transformation of the model is advisable when panel data are employed for the estimation. We have derived theoretically that in the presence of heteroskedasticity the logarithmic transformation causes inconsistency of the estimated coefficients. According to the literature, we have recommended rather the Poisson pseudo maximum likelihood estimation technique for the empirical research of the gravity model. We have also provided an empirical analysis of Czech and German panel data sets based on the comparison of the performance of traditional and Poisson estimation approaches. This analysis confirms Poisson pseudo maximum likelihood estimation method as a more proper method for estimating the coefficients of the gravity equation.

JEL Classification	C13, C23, F10, F11, F12, F14
Keywords	Gravity model, Heteroscedasticity, Jensen's inequality, Panel data, Poisson regression

Author's e-mail	Bozena.Bobkova@seznam.cz
Supervisor's e-mail	vladimir.benacek@fsv.cuni.cz

Abstrakt

Tato diplomová práce zkoumá otázku, zda odhad gravitačního modelu mezinárodního obchodu založeného na logaritmické transformaci modelu je vhodný v případě, že odhad je prováděn na panelových datech. Odvodili jsme teoreticky, že za přítomnosti heteroskedasticity použití logaritmické transformace způsobuje nekonsistentnost odhadnutých koeficientů.V návaznosti na literatu jsme doporučili raději Poissonovský druh odhadu pro empirické zkoumání gravitačního modelu. Také jsme provedli empirickou analýzu na českých a německých panelových datech, která byla založena na srovnání tradiční a Poissonovské metody odhadu. Tato analýza potvrdila, že Poissonovský typ odhadu je více vhodný pro odhad koeficientů gravitačního modelu.

Klasifikace JEL C13, C23, F10, F11, F12, F14
Klíčová slova Gravitační model, Heteroskedasticita, Jensenova nerovnováha, Panelová data, Poissonovská regrese

E-mail autora Bozena.Bobkova@seznam.cz
E-mail vedoucího práce vladimir.benacek@fsv.cuni.cz

Contents

List of Tables

List of Figures

Master Thesis Proposal

Author	Bc. Božena Bobková
Supervisor	Doc. Ing. Vladimír Benáček, Csc.
Proposed topic	Gravity model estimation using panel data - is logarithmic transformation advisable?

Topic characteristics The majority of empirical literature uses the logarithmic transformation of gravity equation for its estimation. Silva and Tenreyro (2006) point out that in the presence of heteroscedasticity is the OLS estimation based on this transformation inconsistent. Moreover, the transformation cannot deal with zero trade observation. They suggest estimating the equation in multiplicative form using Poisson pseudo-maximum likelihood estimation (PPMLE). Westrelund and Wilhelmsson (2007) discuss further the problem of logarithmic transformation when panel data are used. They also conclude that the Poisson estimation is advisable; they concretely suggest estimating the gravity equation by Poisson fixed effects estimation. The thesis will discuss the problem of logarithmic transformation of gravity equation in detail. Using real data; it will compare the results of standard and Poisson panel data estimations (random effects and different types of fixed effects). It will focus on the gravity equation theory, consistency of estimation or fitted values.

Hypotheses 1.The Poisson panel data estimations provide better results from the econometric point of view. 2.The Poisson panel data estimations provide better results from the gravity equation theoretical point of view. 3.The Poisson panel data estimations fit better the data.

Methodology We will perform various panel data estimation technique to estimate the gravity equation using German trade data. We focus on clustered pooled OLS; one way and two way random effects and fixed effects (classical and LSDV). All of these techniques will be performed as standard and Poisson estimation. To compare the models, we will run e.g. some specification tests or estimate the fitted values.

Outline

1. Introduction
2. Literature overview
3. Gravity Model – Theoretical derivation
4. Problems of Logarithmic Transformation and Poisson Estimation
5. Empirical Analysis
6. Conclusion

Core bibliography
Santos Silva, J.M.C. and Tenreyro, Silvana (2006), The Log of Gravity, The Review of Economics and Statistics, 88(4), pp. 641-658

Santos Silva, J.M.C. and Tenreyro, Silvana (2011), Further simulation evidence on the performance of the Poisson pseudo-maximum likelihood estimator, Economics Letters, Elsevier, vol. 112(2), pages 220-222, August

Westerlund J. and Wilhelmsson F. (2007), Estimating the gravity model without gravity using panel data, .Mimeo School of Economics and Management, Lund University.

Author

Supervisor

Chapter 1

Introduction

A considerable amount of empirical and theoretical literature has been published on gravity models during the past 50 years. The conception of gravity models was originally introduced by (Tinbergen 1963). The traditional gravity equation of international trade is a model, which explains the trade flow by home and partner's GDP and trade impediment in the form of distance between the countries. The model became very popular because of its quite simple usage combined with a substantial power of explaining the flows in general. The gravity equation has been exploited as a instrument to model not only international trade flows but also tourism or migration.

There is a large volume of published studies researching the most proper econometric specification of the model. The majority of this studies has been analyzed estimation methods based on the application of cross-sectional data. However, this approach suffers from the producing of biased estimation due to the presence of heterogeneity among countries, which cannot be regulated sufficiently when cross-sectional data are using.

In recent years, there has been an increasing amount of literature (Matyas 1997; Egger 2000; Egger 2002) dealing with the problem by using the panel data instead of cross-sectional data. Applying panel data estimation methods, it is naturally controlled for the heterogeneity among countries.

Further, the majority of studies using either cross-sectional or panel data traditionally estimate the multiplicative gravity equation after the model is log-linearly transformed. This approach allows to employ classical estimation methods. In the cross-sectional framework, the authors usually apply the traditional OLS technique. When the panel data are analyzed, the authors usually introduce the fixed effects or random effects estimation methods.

However, the estimation results based on the logarithmic transformed model could be significantly miss-leading in the presence of heteroscedasticity (Silva & Tenreyro 2006). This conclusion stems from the well-known Jensen's inequality, which states that the expected value of a logarithm of random variable does not equal to the logarithm of expected value. According to the fact, that when estimating the gravity equation by a traditional method, there are strong assumptions on the conditional expected value of the logarithm of error term to ensure the consistency of the estimated coefficients, the Jensen's inequality plays an important role. Namely, Silva & Tenreyro 2006 show that in the presence of heteroscedasticity the assumptions are in general violated. Last but not least, the logarithmic transformation of the model is also struggling how to deal with the zero trade flows.

Thus, the estimation based on the logarithmic transformed model creates a potential significant risk to the properly estimated coefficients. The solution to this problem is to estimate the gravity model directly from the multiplicative form using Poisson pseudo maximum likelihood estimation technique (Silva & Tenreyro 2006). This approach was employed firstly on cross-sectional data and later on panel data as well. For instance,Westerlund & Wilhelmsson 2009 investigate the influence of applying these two different approaches of gravity equation estimation on either simulated or real data.

This thesis is a contribution to this discussion and to relatively scare literature on the problems with estimation of logarithmic transformed gravity equation using panel data. Firstly, we will show theoretically in detail that even when panel data are used the presence of heteroscedasticity makes the traditional estimation biased and inconsistent. Moreover, we will apply our findings on the real panel data sets of the Czech Republic and Germany and we will estimate the gravity equation by traditional and Poisson estimation technique. We compare the performance of both methods in respect to the theory of the gravity equation; the correct specification; and how the predicted flow fits the data.

Our approach is innovative in the way that we will use one-way trade flow for one home country as a dependant variable, which enables us to study the trade in more specific way. Further, we will introduce the influence of the recession into the gravity equation. Next, we will compare the performance of the techniques using four different panel data methods. Last but not least, the performance of these specification have never been compared on the Czech or Germans panel data set. Finally, we will compare how the predicted trade flows fit the original data rather than the transformed data.

The thesis is structured as follows: Chapter 2 summarizes the literature of gravity equation and the studies dealing with the problems of estimation of logarithmic transformed gravity equation. Chapter 3 provides the theoretical derivation of the gravity equation. Chapter 4 discusses in detail all the weaknesses connected to the estimation of the log-linearized equation. Chapter 5 introduces our empirical research, discusses econometric specification, variables and methodology for this research. Chapter 6 provides results of both estimation techniques and compares their performance. Chapter 7 summarizes our findings and discusses which estimation method is more proper.

Chapter 2

Literature review

2.1 A note of the theoretical literature of gravity equation

Since the first introduction of the gravity models,they have been performing very well in empirical applications. However, there was a problem with a lack of theoretical foundations of this concept for a long time.

In 1979, Anderson published a paper in which he provides first serious micro-foundations of the gravity equation based on Armington preferences. However, the Anderson's theoretical concept of gravity models was based on some strong and simplifying assumptions, namely Anderson assumes that each country is fully specialized in production of one good. In the later study (Anderson & van Wincoop 2001), author avoids these weaknesses and enhances the micro-founded theory. This theoretical concept is also adopted by this study and further described in chapter 3. In the literature of international trade, there can be found many other theories based on different basis. For instance, the factor-endowment approach introduced by Deardorff (1995); or increasing returns to scale approach investigated by Helpman & Krugman (1990).

2.2 Discussion of the suitability of the logarithmic transformation in trade literature

Most of the empirical studies of gravity equation are traditionally based on estimation of the log-linearized version of the gravity equation. However, there is a few of studies [Silva & Tenreyro (2006) or Westerlund & Wilhelmsson 2009] demonstrating that this approach suffers from some serious weaknesses.

Firstly, it was investigated that strong assumptions on the error term of the multiplicative model has to be imposed to estimate the model consistently after the logarithmic transformation due to the Jensen's inequality. Secondly, the logarithmic transformation is not able to deal with the zero trade observations, which are in bilateral trade data very common. Last but not least, Arvis & Shepherd (2011) pointed out that the log-linearized model also suffers from the so called adding up problem.

This section briefly summarizes the studies which pursue the problem of logarithmic transformation international trade. It discusses it in terms of the type of data they are using (cross-sectional or panel data). We summarize the studies demonstrating the problems named above, analyzing their impacts and recommending the proper estimation technique. However, a detailed discussion of this problems and their solution are provided in Chapter 4 and Chapter 5.

2.2.1 Crossectional

In 2006, Silva & Tenreyro pointed out as the first the problematic of logarithmic transformation of gravity equation. For the traditional gravity equation in the multiplicative form, they derive that to provide consistent estimation of the logarithmic transformed model there is a need for the error term to be statistically independent of other regressors and its conditional mean value on other regressors has to equal 1.

They suggest that especially the assumption of error term statistical independence on other variables is crucial; after logarithmic transformation it should also hold for the logarithm of error term; but they remind that expected value of logarithm of random variable is a function of the random variable mean and also higher moments of its distribution.

Further, Silva & Tenreyro demonstrate that in the presence of heteroscedasticity (meaning that the variance is a function of other regressors) the crucial assumption of the error term statistical independence of other regressors is violated and the OLS estimation is inconsistent. The authors also claim that analyzing the logarithmic transformed data of international trade, they found the evidence of heteroscedasticity being enormous.

Silva & Tenreyro (2006) also discussed the possible presence of the multilateral resistance term in the model suggested by Anderson & van Wincoop

(2001) and its possible improvement for the estimation. They conclude that in the presence of heteroscedasticity, individual fixed effects may make the problem less significant. Testing this hypothesis empirically, they assert that heteroscedasticity means a serious problem for the estimation of the transformed gravity equation even when controlling for fixed effects. Last but not least, they also point out the problem with zero observation for the bilateral trade because taking logarithm of this variable fails in that case.

Silva & Tenreyro (2006) conclude that the log-linear transformation of the gravity model is not advisable and recommend estimating the equation directly from the multiplicative form. They suggest foremost two type of estimation: non-linear least squares NLS and pseudo-maximum likelihood estimators based on some assumption of functional form of the conditional variance of bilateral trade variable.

Firstly, The non-linear least squares estimator put emphasis more on noisier observations with higher variance. They conclude that this estimation is then supposed to be very inefficient. On the other hand, assuming conditional variance being proportional to conditional mean, authors identify Poisson pseudo-maximum likelihood estimator PPML(usually used for count data analysis) being the appropriate estimator from the theoretical point of view. They claim that even if the proposed assumption is not fulfilled the estimator is more efficient than NLS; moreover, if the conditional mean is correctly specified in the form of exponential function $E(y_i|x) = exp(x_i\beta)$the estimator is consistent even if the dependent variable is uncount or does not evince being Poisson distributed.

Last but not least, the authors consider the gamma PML estimator as the possible appropriate estimator for the gravity equation assuming that in this case is the conditional variance function of higher powers of conditional mean. However, they identify this assumption causes a problem, namely the estimator might give excessive weight to the observation susceptible to measurement errors.

Finally, Silva & Tenreyro (2006) summarize their theoretical foundation finding Poisson PML estimation as the most likely proper estimation of gravity equation. They also provide a simulation study and an estimation of gravity equation on real data to compare the performance of different type of estimators (e.g. OLS, NLS, PPMPL or Tobit) to verify their conclusions. For the simulation, they generate the data for four cases under different patterns of heteroscedasticity. Their results more or less confirm theoretical hypothesis.

For instance, according to the simulations, OLS provide reasonable estimation only when the generated data fulfill the logarithmic transformation error term assumptions discussed earlier; otherwise is the OLS estimation badly biased. Moreover, an upward bias of coefficients for income elasticities or graphical proximity was detected for OLS estimation by comparing them with PPMLE, performing both estimations on real data set. On the other hand, the simulations detect that the PPML estimator performs very well in all cases of suggested heteroscedasticity. Last but not least, the authors pointed out a poor performance of NLS estimator, especially in cases with severe heteroscedasticity in data. The simulations also confirm the significant sensitivity of gamma PML estimator to the measurement errors.

However, the simulations performed by Silva & Tenreyro (2006) suffers from their assumption on generated data, namely the bilateral trade variable was generated strictly positive. There is a doubt if the PPML estimator would perform so well even in the frequent presence of zeros in bilateral trade variable. Motivated by this hesitation, Silva & Tenreyro (2009) provide again the similar Monte-Carlo simulation taking the large zeros frequency into account. The PPMLE suitability for gravity equation estimation was confirmed by this study.

In contrary, the studies of Martinez-Zarzoso (2013) or Martin & Pham (2008) disprove partially the results of Silva & Tenreyro (2006). In both studies, Monte-Carlo simulation is provided as well to test the performance of different types of estimation of gravity equation. On the other hand, they also propose using other types of estimators. For instance, Martinez-Zarzoso suggest using FGLS to deal with heteroscedasticity or Heckman selection estimator to deal with zeros in the data. Moreover, the Heckam selection estimator takes into consideration the probability if the countries would trade or not.

Martin & Pham also investigate the application of a range of estimators such as truncated OLS, different types of Tobit models or also Heckman selection estimator. According to the results of Martinez-Zarzoso, it is argued that in terms of out-of-sample forecast FGLS, OLS or sample selection techniques estimation offer better results than PPMLE. Similarly, the study of Martin & Pham does not confirm PPMLE as an advisable estimator and the author are rather inclined to more traditional estimation technique such as truncated OLS based on the logarithmic transformation. However, Silva & Tenreyro (2009) point out that the findings of both studies are generally useless because the authors generated the data on the base of non-constant income elasticity model.

The problematic of logarithmic transformation of gravity equation also discusses a study by Siliverstovs & Schumacher . The authors provide an empirical analysis to investigate the differences between the OLS and PPML estimation technique results and also compare their findings with Silva & Tenreyro. They estimate the gravity equation using trade data of OECD countries in years 1988 - 1990. Their estimation is split to compare the results for aggregated trade flows and for manufacturing goods as well as for manufacturing trade disaggregated at the three-digit level. Their findings more or less confirm the findings of Silva & Tenreyro.

Last but not least, the study of Burger *et al.* confirms the suggestion of Silva & Tenreyro to estimate rather the multiplicative form of gravity equation. Besides estimate the equation by PPMLE, the authors also advocate applying negative binomial and zero-inflated models as modified Poisson models. The motivation for Burger *et al.* to employ the modification in form of negative binomial model lies in the fact that the assumption for the application of Poisson model is usually violated. Namely, the conditional variance is usually higher than conditional mean. This problem is called over-dispersion of the dependent variable. Burger *et al.* identify that the higher conditional variance is usually caused by the presence of omitted variables and so unobserved heterogeneity. Poisson estimation deals only with observed heterogeneity. The resulting estimation is then still consistent but not efficient. For the estimation, the negative binomial model is in this case preferred over standard Poisson because it can naturally also account for the hidden heterogeneity.

Further, study the problem of excess zeros in the data (when the number of zeros is higher than Poisson or negative binomial model predicts). Referencing the statistics literature, Burger *et al.* argue that this excess zeros demonstrates itself as over-dispersion and its causation is linked to a presence of "non-Poissonness". The authors investigate that the presence of "non-Poissonness" is caused by different types of zeros in the international trade data. Namely, one part of zero trade observations is created by a different process than other zero or non-zero observations. For instance, the absence of trade between two countries caused by a lack of natural resources is different than the absence caused by distance or different specialization. The main difference is that in the first case the probability of trade is essentially zero and in the other case the probability is theoretically different than zero. Burger *et al.* pointed out that this problem solves the application of zero-inflated estimation technique: zero inflated Poisson pseudo maximum likelihood estimation

(ZIPPMLE) and zero inflated negative binomial pseudo-maximum likelihood estimation (ZINPBMLE).

Burger *et al.* also provide an empirical analysis to compare different types of model specifications using data for 138 countries in years 1996 - 2000, estimating export on time average in these years. Besides the estimation technique proposed above, they employ traditional OLS estimation based on logarithmic transformation accounting for zero observation in the way that they treat the zero observation as small positive values. Their results confirm serious bias of OLS estimation caused by transforming zeros. Applying goodness of fit and the relevance of excess zeros as the indicators of a good specification, the authors assess ZIPPMLE as the on average best score estimator.

2.2.2 Panel

The empirical analysis of gravity equation has traditionally been based on cross-sectional data. However, this approach cannot sufficiently account for heterogeneity among countries. On the other hand, using panel data for the estimation allows taking into consideration more general types of heterogeneity. Increasing amount of current literature on gravity equation notices this fact and estimates the equation employing panel data mode. However, the researchers apply logarithmic transformation before the panel data estimation technique performed.

Westerlund & Wilhelmsson point out that the logarithmic transformation of the model for its estimation still causes problems even if panel data estimation methods are used. Namely, the authors identify the problem with zero trade observations and also with the heteroscedasticity present in the model. Firstly, they identify that replacing zeros by some small positive values causes sample selection bias. Secondly, they argue that correct estimation of the log-linearized gravity equation by fixed effects technique requires that the conditional expected value of logarithm of the general error term from the model equals to zero. However, the study demonstrates that this assumption is violated due to Jensen's inequality. In addition, they argue that this problems cause the tradition fixed effects OLS severely biased and inefficient. Further discussion of this problem can be found in Chapter 4.

Westerlund & Wilhelmsson also suggest estimating the gravity equation from its multiplicative form by fixed effects Poisson pseudo maximum likelihood estimator. They choose fixed effects estimator rather than random effects estimator, they point out that in the multiplicative form is generally the assumption of non-correlation of individual specific effect on other regressors generally violated. The authors confirm their suggestion when comparing estimation results of traditional and Poisson fixed effects estimation techniques which were applied on panel data generated by a Monte Carlo simulation (both homoscedastic and heteroscedastic) and on real panel data of bilateral trade of Austria, Finland and Sweden in years 1992 -2002.

The authors show that the performance of the traditional OLS fixed effects approach was so poor on simulated panel data that it was not meaningful to interpret the results. On the other hand, they highlight very good results of Poisson estimation with very small bias and good size accuracy. They detect only one disadvantage of this estimator - downwardly biased estimated of standard errors. W and W recommend using bootstrapped standard errors to fix this problem.

Analyzing real panel data, Westerlund & Wilhelmsson obtain significantly different results for OLS and Poisson fixed effects. In addition, they conclude that the Poisson fixed effects estimator with bootstrapped standard errors is the most advisable.

Chapter 3

Gravity models of trade

The classical conception of gravity model originally reported by (Tinbergen 1963) was inspired by the Newton's law of universal gravitation. This law states[1] that every point mass attracts every other point mass with a gravity force F_g that is directly proportional to the product of their masses M_1 and M_2 and inversely proportional to the square of the distance r between them:

$$F_g = G\frac{M_1 M_2}{r^2}.$$

Gravity model for international trade considers the bilateral trade as the "gravity force" between two countries and suggests the same relationship between this force, masses of the countries proxied by GDP and the distance between them.

3.1 Theoretical Model

The theoretical model of this paper was adopted from the study by Baldwin & Taglioni, in which the author follows the theory concept by Anderson and adjusts it for the possible application of panel data. Based on this theory, the gravity model is basically the expenditure equation with expenditure share identity as the cornerstone of the model. Baldwin & Taglioni divided the derivation of the model into six steps which will be described in detail subsequently. For all the steps, the prices and expenditures are measured in numeraire.

[1]source: www.wikipedia.com

3.1.1 The expenditure share identity

Let us introduce the expenditure share identity for a single good exported from country j to country i. This identity equivalizes the value of the trade flow of country j with the share S_{ji} of expenditure E_i in country i on a typical variety from country j:

$$p_{ji}x_{ji} \equiv S_{ji}E_i. \tag{3.1}$$

The trade flow is defined as the product of the price p_{ji} of imported good in country i and the export x_{ji} of a single variety from country j to country i.

3.1.2 The expenditure function

According to microeconomic theory, the expenditure share is a relation of relative prices and income levels. Simplifying this theory, we assume that the expenditure share depends only on relative prices. Moreover, we assume the CES demand production function and that all goods are traded. The expenditure share is then expressed as:

$$S_{ji} \equiv \left(\frac{p_{ji}}{P_i}\right)^{1-\sigma}. \tag{3.2}$$

The expression $\frac{p_{ji}}{P_i}$ stands for the real price of p_{ji}, where $P_i \equiv \left(\sum_{k=1}^{R} n_k p_{ki}^{1-\sigma}\right)^{\frac{1}{1-\sigma}}$ is the ideal domestic CES price index. σ is the elasticity of substitution among all varieties and is assumed to be higher than one; R stands for the number of importing countries into the domestic country i counting also itself; n_k is the number of varieties exported form country k. The varieties are assumed to be symmetric.

Plugging 3.2 into 3.1, we get the product specific import expenditure equation:

$$p_{ji}x_{ji} \equiv \left(\frac{p_{ji}}{P_i}\right)^{1-\sigma} E_i \tag{3.3}$$

.

3.1.3 Aggregating across individual good

To aggregate the pro-variety exports we multiply the expenditure share equation by n_j, where n_j is the number of symmetric varieties country j supplies:

$$V_{ji} \equiv n_j p_{ji} x_{ji} \equiv n_j \left(\frac{p_{ji}}{P_i} \right)^{1-\sigma} E_i, \tag{3.4}$$

where V_{ji} is the total value of trade from country j to country i.

3.1.4 Adding the pass-trough equation

The price p_{ji} of imported good in country i is obtained as a mark-up price of the production and transportation cost:

$$p_{ji} = \mu p_j \tau_{ji}, \tag{3.5}$$

where p_j stands for the producer price in country j, τ_{ji} covers the trade costs and μ is the bilateral mark-up. Moreover, we assume the perfect competition with Armington good or the Dixit-Stigliz monopolistic competition, so the mark-up μ is assumed to equal one.
Combining 3.4 and 3.5 we get for the total value of trade from country j to country i:

$$V_{ji} = n_j \left(p_j \tau_{ji} \right)^{1-\sigma} \frac{E_i}{P_i^{1-\sigma}}. \tag{3.6}$$

This equation also expresses sales of country j to each market. Summing V_{ji} over all markets yields total sales of country j goods.

3.1.5 Market clearing

Producer price in country j reflects that all the output can be sold either home or abroad. Moreover, the prices and wages in country j are adjusted to production of traded goods and sales of trade goods fulfill the market-clearing condition:

$$Y_j = \sum_{i-1}^{R} V_{ji}, \tag{3.7}$$

where Y_j stands for the output of country j measured in numeraire and we sum the sale of trade goods over all markets, including country j market. Plugging 3.6 for V_{ji} we get:

$$Y_j = n_j p_j^{1-\sigma} \sum_{i-1}^{R} E_i \left(\frac{\tau_{ji}}{P_i} \right)^{1-\sigma}. \tag{3.8}$$

If we solve this equation for $n_j p_j^{1-\sigma}$, we get:

$$n_j p_j^{1-\sigma} = \frac{Y_j}{\sum_{i-1}^{R} E_i \left(\frac{\tau_{ji}}{P_i}\right)^{1-\sigma}}. \tag{3.9}$$

3.1.6 A first-pass gravity equation

The denominator of the right hand side of the equation 3.9 could be denoted as Ω_j, where Ω_j is similar to the market potential and measures the openness of country j's exports to international markets. Plugging for $n_j p_j^{1-\sigma}$ from 3.9 into 3.6 yields the first-pass microfounded gravity equation:

$$V_{ji} = \tau_{ji}^{1-\sigma} \left(\frac{Y_j E_i}{\Omega_j P_i^{1-\sigma}}\right). \tag{3.10}$$

3.1.7 The gravity equation

Let us proxy Y_j by country j's GDP and E_i by country i's GDP. Further, assume that the trade costs τ_{ji} are related only to the distance between two countries. Let us also define an "un-constant" G, where

$$G \equiv \frac{1}{\Omega_j P_i^{1-elasticity}}, \tag{3.11}$$

then the gravity equation gets the form:

$$bilateral\ trade = G \frac{GDP_j GDP_i}{distance_{ji}^{elasticity-1}}. \tag{3.12}$$

The "un-constant" G is also called the multilateral resistance term. According to Anderson, including this term into the empirical model is crucial for proper specification of the model.

3.2 Empirical Application of the Gravity equation

The most important merit of the gravity model lies foremost in its empirical application. In next two sections we show how to transform both the traditional and Anderson theoretical gravity model into the stochastic version that could be estimated by an econometric method.

3.2.1 Estimating Traditional Gravity Equation

For estimating purposes, the traditional gravity model of international trade similar to the equation 3.12 could be rewritten in the form:

$$X_{ji} = \beta_0 \; GDP_j^{\beta_1} GDP_i^{\beta_2} D_{ij}^{\beta_3} \epsilon_{ij}, \tag{3.13}$$

where X_{ji} stands for the bilateral trade between countries i and j; D_{ij} is a distance between these two countries; ϵ_{ij} stands for the error term and β_0, β_1, β_2 and β_3 are parameters to be estimated.
We assume that he error term ϵ_{ij} is statistically independent on the othe regressors; moreover, we further assume that $E\left(\epsilon_{ji} | GDP_i, GDP_j, D_{ij}\right) = 1$. This assumption leads to:

$$E\left(X_{ji} | GDP_i, GDP_j, D_{ij}\right) = \beta_0 \; GDP_j^{\beta_1} GDP_i^{\beta_2} D_{ij}^{\beta_3}. \tag{3.14}$$

However, the gravity model is identified in multiplicative form, which does not permit for employing standard estimation techniques. The traditional way in the literature how to deal with estimation of multiplicative form of the model is to estimate the logarithmic transformed model:

$$ln\left(X_{ji}\right) = ln\left(\beta_0\right) + \beta_1 ln\left(GDP_j\right) + \beta_2 ln\left(GDP_i\right) + \beta_3 ln\left(D_{ij}\right) + ln\left(\epsilon_{ij}\right). \tag{3.15}$$

3.2.2 Estimating Anderson's gravity equation and estimating gravity equation using panel data

Adopting the Anderson's theory concept derived earlier, the aim of the empirical studies is to estimate stochastic version of the equation 3.10. Let us rewrite the deterministic version of this equation:

$$V_{ji} = \beta_0 \; \tau_{ji}^{\beta_1} GDP_j^{\beta_2} GDP_i^{\beta_3} e^{\theta_j} e^{\theta_i}, \tag{3.16}$$

where the expressions θ_j and θ_i stand for the multilateral resistance terms in the form of importer and exporter fixed effects; and again Y_j is proxied by country j's GDP and E_i is proxied by country i's GDP. In empirical literature, τ_{ji} is treated as a function of distance between countries i and j and other stuff creating costs of trade for the countries. The original theoretical model by Anderson assumes the unit-income elasticity model with $\beta_3 = \beta_4 = 1$.

Stochastic version of the model has the form:

$$E\left(V_{ji}|GDP_i, GDP_j, \tau_{ji}, \theta_i, \theta_j\right) = \beta_0\ \tau_{ji}^{\beta_1} GDP_j^{\beta_2} GDP_i^{\beta_3} e^{\theta_j} e^{\theta_i}. \qquad (3.17)$$

Further, let us rewrite the equation to define the regression:

$$V_{ji} = \beta_0\ \tau_{ji}^{\beta_1} GDP_j^{\beta_2} GDP_i^{\beta_3} e^{\theta_j} e^{\theta_i} \epsilon_{ij}, \qquad (3.18)$$

where ϵ_{ij} stands for the error term independent on other regressors with $E\left(\epsilon_{ji}|GDP_i, GDP_j, \tau_{ji}, \theta_i, \theta_j\right) = 1$.

To estimate equation 3.18, the traditional approach consists also in estimation of the logarithmic transformed model:

$$ln(V_{ji}) = ln(\beta_0) + \beta_1 ln(\tau_{ji}) + \beta_2 ln(GDP_j) + \beta_3 ln(GDP_i) + \theta_j + \theta_i + ln(\epsilon_{ij}). \qquad (3.19)$$

However, the equation also contains the importer and exporter unobserved fixed effects. If these effects are correlated with other regressors, then coefficients estimated by traditional OLS are by definition inconsistent. To deal with this problem, the fixed effect is usually proxied by importer and exporter dummy variables. The other method lies in using panel data rather than cross-sectional data. The panel data estimation methods naturally control for the unobserved fixed effect's.

The estimation of gravity equation using panel data is generally based on the estimation of stochastic version of the Anderson's model summarized in equation 3.18. The only difference is that we assume the estimation in a time frame, which leads to the equation:

$$V_{jit} = \beta_0\ \tau_{jit}^{\beta_1} GDP_{jt}^{\beta_2} GDP_{it}^{\beta_3} e^{\theta_j} e^{\theta_i} \epsilon_{ijt}, \qquad (3.20)$$

where t is a time index. The panel data estimation method are also usually applied on the logarithmic transformed model:

$$ln(V_{jit}) = ln(\beta_0) + \beta_1 ln(\tau_{jit}) + \beta_2 ln(GDP_{jt}) + \beta_3 ln(GDP_{it}) + \theta_j + \theta_i + ln(\epsilon_{ijt}). \qquad (3.21)$$

The problems connected with the logarithmic transformation are discussed in the next chapter.

Chapter 4

Problems with the Logarithmic Transformed Model Estimation

This chapter discusses in detail the problems resulting from the logarithmic transformation of the gravity model to be estimated which is widely employed in empirical trade literature. We provide this analysis for both cross-sectional and panel data. Finally, according to the literature summarized in Chapter 2.We propose a solution to these problems in the form of direct estimation of the multiplicative form by Poisson pseudo-maximum likelihood estimator.

4.1 "The Logarithmic Transformation Effects the Nature of the Estimation"

The primary interest of the empirical application of gravity model is to estimate the trade flow between two countries. However, when the logarithmic transformation is applied on the equation for estimation purposes, we estimate the logarithm of the trade flow instead of the trade flow itself. Moreover, taking the antilogarithm of this estimates, the biased estimation of the variable of interested is obtained [Haworth & Vincent (1979)].

Haworth & Vincent explain the bias by the fact, that the application of logarithmic transformation for estimation basically impose an inappropriate assumption on the dependant variable as being log-normally distributed. In fact, the log-normal distribution evinces the positive skewness. These findings are also linked with the well-known Jensen's inequality which implies for the expected value of a random variable Z that:

$$lnE(Z) \leq E(lnZ). \tag{4.1}$$

.

Arvis & Shepherd (2011) demonstrate that this bias leads to over-prediction of especially large flows and thus total flows.

4.2 Zero-valued Observation of Dependant Variable

It is natural, that many countries do not trade with each other. The data of trade flows between countries usually contain a hardly negligible amount of zero-valued observations.The logarithmic transformation is then in this case improper because logarithm of zero is not defined.

To solve the problem of zero-valued trade flows, various methods have been developed in empirical literature. The most common approaches are to add some small positive value to all observations or get rid of the zero-valued observation by deleting them.

However, Flowerdew & Aitkin (1982) demonstrate, that in the case of adding some small value, the resulting estimation highly varies with the choose of such a small number. On the other hand, omitting the observations causes serious problems as well. Firstly, we loose the information encompassed in the deleted data [Eichengreen & Irwin (1996)]. Moreover, the estimation suffers very likely from a sample selection bias caused by omitted zero-valued trade flows observations which are probably non-randomly distributed [Burger *et al.* (2009)].

4.3 Problem of Inconsistent Estimation in the Presence of Heteroscedasticity

Firstly, we demonstrate the problem in a general case of constant-elasticity model. Secondly, we apply the findings from the general case on the gravity equation. In this section, we follow the derivation of this problem proposed by Silva & Tenreyro (2006). We also supplement it by some additional explanation and extend these findings to application of panel data.

4.3.1 Estimation of the Constant-Elasticity Model

The multiplicative constant elasticity model is very common and could be found in many areas of the literature. Silva & Tenreyro investigate the problem wheatear transform or not in the case of general constant-elasticity model.

Let us start our analysis from the deterministic form of a multiplicative constant elasticity model. If the model is based on economic variables, the proposed relationship never holds so accurate in reality and in all cases [1]like in e.g. physic. What we can only expect is that the by model proposed relationship holds on average. This a key assumption for the analysis.

Let us define variables $y \geqslant 0$ as the variable of interest and x as the explanatory variable. If we claim, that y is linked to x by a constant elasticity model in a specific form of:

$$y_i = exp\left(x_i\beta\right), \tag{4.2}$$

we mean that this relation holds on average, so it holds for the conditional expectation of the variable of interest:

$$E\left(y_i|x\right) = exp\left(x_i\beta\right). \tag{4.3}$$

In contrary, for individual realizations of y_i is the equation 4.2 no longer valid. We need to add an error term ϵ_i into the equation, which would measure the deviation of an individual realization of y_i from its conditional mean as $\epsilon_i = y_i - E\left(y_i|x\right)$. This leads to the stochastic version of the model:

$$y_i = exp\left(x_i\beta\right) + \epsilon_i, \tag{4.4}$$

with $E\left(\epsilon_i|x\right) = 0$.

Further, let us define a random variable ω_i, where $\omega_i = 1 + \frac{\epsilon_i}{exp(x_i\beta)}$ and $E\left(\omega_i|x\right) = 1$. Using ω_i, we can rewrite the equation 4.4 in the form:

$$y_i = exp\left(x_i\beta\right)\omega_i. \tag{4.5}$$

We have obtained a stochastic multiplicative model. To gain an estimation of the slope coefficient β from the equation 4.4 , the traditional approach is

[1]For instance, human factor plays its role

to estimate the parameter β from the logarithmic transformed version of this model[2]:

$$ln(y_i) = x_i\beta + ln(\omega_i). \tag{4.6}$$

When we decide for the OLS estimation technique, the crucial assumption to yield the consistent estimation of β is that $ln(\omega_i)$ is not statistically dependant on other regressors. Namely, we require for $E\left(ln(\omega_i)|x\right)$ being zero or some constant.

However, we defined ω_i as a function of ϵ_i and the explanatory variable x_i in the form of $\omega_i = 1 + \frac{\epsilon_i}{exp(x_i\beta)}$. Thus, the only way, how to meet the assumption on $ln(\omega_i)$ is to assume a specific function form of ϵ_i:

$$\epsilon_i = exp(x_i\beta)\psi_i, \tag{4.7}$$

where ψ_i is a random variable, which is not dependent on the explanatory variable x_i. Since in this case ω_i equals to this random variable ψ_i, also ω_i is statistical independent on x_i. Thus, for the $E\left(ln(\omega_i)|x\right)$ holds that it is constant.

Following Silva & Tenreyro , we have now demonstrated that the OLS technique based on the logarithmic transformation provides consistent estimation only by imposing very specific assumptions on error term .

Further, Silva & Tenreyro argue that when ω_i is not dependent on the regressors, the conditional variance of y_i and also ϵ_i is proportional to $exp(2x_i\beta)$. Next, Silva & Tenreyro also investigate that ϵ_i is very likely to be heteroscedastic but there is no reason why the conditional variance of ϵ_i should be exactly proportional to $exp(2x_i\beta)$.

We can summarize these findings that in the estimation of log-linearized version of the constant elasticity model generally inconsistent.

4.3.2 Estimating gravity equation using cross-sectional data

We showed in the previous subsection that the estimation of the logarithmic transformed constant-elasticity model is inconsistent in the presence of heteroscedasticity. We apply these findings on the estimation of gravity model.

In chapter 2, we derived that our desire is to estimate the slope coefficients in the traditional gravity model represented by equation (3.13). As we mentioned,

[2]We assume that y_i is strictly positive.

the traditional approach is to log-linearize this model:

$$ln\left(X_{ji}\right) = ln\left(\beta_0\right) + \beta_1 ln\left(GDP_j\right) + \beta_2 ln\left(GDP_i\right) + \beta_3 ln\left(D_{ij}\right) + ln\left(\epsilon_{ij}\right), \quad (4.8)$$

and estimate by OLS the coefficients of this transformed model. According to the findings derived above, let us discuss more in detail the question of consistency of the estimated coefficients by this approach.

Firstly, let us recall that applying the estimation technique on the transformed model, we would like to evince the conditional mean of the logarithm of the trade flow $E\left(ln(X_{ji})|GDP_i, GDP_j, D_{ij}\right)$. Plugging (4.8) for the $ln(X_{ji})$, we get for the conditional mean:

$$\begin{aligned} E\left(ln(X_{ji})|GDP_i, GDP_j, D_{ij})\right) = E\left(ln(X_{ji})|ln\left(\beta_0\right) + \beta_1 ln\left(GDP_j\right)\right. \\ \left. + \beta_2 ln\left(GDP_i\right) + \beta_3 ln\left(D_{ij}\right) + ln\left(\epsilon_{ij}\right)\right). \quad (4.9) \end{aligned}$$

Applying the additivity property of the expected value, we further obtain for the conditional mean:

$$\begin{aligned} E\left(ln(X_{ji})|GDP_i, GDP_j, D_{ij})\right) = & \; ln\left(\beta_0\right) + \beta_1 ln\left(GDP_j\right) + \beta_2 ln\left(GDP_i\right) \\ & + \beta_3 ln\left(D_{ij}\right) + \beta_2 ln\left(GDP_i\right) + \beta_3 ln\left(D_{ij}\right) \\ & + E\left(ln(\epsilon_{ij})|GDP_i, GDP_j, D_{ij})\right). \quad (4.10) \end{aligned}$$

Thus, all of the coefficients including intercept are estimated consistently only if

$$E\left(ln(\epsilon_{ij})|GDP_i, GDP_j, D_{ij}\right) = 0.$$

Moreover, for the consistent estimation of the slope coefficients, it is sufficient if $E\left(ln(\epsilon_i)|GDP_i, GDP_j, D_{ij}\right)$ is constant.

Next, we know that

$$E\left((\epsilon_{ij}|GDP_i, GDP_j, D_{ij}\right) = 1$$

and so

$$ln(E\left(\epsilon_{ij}|GDP_i, GDP_j, D_{ij}\right) = 0.$$

However, due to Jensen's inequality,

$$ln(E\left(\epsilon_{ij}|GDP_i, GDP_j, D_{ij}\right) \neq E\left(ln(\epsilon_{ij})|GDP_i, GDP_j, D_{ij}\right).$$

On the other hand, it holds for the expected value of a random variable in logarithmic form that it is a function of its mean and also higher-order moments of the distribution. For instance, if ϵ_{ij} is log-normally distributed[3] with σ_{ij}^2 variance, then

$$E\left(ln(\epsilon_{ij})|GDP_i, GDP_j, D_{ij}\right) = ln\left(\frac{1}{1+\sigma_{ij}^2}\right). \tag{4.11}$$

Thus, to fulfill the consistency condition for all of the estimators including intercept, we need $\sigma_{ij}^2 = 0$. However, this will never hold. If we allow only for the intercept to be inconsistently estimated, then we require σ_{ij}^2 to be a constant and thus, we require the error term being homoscedastic. Moreover, if the error term is heteroscedastic and so it is a function of regressors, the OLS estimation is always inconsistent.

We can conclude that in the case of log-normally distributed error term, the OLS estimation is at least inconsistent estimation of intercept in the case of homoscedastic error term, while it is inconsistent estimation of all the regressors in the case of heteroscedastic error term.

4.3.3 Estimating gravity equation using panel data

In chapter 2, we also derived the stochastic version of gravity model, when panel data are employed for the estimation. As we mentioned, the log-linearized model of gravity equation with country-specific unobserved effect is applied for the estimation of the coefficients. Let us recall the equation:

$$ln(V_{jit}) = ln(\beta_0) + \beta_1 ln(\tau_{jit}) + \beta_2 ln(GDP_{jt}) + \beta_3 ln(GDP_{it}) + \theta_j + \theta_i + ln(\epsilon_{ijt}). \tag{4.12}$$

When we choose the fixed effects technique of estimation, we probably will decide for the dummy regression fixed effect to estimate the time invariant variables as well. Namely, we include the dummy variables for the importer and exporter fixed effect and estimate the model by OLS.

Similarly to the estimation of gravity model using cross-sectional data, we require for the transformed error term $E\left(ln(\epsilon_{jit})\right)|GDP_{it}, GDP_{jt}, \tau_{jit}, \theta_i, \theta_j)$ to be a constant to provide consistent estimates of the slope coefficients or to be zero to provide consistent estimates of all regressors including intercept.

For the same reasons as in the cross-sectional data case, this condition is

[3] We know that the mean is $E\left(\epsilon_{ij}|GDP_i, GDP_j, D_{ij}\right) = 1$.

met only in presence of homoscedastic error term implying that only the slope coefficients are estimated consistently.

Let us also discuss the condition for the consistent random effects estimation. When we treat the unobserved effects as random, then the assume a composite error term ψ_{ijt} with a form:

$$\psi_{ijt} = \theta_j + \theta_i + ln(\epsilon_{ijt}). \tag{4.13}$$

Next, for the the conditional expected values of the components of the composite error term holds that

$$E\left(ln(\epsilon_{jit})|GDP_i, GDP_j, \tau_{ji}, \theta_i, \theta_j\right),$$

is a constant and

$$E\left(\theta_i, \theta_j|GDP_i, GDP_j, \tau_{ji}\right),$$

is a constant as well to estimate the slope coefficients consistently. To estimate also the intercept consistently, these conditional values need to equal to zero.

We can see that for the condition of the logarithm of the error term, we can make the same conclusions as in the previous analysis.

4.4 Poisson pseudo-maximum likelihood estimation technique

The alternative approach to the estimation of log-linearized model lies in direct estimation of the multiplicative form of the gravity equation:

$$E\left(V_{ji}|GDP_i, GDP_j, \tau_{ji}, \theta_i, \theta_j\right) = \beta_0 \; \tau_{ji}^{\beta_1} GDP_j^{\beta_2} GDP_i^{\beta_3} e^{\theta_j} e^{\theta_i}. \tag{4.14}$$

According to Hausman *et al.* (1984), similar types of equations, based on non-count data measured in non-negative integers, can be estimated by Poisson pseudo maximum likelihood estimator.

Let us describe the first order condition for maximizing the likelihood function in the case of fixed effects model. To pursue time invariant variables, let us describe the pooled model with cross-country fixed effect. We follow (Wooldridge 2001).

Firstly, the gravity equation could be rewritten in the simplified form of the dependant variable y_{it} and independent variables x_{it} including proxy for fixed effects and for the β parameters to be estimated.

We assume that the the conditional expectation is proportional to the moment function m of x and β. This conditional expectation takes usually a form of $exp(x_t\beta)$. The estimation of the coefficients of interest is a solution to the maximization of the log-likelihood function:

$$l_i(\beta) = \sum_{t=1}^{T} (y_{it} log[m(x_{it}, \beta)] - m(x_{it}, \beta)) \tag{4.15}$$

Chapter 5

Empirical Analysis of Trade

The aim of this and next chapter is to compare the performance of Poisson pseudo-maximum likelihood panel data estimation techniques with the traditional panel data estimation techniques based on logarithmic transformation of gravity equation. Both proposed estimation techniques are applied on real panel data sets of Germany and the Czech Republic. In this chapter, we discuss the data, variables and methodology of the estimation, while Chapter 6, we investigate the results.

Germany was chosen as one of the analyzed home countries because it passes for a free, economically strong country with long tradition in international trade. Moreover, it is a member of the European Union and European currency union. On the other hand, the Czech Republic was chosen as a smaller open country with less developed economy comparing to Germany. The Czech Republic is also not a key player on the field of international trade in the world. Moreover, there exist some countries in the world which were even not or very small trade partners with the Czech Republic in some years.

In addition, analyzing trade data set of these two countries allow us comparing the performance of the estimators on different types of countries.

5.1 Data

We analyze a strongly balanced panel data sets of Germany and the Czech Republic and their 177 trade partners in time period 1995-2009 yielding 2655 observations for each data set. We use the same trade partner—s countries in both sets. The time period is interesting in the point of view that it covers years of economic boom but also years of economic recession. The data used for this empirical analysis were collected from the databases of Eurostat, the World Bank, IMF, European Commission, Geobytes, CEPII, Heritage Foundation, Mannheim enterprise database, WTO, database on Tenders of Public Procurement in the EU and national statistical offices.

5.2 Econometric Specification and Variables

The traditional approach based on multi-country models usually studies the huge trade panel data sets. These data sets suffer from the fact that they incorporate many various information. When we estimate the gravity equation in this framework, meaning in general for all the countries, we probably loose some information.

Our approach is different from the multi-country or bi-country gravity models. We target our analysis on the one-way trade flow of home country. So, we are enabled to study the relationships in the gravity model in more specific way. Defining Germany or the Czech Republic as a single "home country" i, the analysis is based on an econometric estimation of export function from the home country to its trade partners.

According to the gravity theoretical concept, we define German or Czech exports X_{ijt} from country i to country j in year t as a stochastic function of economic variables representing: GDP Y_{it} of country i ; GDP Y_{jt} of country j ; population size L_{jt} of partner's country j; currency volatility C_{ijt}; real exchange rate ER_{ijt}; bilateral resistance to trade represented by the distance D_{ijt} between countries, trade barrier quality T_{ijt} and institutional variables included in vector U_{ijt}; fixed effects term θ_{jt} of partner's country j; and time effect λ_t:

$$X_{jit} = \beta_0 Y_{it}^{\beta_1} Y_{jt}^{\beta_2} L_{jt}^{\beta_3}\ D_{ijt}^{\beta_4} e^{(\beta_5 C_{ijt} + \beta_6 T_{ijt} + \beta_7 ER_{ijt} + \beta_8 U_{ijt} + \theta_j t + \lambda_t)} \epsilon_{ijt}, \tag{5.1}$$

Exponents of these variables are the parameters to be estimated and ϵ_{ijt} stands for the error term.

Our specification also contains additional variables (e.g. C_{ijt}, L_{jt} or T_{ijt}) to the empirical Anderson-Wincoop model defined in equation 3.16. We assume that these variables are covered in the "trade-cost" function τ_{ijt} from equation 3.16.

5.2.1 Variables

Let us discuss in this subsection the meaning of each variable and expected magnitude and sign of its parameter.

Firstly, as we mentioned in chapter three, (Anderson 1979) assume constant unit-elasticity gravity equation. Applying this assumption to our specification, it implies that $\beta_1 = \beta_2 = 1$. Further, the variable population size L_{jt} is considered as a proxy variable for the "openness" of the country. We assume that the larger and more populated country, the less open is to the international trade. Thus the sign of β_3 is expected to be negative.

A sign of the parameter for the distance variable D_{ijt} is naturally presumed to be negative because the variable originally finds out in the denominator of the gravity equation. Further, currency volatility C_{ijt}, represented particularly by euro, is involved in the model. The expected sign of its parameter β_5 is essentially positive.

Last but not least, a variable T_{ijt} qualifies trade barriers between countries, which it essentially implies the expected influence on the trade flow between countries being negative. The negative influence on the trade flow is also naturally presumed by the real exchange rate variable ER_{ijt}.

Finally,U_{ijt} is a vector which consist of institutional and policy factors of trade related to transaction costs, externalities and risks, such as the integration alignment, shared policies, impacts of the currency, the intensity of impediments to cooperation, infrastructure, government procurement, cultural and "psychological" distance, extending so the existing literature [Babecká Kucharčuková *et al.* (2010);Bussiere *et al.* (2008); etc.]. However, we assume high correlation among these variables, so only a few of the are really employed for the estimation.

We also use dummy variables for year 2008 and 2009 as an indicator of the economic recession. We assume that these variables have negative influence of the export.

5.3 Methodology

To perform econometric estimation of equation 5.1, we employ various panel data estimation techniques. All of the procedures are applied in traditional way and also as Poisson estimation. We provide both random effects and fixed effects estimation techniques.

In empirical literature, the econometric model of gravity equation contains many time-invariant or nearly time-invariant variables. For instance, in our model some important variables such as distance, population, currency volatility or trade barriers evince very small within variation in our panel data sets. Applying the traditional fixed effects method of estimation, all of these variables would be omitted during the regression. However, we would like to determine the influence of this variables on the trade flow. Moreover, we assume a high explanatory power of these variables.

To solve this problem, there is a need to estimate the gravity equation by a different type of fixed effects estimation technique. According to our econometric specification, we offer dummy regression as fixed effects model with importer fixed effects a possible appropriate estimator. This technique is based on the assumption, that the fixed effects of partner's countries could be proxied by the bunch of country specific dummy variables. Taking this assumption, we can estimate the gravity equation by not exactly panel data estimation methods. OLS in the classical framework and PPML for the Poisson framework are applied. To estimate the standard errors consistently, we need to control for serial correlation among observation using clustering method.

Next, the model also considers the presence of a time effect component. We proceed similarly as in the previous case by incorporating this component into the regression. We create year specific dummy variables as proxy variables. We add these dummy variables to the dummy model to estimate two-way fixed effects. We also assume the case that only the time effect is presented in the model by estimating the equation as a pooled model (by OLS and by PPML).

In the case, that the panel data's time series and cross-sectional series were incorrectly pooled, distinguish between the country-specific and time effect enable us to correctly determine how the decision-making of agents evolves in time and how it differs between countries.

We also estimate the gravity equation by random effects. For the PPML estimations, there is a need to control for heteroscedasticity by estimating the standard errors by bootstraping.

To compare the correct specification of fixed effects and pooled models, the two proposed estimation methods, we perform the traditional RESET (Ramsey Regression Equation Specification Error Test). We predict the fitted values for the various specifications. Next, we include a higher order of fitted values into the regression. If the model is correctly specified, the fitted values term would be confirmed as insignificant[1].In the case of random effects model, we perform Hausman specification test.

Finally, we calculate the goodness-of-fit statistics[2] for the predicted export. This statistics is calculated for the untransformed model. This method allows us to see, how is the logarithmic transformation distortive. The statistics is calculated in following way:

$$\chi^2_{red} = \frac{1}{N-k-1} \sum \frac{(X_{ijt}fitted - X_{ijt})^2}{\sigma^2}, \tag{5.2}$$

where $X_{ijt}fitted$ stands for the fitted values of export; X_{ijt} is the actual export; σ^2 is the variance of the measurement error; N stands for the number of observation and k is the number of parameters in the model. If the statistics is close to one,our model fits the data very well.

[1]We test this hypothesis by F-test in the case of OLS estimation and by χ^2 in the case of Poisson regression

[2]source of the formula: www.wikipedia.com

Chapter 6

Results

This chapter is divided in two sections by country we are analyzing. We target on investigating differences between the estimation based on logarithmic transformation and the Poisson pseudo-maximum likelihood estimation. Our primary interest is to determine, weather the estimated coefficients are valid according to the theory of gravity models. We also concern the proper model specification by RESET test and evaluate a fit of the estimated model.

As we mentioned before, we consider also the institutional variables measuring a quality of institutes in the partner's countries as variables in the model. We detected that these variables are often correlated. For the estimation, we choose only some of them: fiscal freedom, government spending, financial freedom, property rights, freedom corrupt and education.

6.1 Germany

6.1.1 Pooled Estimation with Time Effect

The resulting estimated coefficient of the pooled model with time effect are summarized in Table 6.1. We can conclude that impact of GDP of the partner's country is estimated strongly significant for both OLS and PPML. On the other hand, PPML estimation suggests higher elasticity of partner's GDP than OLS predicts. Moreover, PPML estimates the elasticity partner's GDP 0.9, which is closer to the by theory suggested unit elasticity.

Similarly, both types of estimates find the distance as the most important trade deterrent. PPML predicts the negative influence of distance a little bit lower; its predicted elasticity is by 15 percent higher. Next, both estimation confirm that exchange rate and the recension years decrease export. However, we can see that the coefficients estimated by OLS are lower than the coefficients assumed by PPML.

However, only PPML suggests the currency (euro) volatility being significant and positive. Moreover, the biggest difference between these two estimation technique is that due to OLS higher partner's population increases the expected value of export. In contrary, we assume that higher partner's population decreases the expected export because the more populated and bigger countries are less open to the world.On the other hand, PPML estimates the influence of population according to to our assumptions but does not predict it to be significant.

Last but not least, we can also notice that only OLS predicts the institutional variables as significant for the explanation of German exports. In our opinion, this is caused by the heterogeneity in the data. While PPML naturally deals with heterogeneity, OLS uses these institutional variables to catch it up.

Next, we can conclude that the the PPML model is incorrectly specified; RESET test rejected the hypnoses of no omitted variables and proper specification. OLS model is due to RESET specified properly. On the other hand, when we employ higher order of fitted values for the RESET test, OLS model is also rejected as a correct specification of gravity model.

Finally, we also measured the goodness of fit for both models. We can see that both reduced χ^2 statistics are very similar and close to one. We can conclude that both models fit the data very well. However, the OLS estimation fits data a little bit better.

6.1.2 Country-specific Fixed Effects Model Estimation

The performed regression for both estimation technique in summarized in Table 6.2. When we compare the results of the estimated LSDV regression with the estimated pooled model with time effect, nearly all of the coefficients are in the case of this model estimated on lower level. The PPML estimates remain more or less very similar. However, coefficients of OLS change more significantly. For instance, the elasticity of distance significantly increases by about 14 per cent and comes closer to PPML estimation. Similarly, the influence of exchange rate predicts OLS to be higher in that case. Moreover, also the coefficient of population increases under the LSDV specification.

Further, comparing the results of RESET tests, we argue that only PPML

Table 6.1: Germany: Pooled Estimation when Controlling for Time Effect

Variable	OLS	PPML
$log\,(y_{it})$	.13136498	.38131167*
$log\,(y_{jt})$	.65642439***	.89836004***
$log\,(l_{jt})$	.31602936*	-.07778971
$log\,(d_{ij})$	-.95918182***	-.81792587***
c_{ijt}	.13963835	.16516108*
t_{ijt}	.01179028	.02101356
er_{ijt}	-.29820327**	-.19631704*
fis	.00696085*	.00444946
gov	.0035757	-.00012891
fin	-.00383187	.00042539
pro	.00688462*	.00533965
cor	.01137964*	-.00171886
edu	.02221316***	.00994845
$year$ 2008	-.07185233*	-.04223227***
$year$ 2009	-.22393132***	-.1893204***
$RESETtest$	F(1,176) = 2.15	chi2(1) = 7.97
	Prob > F=0.1443	Prob > chi2 = 0.0048
$Goodnessoffit$	$chi2_{red}$ =1.012117122	$chi2_{red}$=1.012841785

Table 6.2: Germany: Fixed effects estimation with importer fixed effects

Variable	OLS	PPML
$log\,(y_{it})$	.32136962	.36592425*
$log\,(y_{jt})$	.54254974**	.82029949***
$log\,(l_{jt})$	.38819104*	-.02436987
$log\,(d_{ij})$	-.8148262***	-.80074789***
c_{ijt}	.12498686	.15069653
t_{ijt}	-.04979382	.01292356
er_{ijt}	-.16064819	-.1905623*
fis	.00517944	.00243219
gov	.00411581	.00108627
fin	-.00103664	.00074441
pro	.00782706*	.00611548
cor	.00525435	-.00397649
edu	.03227559***	.01747421**
$year$ 2008	-.06787669	-.03455236
$year$ 2009	-.2325794***	-.18697065***
$RESETtest$	F(1,176) = 5.34	chi2(1) = 3.41
	Prob > F=0.0220	Prob > chi2 = 0.0649
$Goodnessoffit$	$chi2_{red}$ =1.047613362	$chi2_{red}$=1.039996187

model is correctly specified. Finally, the lower value of statistic of goodness of fit also confirm Poisson estimation to fit better the data.

6.1.3 Country-specific Fixed Effects Model Estimation Controlling for Time Effect

Table 6.3. summarizes the LSDV estimation when we add the time effect into the model. Including the time effect, the elasticity of partner's GDP decreases significantly to 0.79 for the PPML estimation but still it is close to the theoretical unit-elasticity hypothesis. The estimated elasticity of population again increases for both estimation techniques. The OLS predicts this elasticity to be significant with value about 40 percent. Furthermore, PPML also estimates the population elasticity to be positive but not significant.

Next, the incorrect specification of the model was detected by the RESET test for both estimation techniques. It seems that including additional time dummies deteriorates the model specification by loosing additional degrees of freedom. Including only recession dummies provides better model specification.

Table 6.3: Germany:Fixed effects estimation with importer fixed effects Controlling for Time Effect

Variable	OLS	PPML
$log\,(y_{it})$	.21151185	.44448059*
$log\,(y_{jt})$	.53309847**	.78651974***
$log\,(l_{jt})$	.40003714*	.00843624
$log\,(d_{ij})$	-.81949468***	-.80261999***
c_{ijt}	.1311927	.16605849*
t_{ijt}	-.04549316	.01718478
er_{ijt}	-.19128299	-.2278457*
fis	.00561118	.00356088
gov	.00367317	.00049461
fin	-.00094969	.00108441
pro	.00756823*	.00559386
cor	.0050959	-.00364016
edu	.03221836***	.01777586***
$year$ 2008	-.05255841	-.04379804***
$year$ 2009	-.22048987***	-.18856919***
$RESETtest$	t = 2.28	z = -2.38
	Prob > t=0.024	Prob > z = 0.016
$Goodness of fit$	$chi2_{red}$=1.052370993	$chi2_{red}$=1.039996187

Finally,the Poisson estimation provides better ability to fit the data in this model comparing with OLS estimation.

6.1.4 Random Effects Estimation

Under assumption that the country's specific fixed effect is a random variable which is not correlated with other regressors, random effects estimates are consistent and effective. We also estimate gravity equation by random effects estimation technique; the results are found in Table 6.4.

We can conclude that under random effects specification the partner's GDP elasticity reaches its highest value 0.78 among other estimated elasticities and is strongly significant. The value of partner's GDP elasticity estimated by PPLM even exceeds the unity and takes the value of 1.25. The elasticity of population remains positive and significant and decreases significantly in the case of standard random effects estimation techniques. Whereas PPML estimation constantly suggests the elasticity of population being negative and

Table 6.4: Germany: Random Effects Estimation

Variable	RE standard	RE poisson
$log\,(y_{it})$	.54003024***	.06323254
$log\,(y_{jt})$	.78004786***	1.2450216***
$log\,(l_{jt})$	.1365572**	-.70893665*
$log\,(d_{ij})$	-.96635855***	-.37334565
c_{ijt}	-.03164555	.00407415
t_{ijt}	-.03053715	-.02444743
er_{ijt}	-.16628333***	-.15990861***
fis	.00191678	-.00041756
gov	.00396688***	.00471652**
fin	.00244575*	.00098621
pro	.00594064***	-.00031178
cor	.00233537	.00001835
edu	.00791231***	.00164016
$year$ 2008	-.08553626*	-.05957013**
$year$ 2009	-.22241144***	-.2223397***
$Goodnessoffit$	$chi2_{red}$=1.005895064	$chi2_{red}$=1.009830317

in the case of random effects estimation it has strong negative influence on the estimated exports. Other estimated coefficients remain very similar.

However, we reject by Hausman specification test that the assumption of random effects model are fulfilled. That means that the estimates performed by both standard and Poisson random effects model estimation are not consistent.

At last, the traditional random effects estimation was concluded to fit better the data than Poisson random effects.

6.2 Czech Republic

6.2.1 Pooled Estimation with Time Effect

Table 6.5. presents the estimated coefficients of the gravity equation when we control only for the time effect. We can recognize that both method confirm a significant impact of the home and partner's GDP on the value of exports. On the other hand, the predicted impact proposed by OLS is about one half smaller than the impact predicted by PPML. Moreover, PPML suggests nearly unit elasticity of GDP for both home and partner country, according to Anderson.

Furter, we can observe very different results also for the partner's population. While OLS predicts strongly significant positive impact on exports with nearly 46 percent elasticity, PPML suggests insignificant negative influence on exports with -8 percent elasticity. Let us recall that the expected sign of the population elasticity is negative.

Last but not least, OLS confirms negative impact of the real exchange rate with the value and PPML confirms that if the partner is a part of the currency union,it significantly increases exports of the home-country. And finally, both techniques suggest that distance passes for the most significant trade deterrent, however, OLS predicts the elasticity of distance higher in absolute value by 8 percentage points.

On the other hand, the other estimated coefficients are similar for both techniques, only OLS predicts some of the institutional variables as significant. We explain the significance of the institutional variables in the same way as in the case of Germany.

Last but not least, we can identify that only OLS provides proper estimation results, according to RESET test. On the other hand, RESET test based on higher order fitted values rejected the hypothesis of correct OLS specification.

Finally, we also calculated the goodness of fit statistics for both models. We can see that both reduced χ^2 statistics are very similar and close to one. We can conclude that both models fit the data very well. However, the Poisson estimation fits data a little bit better.

6.2.2 Country-specific Fixed Effects Model Estimation

As shown in Table 6.6., the results of the estimation of a country-specific fixed effects model compared techniques are very similar to the results in the model with only Time effect. However, we can detect some differences. For instance, estimating the gravity model under this fixed effect specification, only OLS model predicts a significant positive impact of the currency volatility on exports.

On the other hand, only PPML estimation deduces that in the recession year 2009 the exports of the Czech Republic significantly decreased. Moreover, the small difference in the predicted elasticity of distance between the two estimation technique is in this specification nearly suppressed to the similar value of about -1.6.

Last but not least, only PPML model was identified as a correct specifica-

Table 6.5: Czech Republic:Pooled Estimation when Controlling for Time Effect

Variable	OLS	PPML
$log\,(y_{it})$	.47923292*	1.241171*
$log\,(y_{jt})$	.57535144***	1.0126592***
$log\,(l_{jt})$	.46199674***	-.08266925
$log\,(d_{ij})$	-1.657342***	-1.594765***
c_{ijt}	.21747735	.1924412*
t_{ijt}	.04686773	.02219882
er_{ijt}	-.20111515*	.16935776
fis	.01257924**	.01761892
gov	.00764017*	.00253465
fin	.00031383	.00402042
pro	.00597383	.00104031
cor	.00495309	.00354171
edu	.03491408***	.01168062
$year$ 2008	.01906898	.02791821
$year$ 2009	-.03793476	-.10462332
$RESET test$	F(1,176) = 0.13	chi2(1) = 5.65
	Prob > F=0.7170	Prob > chi2=0.0175
$Goodness of fit$	$chi2_{red}$ =1.011493261	$chi2_{red}$=1.010779813

Table 6.6: Czech Republic: Fixed effects estimation with importer fixed effects

Variable	POLS	PPML
$log\,(y_{it})$	.75454356***	1.2556946*
$log\,(y_{jt})$	.58570432***	1.0001173**
$log\,(l_{jt})$	.41703476**	-.07472594
$log\,(d_{ij})$	-1.6213002***	-1.6044091***
c_{ijt}	.26115109*	.17236826
t_{ijt}	.05148159	.00875241
er_{ijt}	-.0925616	.22906122
fis	.00879	.01455575
gov	.0069102	.00413351
fin	.00179641	.00338205
pro	.00592723	.001415
cor	-.00098126	.00203985
edu	.03771312***	.01989994
$year$ 2008	.07473072	.03814763
$year$ 2009	.01749076	-.10618543***
$RESETtest$	F(1,176) = 5.29	chi2(1) = 1.87
	Prob > F=0.0227	Prob > chi2 = 0.1715
$Goodnessoffit$	$chi2_{red}$ =1.046556251	$chi2_{red}$=1.041720431

tion of the gravity equation. The hypothesis of proper specification cannot be rejected at a high significance level.

Finally, the ability of the model to fit the data can be concluded very similar and powerful for both types of estimation techniques. However, the Poisson predicts the coefficient to fit the the data a little bit better than OLS.

6.2.3 Country-specific Fixed Effects Model Estimation Controlling for Time Effect

When we add also the time effect into the model with country-specific fixed effects, the difference between different types of estimation methods remain very similar. The estimation results are presented in Table 6.7.It is apparent from this table that foremost the difference in estimated elasticity of home country is deepened by adding the time effect into the regression, since OLS predicts the elasticity only at the value of 0.62

At last, the results of RESET test and goodness of fit statistics lead to the same conclusion as in the previous model without time effect.

Table 6.7: Czech Republic: Fixed effects estimation with importer fixed effects Controlling for Time Effect

Variable	OLS	PPML
$log\,(y_{it})$	.62479365**	1.2503933*
$log\,(y_{jt})$	.58918253***	.95815568**
$log\,(l_{jt})$	.41260559**	-.03034456
$log\,(d_{ij})$	-1.6194507***	-1.6102356***
c_{ijt}	.26590142*	.19776868
t_{ijt}	.05019193	.01500725
er_{ijt}	-.07650519	.20270225
fis	.00861305	.01552194
gov	.00721709	.00388013
fin	.00226652	.00416674
pro	.00577887	.00081442
cor	-.00093766	.00264785
edu	.03790683***	.02039469
$year$ 2008	.03632022	.03044637
$year$ 2009	-.02993371	-.11210392
$RESETtest$	F(1,176) = 5.11	chi2(1) = 1.85
	Prob > F=0.0249	Prob > chi2 = 0.1742
$Goodnessoffit$	$chi2_{red}$=1.04775547	$chi2_{red}$=1.0423856

6.2.4 Random Effects Estimation

We also compare the performance of the two estimation techniques using random effects models. However, we recognized by the Hausman specification test that the assumptions on consistency of RE estimator were significantly not met for both standard and Poisson estimation.

Table 6.8.summarizes the estimated coefficients for both techniques. The PPML predicts very strange elasticity of the home GDP, when the estimated coefficient reaches value higher than 2. On the other hand, the standard random effects method produces very reasonable results in the case of home and partner's GDP with the values of 84 per cent for home and 77 per cent for partner's country.

In the case of OLS estimating technique, the partner's population elasticity is still positive and very significant but it reaches its lowest value among other standard panel data models for the Czech Republic. The PPML estimation method also suggests positive partner's population elasticity. However, the estimated impact is not significant.

The significant positive impact of euro adoption was predicted only by the standard RE model. Further, standard RE model suggests that the exchange rate decreases trade at very high significance level. The PPML random effects model also predicts the negative influence of exchange rate but suggests a lower coefficient estimated on a lower significance level.

Finally, we also compare the goodness of fit for both the estimation methods. We can see that the χ^2 statistics is lower than one in the case of Poisson regression. This happens when the model over-fits the data. On the other hand, the traditional random effects estimation method provides estimates which fit the data very well.

Table 6.8: Czech Republic: RE Estimation

Variable	RE standard	RE poisson
$log\,(y_{it})$	.84250179***	2.2939026***
$log\,(y_{jt})$	.76518051***	.63190803
$log\,(l_{jt})$	.27178725***	.43530609
$log\,(d_{ij})$	-1.6931895***	-1.4908341***
c_{ijt}	.2995425***	.05487512
t_{ijt}	.08936008**	.03521743
er_{ijt}	-.23750229***	-.11313064*
fis	.00269996	-.00634849
gov	.00858404***	.00273458
fin	.00496691**	-.00039917
pro	.00755392***	.00078416
cor	-.00407051	-.00429402
edu	.02178193***	-.01884197
$year$ 2008	-.01254164	-.05181161*
$year$ 2009	-.03912757	-.11281874***
$Goodnessoffit$	$chi2_{red}$ =1.048113139	$chi2_{red}$=1.18794E-06

Chapter 7

Conclusion

This thesis has analyzed the potential negative effect of logarithmic transformation on the quality of estimation of gravity model of international trade using panel data. We have discussed this issue from both the theoretical and empirical point of view. In the theoretical framework of estimation of gravity model, we have shown that in the presence of heteroscedasticity the estimation of logarithmic transformed model causes inconsistency of the estimation. Thus, we recommend, according to the literature, the Poisson pseudo maximum likelihood estimator, which is applied immediately on the multiplicative form of gravity equation. We have also tested these conclusions on the real panel data sets of the Czech Republic and Germany. We have provided estimation of four types of panel data models: pooled model with time effect; importer fixed effects; importer fixed effects with time effect; and random effects. We have applied the traditional estimation technique based of log-linearization of the model and Poisson pseudo maximum likelihood estimation technique to compare these two approaches.

After the analysis of German and Czech data, we can conclude that both estimation techniques have their advantages and disadvantages. The Poisson estimation technique has predicted coefficients of statistically significant variables, which have been always with expected signs and usually with value close to the value suggested with the theory of gravity models. On the other hand, the OLS estimations have sometimes suggested signs of the coefficients, which are not in accordance with our expectation, even if the connected variables were statistically significant due to the estimation. For instance, OLS and other traditional estimation techniques have failed in all cases when they predict the impact of the partner's population on home export. The theory suggests a negative impact of partner's population because it is expected that population could be understand as a proxy for the openness of the country. We assume that bigger country is less open. However, the traditional estimation methods have predicted a significant positive impact of the partner's population.

Last but not least, both estimation methods have suffered in some panel data models from improper specification detected by the RESET test. This test investigates whether if we add a higher order of the fitted values into regression, these fitted values are significant for the estimation. According to our expectations, both methods have provided an inconsistent estimation of random effects model. Further, for the German data, the estimation of the model with fixed effects of importer and with time effect has been also detected that it suffers from improper specification for both estimation methods. However, in the case of Czech Republic, the only Poisson estimation results have been concluded as econometrically correct for the fixed effect model with importer fixed effects and time effect. For both countries, the OLS technique has been suggested that it is correctly specified for the pooled model with only time effect. On the other hand, for both countries the Poisson estimation technique has been chosen as correctly specified for the model with importer fixed effects.

Moreover, we can further conclude from the German and Czech estimation results that the better score of fitted values has confirmed OLS as a better method in the case of pooled model, while Poisson estimation method has been superior for the fixed effect model. The Poisson estimation technique has also performed a better fitting ability to the model with importer fixed effects and time effect. The goodness-of-fit statistics for this investigation has been calculated in the way to evaluate the how the predicted trade flow fits the original rather than transformed data.

In addition, the only estimation method, which has been performing correct specification, has provided good estimation results according to the theory of gravity model and also its results have fitted the data well is the Poisson pseudo maximum likelihood estimator, which is applied on the fixed effects model with importer fixed effects. This conclusion holds for both countries. Moreover, for the Czech data analysis, we can also evaluate the Poisson estimation method, which is applied on the fixed effects model with importer fixed effects and time effect, as the proper estimation technique.

In conclusion, we have demonstrated theoretically that the estimation of gravity equation based on the logarithmic transformation is in the presence of heteroscedasticity highly miss-leading, even when we concern on the panel data usage. The empirical analysis of the real Czech and German panel data sets has also evaluated the Poisson estimation technique, which is based on the estimation of the multiplicative form of the gravity equation, as more advisable than the traditional methods based on logarithmic transformation.

Bibliography

ANDERSON, J. E. (1979): "A theoretical foundation for the gravity equation." *American Economic Review* **69(1)**: pp. 106–16.

ANDERSON, J. E. & E. VAN WINCOOP (2001): "Gravity with gravitas: A solution to the border puzzle." *NBER Working Papers 8079*, National Bureau of Economic Research, Inc.

ARVIS & SHEPHERD (2011): "The poisson quasi-maximum likelihood estimator: A solution to the â€oadding upâ€? problem in gravity models." *MPRA Paper 34334*, University Library of Munich, Germany.

BABECKÁ KUCHARČUKOVÁ, O., J. BABECKÝ, & M. RAISER (2010): "Gravity approach for modelling international trade in south-eastern europe and the commonwealth of independent states: The role of geography, policy and institutions." *Open Economies Review* .

BALDWIN, R. & D. TAGLIONI (2006): "Gravity for dummies and dummies for gravity equations." *CEPR Discussion Papers 5850*, C.E.P.R. Discussion Papers.

BURGER, M., F. v. OORT, & G. LINDERS (2009): "On the specification of the gravity model of trade: Zeros, excess zeros and zero-inflated estimation." *Research Paper ERS-2009-003-ORG*, Erasmus Research Institute of Management (ERIM), ERIM is the joint research institute of the Rotterdam School of Management, Erasmus University and the Erasmus School of Economics (ESE) at Erasmus Uni.

BUSSIERE, M., J. FIDRMUC, & B. SCHNATZ (2008): "Eu enlargement and trade integration: Lessons from a gravity model." *Review of Development Economics* **12(3)**: pp. 562–576.

Deardorff, A. V. (1995): "(determinants of bilateral trade: Does gravity work in a neoclassical world?" *Working Paper 5377*, National Bureau of Economic Research.

Egger, P. (2000): "A note on the proper econometric specification of the gravity equation." *Technical report*.

Egger, P. (2002): "An econometric view on the estimation of gravity models and the calculation of trade potentials." *The World Economy* **25(2)**: pp. 297–312.

Eichengreen, B. & D. A. Irwin (1996): "The role of history in bilateral trade flows." *NBER Working Papers 5565*, National Bureau of Economic Research, Inc.

Flowerdew, R. & M. Aitkin (1982): "A method of fitting the gravity model based on the poisson distribution*." *Journal of Regional Science* **22(2)**: pp. 191–202.

Hausman, J., B. H. Hall, & Z. Griliches (1984): "Econometric models for count data with an application to the patents-rd relationship." *Econometrica* **52(4)**: pp. 909–38.

Haworth, J. M. & P. J. Vincent (1979): "The stochastic disturbance specification and its implications for log-linear regression." *Environment and Planning A* **11(7)**: pp. 781–790.

Helpman, E. & P. Krugman (1990): "Trade policy and market structure : Elhanan helpman and paul krugman, (mit press, cambridge, ma) pp. 191, 24.95." *Journal of International Economics* **29(1-2)**: pp. 187–189.

Martin, W. & C. S. Pham (2008): "Estimating the gravity model when zero trade flows are frequent." *Economics Series* $2008_0 3$.

Martinez-Zarzoso, I. (2013): "The log of gravity revisited." *Applied Economics* **45(3)**: pp. 311–327.

Matyas, L. (1997): "Proper econometric specification of the gravity model." *The World Economy* **20(3)**: pp. 363–368.

Siliverstovs, B. & D. Schumacher (2007): "Estimating gravity equations: To log or not to log?" *Technical report*.

Silva, J. M. C. S. & S. Tenreyro (2006): "The log of gravity." *The Review of Economics and Statistics* **88(4)**: pp. 641–658.

Silva, J. S. & S. Tenreyro (2009): "Further simulation evidence on the performance of the poisson pseudo-maximum likelihood estimator." *Economics Discussion Papers 666*, University of Essex, Department of Economics.

Tinbergen, J. (1963): "Shaping the world economy." *The International Executive* **5(1)**: pp. 27–30.

Westerlund, J. & F. Wilhelmsson (2009): "Estimating the gravity model without gravity using panel data." *Applied Economics* **43(6)**: pp. 641–649.

Wooldridge, J. M. (2001): *Econometric Analysis of Cross Section and Panel Data*, volume 1 of *MIT Press Books*. The MIT Press.