AN APPLICATION OF THE GRAVITY MODEL TO ECOWAS COUNTRIES

Master Thesis

Amadou Ndongo Thiam

Eskisehir, August 2019

AN APPLICATION OF THE GRAVITY MODEL TO ECOWAS COUNTRIES

Amadou Ndongo Thiam

MASTER THESIS

Department of Economics

Supervisor: Prof. Dr. Mustafa Özer

Eskisehir

Anadolu University

Graduate School of Social Science

August 2019

This study was supported by Anadolu University Scientific Research Projects Commissions under the grant no: 10284043

FINAL APPROVAL FOR THESIS

This thesis titled "**An Application of the Gravity Model to ECOWAS Countries**" has been prepared and submitted by **Amadou Ndongo THIAM** in partial fullfillment of the requirements in "**Anadolu University Directive on Graduate Education and Examination**" for the Master of Arts in **Department of Economics** has been examined and approved on **19/08/2019.**

<u>**Committee Members**</u>		<u>**Signature**</u>
Member (Supervisor)	: Prof.Dr. Mustafa ÖZER	....
Member	: Assoc.Prof.Dr. Oytun MEÇİK	....
Member	: Assist.Prof.Dr. Sevilay KÜÇÜKSAKARYA	....

19/08/2019
Date

Prof.Dr.Bülent GÜNSOY
Director
Graduate School of Social Sciences

ABSTRACT

AN APPLICATION OF THE GRAVITY MODEL TO ECOWAS COUNTRIES

Amadou Ndongo Thiam

Department of Economics

Anadolu University, Graduate School of Social Science, August 2019

Supervisor: Prof. Dr. Mustafa Özer

This thesis examined the theoretical foundations of the gravitational equation, its specifications and different econometric approaches for its estimation. We adopt Anderson and Van Wincoop (2003) specification to see the effects of GDP, distance and population variables on the trade flow of ECOWAS countries. After substitution of the multilateral resistance effect by unilateral individual effects or remoteness indices, and the zero trade flows by bilateral effects, we show that the PPML method and the Heckman selection procedure remain the most efficient because they give us unbiased estimators and relatively high explanatory power. However, from the point of view of estimation by cointegration techniques, the results are similar to PPML estimation and the Heckman selection procedure and are both consistent with the theory. They only differ in predicting the level of development of ECOWAS countries. By using the Heckman estimation method to take into effects of qualitative variables such as the common border, common language and sharing a monetary union, we found that they may play a role in increasing the volume of trade in ECOWAS countries, while transport cost reduce it.

Keywords: Gravity model, Panel data methods, Cointegration, ECOWAS, Zero trade flow, PPML method, Heckman selection procedure.

ÖZET

Bu tezde, gravite eşitliğinin teorik temellerini, özelliklerini ve tahmin etmekte kullanılan farklı ekonometrik yöntemler incelenmektedir. Tezde, GSYİH, uzaklık ve nüfus değişkenlerinin ECOWAS ülkelerinin dış ticaret akımlarına etkilerini analiz edebilmek için, Anderson ve Van Wincoop tarafından geliştirilen model kullanılmıştır. Tek taraflı bireysel etkiler veya uzak değer indekslerinin çok taraflı direnç etkisinin yerine kullanılması ve sıfır ticaretin iki taraflı etkilerini dikkate larak, PPML yönteminin ve Heckman model seçim sürecinin en etkin yöntem olduğunu, yansız tahminciler ürettiğini ve daha yüksek açıklama gücü gösterdiğini ortya akoyduk. Bununla birlikte, eşbütünleşme yöntemleri kullanılarak yapılan tahminler yapılan tahminler, PPML tahminleri ile Heckman model seçim sürecine benzer sonuçlar vermekte ve her ikisi de teorik beklentilerle tutarlıdır. Farklı yöntemlerle elde edilenn tahminler ararsındak tek fark, ECOWAS ülkelerinin gelişmişlik düzeyini tahmin etmeden kaynaklanmaktadır. Ortak sınır, ortak dil ve parasl birliğin etkilerini görmek amacıyla kukla değişkenler kullanıp ve modeli Heckman yöntemi ile tahmin ederek, bu niteliksel etkilerin ECOWAS ülkerinde ticaret hacmini artırıcı rol oynayabileceğini bulduk. may play a role in increasing the volume of trade in ECOWAS countries, while transport cost reduce it.

Anahtar sözcükler: Gravite modeli, panel verisi yöntemleri, eşbütünleşme, ECOWAS, sıfır ticaret akımı, PPML yöntemi, Heckman seçim süreci

Önsöz, Teşekkür veya İthaf Sayfası (isteğe bağlı)

20/.08/20.19

STATEMENT OF COMPLIANCE WITH ETHICAL PRINCIPLES AND RULES

I hereby truthfully declare that this thesis is an original work prepared by me; that I have behaved in accordance with the scientific ethical principles and rules throughout the stages of preparation, data collection, analysis and presentation of my work; that I have cited the sources of all the data and information that could be obtained within the scope of this study, and included these sources in the references section; and that this study has been scanned for plagiarism with "scientific plagiarism detection program" used by Anadolu University, and that "it does not have any plagiarism" whatsoever. I also declare that, if a case contrary to my declaration is detected in my work at any time, I hereby express my consent to all the ethical and legal consequences that are involved.

AMADOU NDONGO THIAM

TABLE OF CONTENTS

TABLES

FIGURES

SYMBOLS AND ABBREVIATIONS

ADF:	Augmented Dickey-Fuller
BPML:	Binomial Pseudo Maximum Likelihood
CD:	Cross Dependence
ECOWAS:	Economic Community Of West African States
EU:	European Union
DF:	Dickey-Fuller
DOLS:	Dynamic Ordinary Least Square
FDI:	Foreign Direct Investment
FTA:	Free Trade Agreement
FMOLS:	Fully Modified Ordinary Least Square
GDP:	Gross Domestic Product
GLM:	General Linear Model
GPML:	Gamma Pseudo Maximum Likelihood
HO:	Heckscher-Ohlin
HOS:	Heckscher-Ohlin-Samuelson
IMR:	Inverse Mills Ratio
IMR:	Inward Multilateral Resistance
LM:	Lagrange Multiplication
LSDV:	Least Square Dummy Variable
NBPML:	Negative Binomial Pseudo Maximum Likelihood
NLS:	Non-Linear Least Square
OLS:	Ordinary Least Square
OMR:	Outward Multilateral Resistance
PPML:	Poisson Pseudo Maximum Likelihood
RE:	Random Effects

REIA: Regional Economic Integration Agreements

VIF: Variance Inflation Factor

INTRODUCTION

The gravity model is an applied econometric model that can be used to explore bilateral trade between two or more partner countries. This model has been highly successful with international trade analysts because of its great ability to explain trade. It is the most successful empirical instrument that has become unavoidable in the field of international trade. Its first application dates back to 1962 and was conducted by Tinbergen. After this initiation, there were other works like those of Pöyhönen (1963), Pullainnen (1963), and Linnemann (1966).

The gravity model is originally a model without theoretical foundations. Newton's law of universal gravitation inspired the gravity equation. According to this law, two-point bodies of respective masses M_I and M_J attract with forces of the same values, positively to the products of the two masses, and negatively to the square of the distance, D that separates them. In international trade, it will be said that the commercial flows depend positively on the Gross Domestic Product or the Gross National Product of each partner country and negatively on the distance that separates them.

Anderson (1979) was the first to bring a gravitational theoretical approach, based on monopolistic competition. Subsequently, there was the approach of Bergstrand (1989), which is based on Krugman-Helpman's[1] (1985) theory of international trade in imperfect competition with firm homogeneity. The microeconomic underpinnings of the gravitational equation were initiated by Deardorff (1998) who rejected Krugman and Helpman's thesis. The gravity model can not be theoretically grounded by applying the Heckscher-Ohlin[2] approach. After this approach, there was that of Eaton and Kortum (2002) largely inspired by Ricardo's[3] theory of the differences in production technology.

Of all these approaches, only Anderson (1979) and Deardorff (1998) reach a final conclusion that trade flows are negatively dependent on the economic distance between the two partner countries,

[1] The Krugman and Helpman model (1985) is an extension of Krugman's models (1979, 1980, 1981).
[2] It is a model that owes its name to two Swedish economists. Eli Heckscher was the first economist to introduce the notion of specialization in terms of factor endowment appeared in his article published in 1919 entitled 'The effect of Foreign Trade on distribution on income.' In 1933, his successor economist and political scientist Bertil Ohlin published a new article called 'Interregional and International Trade' and deduced the HO theorem. In 1941, this model becomes model HOS following the Stopler-Samuelson theorem.
[3] Ricardo's model appeared in 1847 in an article called 'Principles of Political Economy and Tax.' This is a response to critics of Adam Smith's absolute benefit theory. The model is also called Ricardian model, he had several critics whose model HOS is a solution to its limits.

in relation to global spending and distance between a country and its other trading partners in relation to international trade. The development of the theoretical foundations of the gravitational equation was seen in the early 2000s, with the appearance of the article of Anderson and Van Wincoop (2003). These two authors solved a problem that had grown in the empirical analysis of the gravity model. This problem also called puzzle was the impact of the border effect on the trade flows estimated by McCallum (1995) on the Canadian provinces and the United States that sharing a common border has no effect on the commercial exchanges. If this were the case, then the negative relationship between distance and trade flows no longer made sense. As a result, Anderson and Van Wincoop (2003) replicated McCallum's (1995) estimations and found that his estimations were biased. According to them, McCallum had ignored very important factors that were internal and external multilateral resistance; what Anderson (1979) and Deardorff (1998) named as the economic distance.

The theoretical approach of Anderson and Van Wincoop (2003) is no more than an extension of Anderson's (1979) theoretical model. According to them, the trade flows between two partner countries depend not only on the bilateral resistance approximated by the cost of bilateral trade but also on the multilateral resistance which represents the cost of multilateral trade of a country vis-à-vis to the whole of its trading partners.

Multilateral resistance has had measurement problems. It was often approximated by price indices; which was not practical because price indices are unobservable and need to be estimated. The best way to measure multilateral resistance is to approach it by the remoteness index easily obtained through distance, GDP and global GDP. Multilateral resistance is, therefore, an unavoidable variable in the gravitational specification of trade and holds very important properties.

The theoretical models we have cited are models that consider the homogeneity of firms [4]. Except for the theoretical approach of Eaton and Kortum (2002) which is part of the heterogeneity of firms. Indeed, the gravitational approach of Anderson and Van Wincoop (2003) has further expanded the theoretical foundations of Eaton and Kortum (2003). This theoretical model leads to the same specification of the gravity equation for Anderson and Van Wincoop. This allowed introducing new notions into the theoretical foundations of Eaton and Kortum (2002). This extension of their model allows to determine the supply and the demand of the commercial flows and thus to study the partial or general equilibrium starting from the model of gravity.

[4] The homogeneity of the firms implies a homogeneity of the goods produced by the firms. We are in a situation of pure and perfect competition. Unlike homogeneity, heterogeneity consists of producing a variety of economic goods within several firms.

The theoretical developments of the gravity model constantly bring new avenues of research to international trade analysts. The gravity model was for a long time estimated in its log-linear form; demand is considered homogeneous of degree zero. Thus, Novy (2010) estimated the gravity model in the framework of heterogeneity of firms by a trans-log[5] demand function and found much more robust results when it comes to a log-linearized demand function. This has also introduced the notions of intensive margins and extensive margins.

The applications of the gravity model are many and varied. The gravity model has become the most successful model in international economics. Its fields of application are quite extensive. Indeed, the gravitational equation can be used to determine bilateral trade between two partner countries, the effect of trade policy in regional integration agreements or a monetary union, foreign direct investment and migration flows. However, its fields of application are not limited to the international economy. The gravity model can also be used to measure financial asset flows, international banking flows, and bank credit stocks.

Research on the gravity model has been numerous theoretically as well as empirically with problems of specification and estimation. Indeed, the estimation of the gravity model can be done in its multiplicative or linear form, in cross-section or in panel data. However, the panel data estimate remains unparalleled because it gives the model the most satisfactory results.

However, we are often faced with econometric problems on an empirical level. First, the measurement of multilateral resistance is still a challenge in estimating a gravity model with fixed or random effects. Many empirical studies show that internal and external multilateral resistance can be considered as unilateral individual fixed effects or as a bilateral fixed effect. Therefore, we are faced with a problem of collinearity; this leads to the omission of collinear variables with unilateral individual fixed effects such as GDP and collinear variables with a bilateral fixed effect such as distance; so that the estimate is made in cross-section. The only solution to avoid this collinearity phenomenon is to estimate in cross-section the gravity model with random effect.

Panel data estimation offers a much broader view in the empirical analysis because it captures all the sources of individual, temporal, inter-individual and both individual and temporal heterogeneities of the chosen gravitational specification. Nevertheless, the estimation of such a panel

[5] It is a very flexible demand function because it does not impose any restrictions, including specific structural assumptions such as the Cobb-Douglas or CES demand. The Trans-log function can also be a second-order approximation, which is twice differentiable and can be considered of any technology (McFadden, Chambers, Mundlak ...).

data model is impossible because the number of observations is too high. For this purpose, many estimable specifications have been developed and take into account multilateral resistances by approximating them either by unilateral effects depending on time or not or by a bilateral effect. The estimation is done under fixed or random effects, where Wald, LM Breush Pagan, and Hausman tests will allow us to determine the choice of estimation results. The estimation method is done by the OLS if the specification chosen is a log-linear gravity model and by non-linear least-squares, if the specification is a multiplicative gravitational model.

We draw attention to the estimation method because estimating the gravity equation by least-squares is not always allowed, especially when it comes to the log-linearized specification. Indeed, estimating a log-linearized transformation in the presence of heteroscedasticity can lead to unsatisfactory estimation results. Several authors have worked on this question and have tried to determine the cause related to the problem of heteroscedasticity. We can cite Silva Santos and Tenreyro (2006) who justified it by the fact that the error term is a function of the gravity variables and the log-linearization of the gravity equation in the presence of heteroscedasticity implies inconsistent estimators. Because the expectation of the logarithm of the random variable is a function of the moments of its distributions. Moreover, Kalirajan (2008) shows that the presence of heteroskedasticity is linked to an omission of relevant variables, which is justified by the fact that trade costs are generally approximated by geographical distance. The estimate is then done by the GLS.

Another very common problem in the empirical analysis is the nature of the data. In a trade flow database, there are often zeros that represent zero trade flows between some countries, often related to the distance between them or to a lack of information on trade flows. Then the estimation of the linear gravity model with such data by the OLS is no longer possible because the logarithm is not defined at zero and any elimination of zero trade flows in the estimation program systematically leads to a bias of estimation linked to a loss of information. Several solutions have been made to deal with zero trade flow. The most widely used and commonly accepted solution is the estimation of generalized linear models whose estimation method is done by the Maximum Probability Pseudo from a law of the exponential family (Poisson, Gamma, Binomial, Binomial Negative, Normal and Normal inverse). The Poisson law remains the most practiced law thanks to the quality of its estimators. There are other effective ways to deal with zero trade flows. We can quote the Heckman selection procedure in two stages.

The choice of variables is very important for the estimation of the gravity equation. In addition to the gravity variables (GDP of the two partner countries, bilateral resistance and multilateral resistance), we can increase other variables such as the common border, the common language, the colonial past, regional integration agreements (Free Trade Zone, Customs Union, Common Market, Economic Union, Monetary Union) and so on, as indicator variables to avoid the problem of endogeneity, despite the fact that panel data models are solutions to this kind of hypothesis violation. Indeed, the endogeneity of the gravity variables is a problem that is not frequent in the estimation of the gravity equation. The endogeneity of an explanatory variable result in a non-zero correlation of the explanatory variable with the error term of the model to be estimated. A measurement error of the gravity variables, the omission of relevant variables in the model specification or the simultaneity of the gravity variables can cause the problem of endogeneity in the gravity equation.

The estimation of the gravity model is often done with panel data. So the temporal dimension is involved in the empirical estimates. However, in recent years, new gravitational econometric approaches have been developed: the estimation of the gravity model by cointegration techniques. Because it is above all a question of studying the stationarity of the gravity variables so that the problem of stationarity in the panel is less important than in time series. However, it is important to know that the estimation of a gravitational relation whose variables are cointegrated on a small period panel will not have a very important impact on the results of estimations. Because the commercial flows and the GDP evolve often rather slowly and a small period of a panel would not assess the effect of their non-stationarity on the estimates by standard econometric techniques.

This thesis is the subject of a study of the gravity model in case of homogeneity of firms and consists of three chapters. The first chapter is entitled **Theoretical Foundations of the Gravity Model Applied to International Trade**. In this chapter, we will examine the origin of the gravity model, the theoretical foundations and their developments as well as some fields of application of the gravitational equation.

The second chapter treats the **Specification and Estimation of the gravity model** and in which we will see the different specifications of the gravitational equation and the methods of estimation of the model according to the multilateral resistance, zeros trade flows, heteroscedasticity, endogeneity, stationarity, and cointegration.

Finally, the third chapter is an empirical analysis of the gravity model applied to international trade by different econometric approaches. The empirical study focuses on a panel of 15 ECOWAS countries. In this chapter, we will examine the different econometric approaches proposed in the theoretical part in order to compare the results and to bring a robust analysis of our empirical study.

CHAPTER ONE

1. THEORETICAL FOUNDATION OF THE GRAVITY MODEL APPLIED TO INTERNATIONAL TRADE

The gravity model is one of the most applied models in international economics to explore bilateral trade flows. Despite its empirical success with analysts in the international economy, the gravity model has remained almost a decade without theoretical foundations.

In its traditional form, the gravity model explains the bilateral trade between two partner countries from their different sizes, expressed by their gross domestic product (GDP) and the geographical distance that separates them. According to this equation, the volume of bilateral trade between two countries is proportional to their GDP and inversely proportional to the distance between them.

However, the intensification of bilateral trade and the emergence of different forms of regional integration will revolutionize the theoretical foundations of the gravitational equation with the introduction of several other factors such as border effects and multilateral resistance, see the approach of Anderson and Van Wincoop (2003).

In this chapter, we will examine, in a much broader framework, first the origin of the gravity equation and its extension, then the theoretical foundations of the gravitational model and finally its various fields of application.

1.1. Origin of the gravity model

Newton's law of universal gravitation inspired the gravity equation. According to this law, two-point bodies of respective masses M_I and M_J attract each other with forces of the same values. These forces depend positively on the products of the two masses, and negatively on the square of the distance D that separates them.

Formally, we have:

$$F_{IJ} = F_{JI} = G\frac{M_I M_J}{D^2} \qquad (1.1)$$

The gravitational equation applied to bilateral trade expresses the volume of bilateral trade between two countries i and j as being proportional to their respective gross domestic products and inversely proportional to the distance that separates them.

Note by X_{ij} the bilateral trade volume of country i to country j, Y_i and Y_j the respective sizes of the two partners measured by their GDP, D_{ij} the distance that separates them and A a constant. Formally, the basic gravitational equation is written:

$$X_{ij} = A\frac{Y_i Y_j}{{D_{ij}}^2} \qquad (1.2)$$

Tinbergen (1962) applied the gravity equation for the first time in the field of international trade. He supposed the following gravitational relation:

$$X_{ij} = A\frac{Y_i^{\alpha} Y_j^{\beta}}{D_{ij}^{\delta}} \qquad (1.3)$$

Positive exponents α, β, and δ can take values different from 1, and are respectively interpreted as the elasticity of exports of country i with respect to Y_i, the elasticity of exports of country j with respect to Y_j and export elasticity of country i in relation to the distance between country i and country j. The sizes Y_i and Y_j of countries, i and j can be measured by gross domestic product (GDP), gross national product (GNP) or per capita income.

Bilateral trade between two partner countries depends positively on the size of the partner countries and negatively on the geographical distance between them. The greater the distance between the two countries, the lower their volume of bilateral trade.

This specification of the gravity equation has been used in several other works such as Pöyhönen (1963), Pullainnen (1963), and Linnemann (1966).

The gravity model has remained more than a decade without theoretical foundations. The first theoretical approach was proposed by Anderson (1979), which will be followed by other, much more in-depth approaches through the development of international trade, resulting in the emergence of forms of regional integration and the intensification of trade.

1.2. Theoretical Foundation of the gravity model

The lack of rationality in the gravity equation has been theoretically justified from several neoclassical approaches to international trade. Indeed, the theoretical foundations of the gravity model have undergone several improvements following the development of the theory of international trade that is justified by an intensification of trade, the emergence of forms of regional integration and the development of means of transport. Among the different approaches to the fundamentals of the gravity model, there are neoclassical models of monopolistic competition based on Ricardo and Heckscher-Ohlin comparative advantage theory, imperfectly competitive models based on the homogeneity of Firms (the Krugman and Helpman model) and the heterogeneity of firms (the Melitz model).

According to Ricardo's theory of comparative advantages, all countries, even the least competitive, have an interest in participating in the international exchange. Each country specializes in the production of the good for which it has a relative productive advantage and purchases the good it produces with a disadvantage. The exchange between countries is Pareto-optimal because it allows countries to use their factor endowments more efficiently so to consume more at equilibrium. The Ricardian model is a single-factor model, unlike the Heckscher-Ohlin (HO) model that considers a two-country economy producing two goods from two factors of production.

In the HO model, the comparative advantage is explained by factor endowments. When the production of a good requires several productive factors, in different proportions according to the sectors, the countries specialize in the production of the goods which require factors of which they dispose of abundance.

The Ricardo and HO models explain the trade of different goods between countries with different production technology or factor endowments, while the new theories of international trade, initiated by Krugman (1979, 1980), explain the cross-trade of differentiated goods between similar countries in terms of production technology with increasing returns to scale, which makes the exchange a source of efficiency. Specialization is done on horizontally differentiated goods[6]. Krugman's model has predictions that are consistent with the gravity model. This model is based on assumptions of economies of scale, monopolistic competition, iso-elastic preferences, and proportional transport costs.

[6] Horizontal differentiation means a production of goods having a common market and whose use is the same or similar. It is also a differentiation of variety. Indeed, other forms of differentiation exist: vertical or quality differentiation and spatial differentiation.

The Krugman and Helpman model is an extension of the Krugman model which takes into account only the hypothesis of a single sector producing a single well-differentiated good, which did not allow the introduction of specialization phenomena. Indeed, the Krugman and Helpman model considers an economy with two productive sectors that highlights the exchange of differentiated varieties of the same good. Intra-industry trade may subsequently affect the cross-sectoral specialization of countries that can be enhanced by lower transport costs. The Krugman and Helpman model is a model with firms competition, unlike the Melitz model, which is based on imperfect competition[7] with heterogeneous firms.

1.2.1. Anderson’s model (1979)

Anderson's model is a monopolistic model based on product differentiation. Initially, Anderson (1979) considers an economy consisting of two countries i and j, each producing a commercial good. The characteristics of the partner countries are determined by identical Cobb-Douglas homothetic preferences. The balance of trade is in equilibrium and the goods are marketed between the two countries. The revenue share of country i from its exportation to country j is given by k_i. Then the volume of trade between country i and country j is expressed as follows:

$$X_{ij} = k_i Y_J \qquad (1.4)$$

It is assumed that there are no transport costs and any other barriers that could negatively affect the exchange between the two partner countries. If country i has several trading partners, its size is given by the constraint below which expresses the gross domestic product in proportion to the world gross domestic product:

$$Y_i = k_i \sum_j Y_j \qquad (1.5)$$

We can directly deduce the expression from the country i coming from exports to its various partners j:

$$k_i = \frac{Y_i}{\sum_j Y_j} \qquad (1.6)$$

[7] Perfect competition is a non-existent model in reality. Moreover, the concentration of the companies puts into question the assumption of atomicity in pure and perfect competition. Therefore, we will talk about imperfect competition. Monopoly is a form of imperfect competition.

Substituting the relation (1.6) in relation (1.4), we find the standard gravity equation, also called the traditional gravity equation:

$$X_{ij} = \frac{Y_i Y_j}{\sum_j Y_j} \qquad (1.7)$$

Anderson's initial model has very restrictive assumptions. Each of the two countries produces only one good. The introduction of a second product will be the subject of an extension of the Anderson model.

Anderson introduces into his initial model a utility function that depends on two types of goods: marketed goods and non-marketed goods. Thus, each country chooses a basket of marketed goods respecting their budget constraints in order to maximize their utility. In the relationship obtained, there are no distance and transport costs. The theoretical foundations of the gravity model proposed by Anderson (1979) are based on the assumption of the absence of barriers related to international trade. In the standard gravity model, distance is a very important variable in the gravity equation.

Thus, Anderson extended his model again by introducing the bilateral resistance factor as the cost of transport. The resultant gravity equation represents Anderson's last result (1979) and is written as two economic distances: the economic distance between country i and country j, in relation to total world expenditure on marketed goods and the economic distance between country i and all other trading partners in relation to international trade.

In summary, Anderson's model is the first model to have given a theoretical approach to the gravity equation applied to international trade. Anderson (1979) assumes an economy consisting of two identical Cobb-Douglas homothetic preference countries, each producing a single good. In addition, other specialists have introduced other variables. It is in the image of leamer (1974) who introduced the endowment factors and Bergstrand (1985) included the price variable and Eaton and Kortum (2002) who proposed a theoretical approach to the gravity model based on the difference of technology in the regard of Ricardo.

1.2.2. Eaton and Kortum's model (2002)

The study of the compatibility of the Ricardian theory of international trade (based on the difference in production technology) with the gravity model was conducted by Eaton and Kortum (2002). In this model, we consider two countries i and j, each produce a good. Work is the only factor of production used and is mobile within countries and still between countries. Eaton and Kortum

(2002) introduce into their model geographic barriers according to Samuelson's standard and practical 'iceberg' hypothesis, which states that the delivery of a unit of goods from country i to country j requires the production of d_{ij} units in the country i. We set $d_{ii} = 1$ and the geographical barriers are positive $d_{ij} > 1$ for $i \neq j$. We also assumed that cross-border arbitration forces of effective geographic barriers verify triangular inequality: for three countries i, j, and k, we have: $d_{ij} \leq d_{ik}d_{kj}$.

According to these hypotheses on the geographical barriers, the cost of production of a unit of good s is equal to $\frac{w_i}{y_i(s)}$ and the price of a unit this good is worth $P_{ij}(s) = \frac{w_i d_{ij}}{y_i(s)}$. w_i is the cost of the labor factor or price of the input and $y_i(s)$ is quantity produced of good in country i, with $s \in [0; 1]$.

Eaton and Kortum (2002) assume that competition is perfect and that the market price is determined by:

$$P_j(s) = min\{P_{ij}(s); i = 1, \dots, N\} \qquad (1.8)$$

Consumers in the importing country j choose a basket of goods on the basis of the above price system and their budget constraint Y_j in order to maximize their utility given by a Constant Elasticity of Substitution(CES) type preference function:

$$U^j = \left[\int_0^1 x(s)^{(\sigma-1)/\sigma} d_s\right]^{\sigma/(\sigma-1)} \qquad (1.9)$$

$\sigma > O$ is the elasticity of substitution and $x(s)$ is the consumer demand for good s of country j.

Producers maximize their profits under their production technology constraint $F_i(y)$. These authors consider that the production of good s of country i is a random realization of the produced quantity variable y_i. So the production technology is represented by Fréchet's [8] probability distribution :

$$F_i(y) = Pr[y_i \leq y] = e^{-T_i y^{-\theta}} \qquad (1.10)$$

In equation (1.10), $T_i > 0$ determines the location of the country-specific distribution and designates the state of their technology, and $\theta > 0$ is the parameter that regulates the heterogeneity of goods produced with respect to technical efficiency and determines the comparative advantage in this context.

[8] The law of Fréchet is a particular law in Probability theory and Statistics. This law is that of Maurice Fréchet, appeared in his article of 1927. The law of Fréchet is applicable in several domains like the economy, the trade, the hydrology ...

Prices depend on the random level of production. Then the probability distribution associated with the price $P_{ij}(s)$ is obtained as follows:

$$P_{ij}(s) = \frac{w_i d_{ij}}{y_i(s)} \qquad \text{then} \qquad y_i(s) = \frac{w_i d_{ij}}{P_{ij}(s)} \tag{1.11}$$

We replace the expression (1.11) in (1.10) and we obtain the following probability distribution:

$$F_i(y) = F_i\left(\frac{w_i d_{ij}}{P_{ij}}\right) = Pr(y_i \leq y) = Pr\left(\frac{w_i d_{ij}}{P_{ij}} \leq y\right) = Pr\left(P_{ij} \geq \frac{w_i d_{ij}}{y}\right) \tag{1.12}$$

Let $G_{ij}(P) = Pr\left(P_{ij} \leq P\right)$, the probability distribution of prices. This distribution is a function of the probability distribution of the quantity produced:

$$G_{ij}(P) = 1 - F_i(y) = 1 - e^{-\left[T_i\left(w_i d_{ij}\right)^{-\theta}\right]P^{-\theta}} \tag{1.13}$$

The probability distribution of prices in country j is determined from relation (1.13):

$$G_j(P) = 1 - \prod_i G_{ij}(P) = 1 - e^{-\varphi_j P^{\theta}} \text{ , avec } \varphi_j = \sum_i T_i\left(w_i d_{ij}\right)^{-\theta} \tag{1.14}$$

The parameter φ_j covers several factors. This parameter of the price distribution makes it possible to know the state of the technology in the world, the cost of the work factor in the world and the geographical barriers. Eaton and Kortum thus exploit three properties of the probability distribution of prices that allowed them to obtain a gravity equation in this form:

$$X_{ij} = \frac{\left(\frac{d_{ij}}{P_j}\right)^{-\theta} Y_j}{\sum_j \left(\frac{d_{ij}}{P_j}\right)^{-\theta} Y_j} Y_i \tag{1.15}$$

Relationship (1.15) is a standard gravitational relationship in which exports are expressed in terms of the income of the exporting country, geographical barriers between country i and its partners which are deflated by the price of importing countries.

In summary, the theoretical foundations of the gravity model have been developed on the basis of several neoclassical theories of international trade that sometimes lead to different or somewhat similar results in the work of Anderson (1979) and Deardorff (1998). Which equation of gravity involves a comparison of bilateral trade between two countries and the adjusted trade weight of other partner countries. The estimation of the gravity model by these authors is often done in logarithmic form. However, with the intensification of trade between countries, several economists have introduced into the gravity equation other qualitative variables such as the common language, the colonial link, the colonial past, and the common border.

Indeed, the existence of a common border has been a highly debated issue that has gained momentum in the literature of the gravitational equation. This issue of the border effect was studied for the first time by McCallum[9] (1995) and the solution was brought by Anderson and Van Wincoop (2003). This is what has revolutionized the theoretical foundations of the gravity model with the introduction of a very important variable called multilateral resistance.

1.3. Recent Development of the Theoretical Foundation

The response of Anderson and Van Wincoop (2003) on the border effect has led to considerable development on the theoretical foundations of the gravity equation. The border effect means that sharing a common border leads to a significant reduction in trade. In other words, the border effect exists when two countries are trading more with each other than two countries separated by a single border.

According to McCallum (1995), national borders do not matter in international trade. According to his empirical study of the Canadian provinces and the United States, it shows that a Canadian province traded more with the other Canadian provinces. On average 20 times more than with an American state of size and distance more or less equal.

This result by McCallum (1995) was not logical according to some analysts like Anderson and Van Wincoop (2003) because the United States and Canada share more than one border. It is also a sharing of common language and a very close culture. According to these economists, the results of McCallum (1995) are biased. He made his estimates without taking into account other variables that may further explain bilateral trade between Canada and the United States. Moreover, the resumption of his estimates with a more adapted theory gave lower results of the border effect. For example, Anderson and Van Wincoop (2003) solved the problem of the border effect by introducing into their model a new variable called the multilateral resistance factor that neutralized all the effects from trading partners.

[9] It is an article that has revolutionized the theories of the gravity model. For further details see McCallum (1995) National Borders Matter: Canada-USA, Regional Trade Petters, American Economic Review, (3) 613-623.

1.3.1. Anderson and Van Wincoop's model (2003)

The Anderson and Van Wincoop model (2003) is an extension of Anderson's model (1979) whose last theoretical result was a gravity equation based on an economic distance between country i and country j and the average weight of The economic distance of country i from other trading partners. Thus, they define a variable that is very relevant to trade barriers. As a result, they have been interested in comparing the equilibrium within borders and equilibrium in the frontier world, and incorporate in their model multilateral resistance that measures the cost of multilateral trade. This variable neutralizes the border effect and determines the choice of a country vis-à-vis its trading partners. Intuitively, when the bilateral resistance factor between two countries i and j increases relative to the multilateral resistance factor, then trade decreases between countries i and j.

Formally, Anderson and Van Wincoop (2003) assume an economy with two countries. The exporting country produces a good exported at the price p_{ij} and consumed locally at the price p_i. The importing country j chooses a basket of goods that respects its budget constraint in order to maximize its utility. On the other hand, it is assumed that countries have identical and homothetic preferences. Indeed identical preferences imply that all importing countries have the same demand for consumption and spend the same part of their income on the good produced by the exporting country. Homothetic preferences mean that the proportion in the consumption demand of the good depends not on the income but on the relative price of that good. In addition, it is assumed that the preferences are of CES type and depend on the consumption x_{ij} and the elasticity of substitution $\sigma > 0$. The utility maximization program is as follows:

$$(x_{ij})\Big/ \qquad Max \quad U^j = \sum_i \left(x_{ij}{}^{(\sigma-1)/\sigma} \right)^{\sigma/(\sigma-1)} \qquad (1.16)$$

$$st: \quad \sum_i t_{ij} p_i x_{ij} = Y_j$$

The budget constraint is equal to the total expenditure of country j in the consumption of the exported good. In this constraint the consumer price in the country i is p_i and the cost of bilateral trade between country $i \; and \; j$ is given by $t_{ij} > 1$.

The solution of this maximization program is done by the Lagrangian:

$$\mathcal{L}(x_{ij}, \lambda) = U^j - \lambda(Y_j - \sum_i t_{ij} p_i x_{ij}) \qquad (1.17)$$

$$\Delta\mathcal{L} = 0 \ \ thus \begin{cases} \frac{\partial\mathcal{L}}{\partial x_{ij}} = 0 & (i) \\ \frac{\partial\mathcal{L}}{\partial\lambda} = 0 & (ii) \end{cases} \tag{1.18}$$

the relationship (i) gives us the following equation:

$$\lambda t_{ij} p_i = \left(\sum_i x_{ij}^{(\sigma-1)/\sigma}\right)^{1/(\sigma-1)} x_{ij}^{-1/\sigma} \tag{1.19}$$

By multiplying the above relationship by x_{ij} and by summing on i we arrive at the following equality:

$$\lambda Y_j = \left(\sum_i x_{ij}^{(\sigma-1)/\sigma}\right)^{1/(\sigma-1)} \sum_i x_{ij}^{(\sigma-1)/\sigma} \tag{1.20}$$

The relation (1.19) gives us the expression of the multiplier Lagrange that we will substitute in (1.20) and we thus obtain the export price of the goods:

$$p_{ij} = t_{ij} p_i = \frac{x_{ij}^{-1/\sigma} Y_j}{\sum_i x_{ij}^{(\sigma-1)/\sigma}} \tag{1.21}$$

It is assumed that bilateral trade costs are paid and the nominal value of exports from country i to country j is:

$$X_{ij} = p_{ij} x_{ij} \tag{1.22}$$

Thus, the total income of the exporting country i is equal to the sum of the nominal values of its exports on all its trading partners' j:

$$Y_i = \sum_j X_{ij} \tag{1.23}$$

Anderson and Van Wincoop (2003) define country price index j as being equal to:

$$P_j = \left(\sum_i (p_i t_{ij})^{(1-\sigma)}\right)^{1/(1-\sigma)} \tag{1.24}$$

They get from the results of the maximization program the import demand of the country j which is equal to:

$$X_{ij} = \frac{(t_{ij} p_i)^{1-\sigma} Y_j}{P_j^{1-\sigma}} \tag{1.25}$$

From the relation (1.25) with some mathematical manipulations on the basis of relations (1.23) and (1.24), we obtain the last result of the gravitational equation proposed by Anderson and Van Wincoop (2003):

$$X_{ij} = \frac{Y_i Y_j}{Y^W}\left(\frac{t_{ij}}{P_i P_j}\right)^{(1-\sigma)} \quad avec \begin{cases} P_i^{(1-\sigma)} = \sum_j \left(\frac{t_{ij}}{P_j}\right)^{(1-\sigma)} \frac{Y_j}{Y^W} \\ P_j^{(1-\sigma)} = \sum_i \left(\frac{t_{ij}}{P_i}\right)^{(1-\sigma)} \frac{Y_i}{Y^W} \end{cases} \quad (1.26)$$

Where $Y^W = \sum_j Y_j$ is the world income.

This gravity equation obtained by Anderson and Van Wincoop (2003) expresses the volume of trade between countries i and j according to three resistance factors:

- ✓ The bilateral resistance between countries i and j given by t_{ij} which represents the cost of bilateral trade between the two partner countries and which is approximated in the literature by the geographical variety and the commercial policy variables like the bilateral distance, the tariffs or the Existence of a form of regional integration;
- ✓ The outward multilateral resistance P_i of country i in relation to all its other trading partners, which indicates its choice of markets for its goods for export;
- ✓ The inward multilateral resistance P_j of the country j in relation to all its other trading partners, which designates its choice in foreign markets.

This specification of the gravity equation can be obtained from several gravitational theories. The model of Eaton and Kortum (2002) offers a much broader dimension in terms of supply and demand for a good. You can refer to Anderson's article (2011), which outlines the gravitational foundations of supply and demand. The equations of gravity with terms of multilateral resistance obtained from the model of Eaton and Kortum (2002) are isomorphic and simply differ by the parameters $1 - \sigma \; and \; \theta$.

$$Supply\ of\ goods: X_{ij} = \frac{Y_i Y_j}{Y^W}\left(\frac{t_{ij}}{P_i P_j}\right)^{-\theta} \quad with \begin{cases} P_i^{-\theta} = \sum_j \left(\frac{t_{ij}}{P_j}\right)^{-\theta} \frac{Y_i}{Y^W} \\ P_j^{-\theta} = \sum_i \left(\frac{t_{ij}}{P_i}\right)^{-\theta} \frac{Y_j}{Y^W} \end{cases} \quad (1.27)$$

$$Demand\ of\ goods: X_{ij} = \frac{Y_i Y_j}{Y^W}\left(\frac{t_{ij}}{P_i P_j}\right)^{1-\sigma} with \begin{cases} P_i^{1-\sigma} = \sum_j \left(\frac{t_{ij}}{P_j}\right)^{1-\sigma} \frac{Y_i}{Y^W} \\ P_j^{1-\sigma} = \sum_i \left(\frac{t_{ij}}{P_i}\right)^{1-\sigma} \frac{Y_i}{Y^W} \end{cases} \quad (1.28)$$

1.3.1.1. Measuring and interpreting multilateral resistance

The gravitational equation of Anderson and Van Wincoop (2003) is a simplification of Anderson's gravity equations (1979), which made a comparison of the economic distances called resistances in the Anderson and Van Wincoop model (2003). Indeed, the multilateral resistances are given in this model by the price index and depend on the bilateral resistance between the two partner countries. An increase in trade barriers or bilateral resistance leads to an increase in the price index, and therefore rises multilateral resistances. For two given countries i and j, an increase in multilateral resistance between the country j and its other trading partners implies a decline in the relative price of the good produced by country i and increases the imports of country j from country i. In other words, when the bilateral resistance between countries i and j decreases in relation to their multilateral resistance to all their partners, trade between the two countries will increase. This explains the choice of a country in trade. The implication of the Anderson and Van Wincoop model (2003) in the theory of the gravity equation is that trade between two partners is determined by relative trade barriers. Trade between the two countries depends on the bilateral trade barrier relative to the average trade barriers of these countries with their other trading partners. Indeed, there are three implications for the impact of trade barriers on trade:

- Implication 1: refers to international trade between the two countries.

Trade barriers are decreasing trade between large countries in trade between small countries. That is, multilateral resistance among small countries increases more than multilateral resistance among large countries. This is justified by the fact that small countries are much more dependent on international trade;

- Implication 2: Refers to intra-regional trade.

Trade barriers increase trade in a small economy relative to trade in a large economy. This is due to an increase in the multilateral resistance that small economies face in international trade;

- Implication 3: Analyzes the ratio of international trade to intra-regional trade.

Trade barriers increase the trade ratio in a country i / trade between a country i and a country j if and only if the country i is a small country and the country j is a large country.

The price index is used to measure multilateral resistances but these proxies are unobservable and must be estimated. For example, Anderson and Van Wincoop (2003) used observed variables such as absolute distance and income shares of importing countries as an alternative solution for measuring multilateral resistance. Indeed these resistances have been approximated by the variable remoteness[10] R_i, which is a relative distance equal to the average of the weighted distances:

$$R_i = \sum_j \frac{D_{ij}}{\left(\frac{Y_j}{Y^W}\right)} \qquad (1.29)$$

However, the calculation of this variable often remains a difficulty because it depends on the bilateral distance between the two partner countries which poses difficulties on the empirical analysis. In Chapter 2 we will see the problems of multilateral resistance in estimating the gravity equation and the solutions that have been applied to measure these terms of resistance.

1.3.1.2. Properties of multilateral resistance

Multilateral resistance is one of the variables that has revolutionized gravitational theory. Indeed, the model of Anderson and Van Wincoop (2003) offers an ideal framework of analysis of the balances in the international trade thanks to the multilateral resistances. In a context of general equilibrium, multilateral resistances have the following properties:

P1: *Multilateral resistance is intuitively approximated by the costs of bilateral and multilateral trade.*

The intuition in the interpretation of multilateral resistance is that when two countries market more it is because they remain more distant to the rest of the world. In other words, two countries trade according to a bilateral trade cost related to the costs of multilateral trade.

P2: *Multilateral resistances are compatible aggregates with the gravitational theory of bilateral trade costs.*

This property is verified in the gravity equation of Anderson and Van Wincoop (2002) and in all other gravitational theories like that of Eaton and Kortum (2002). Multilateral resistances are compatible with all possible bilateral trade cost aggregates.

P3: *Multilateral resistances are very effective for practical purposes.*

P4: *Multilateral resistances are easy to calculate.*

[10] It is an Anglo-Saxon term which denotes the distance of a country vis-à-vis all of its trading partners.

P5: *The equivalence of price indices and multilateral resistances are verified at the general equilibrium.*

Since the development of international trade is justified by inter-country interaction, it is important to study the change in trade when trade barriers and prices change. Multilateral resistance is at the heart of the general equilibrium. These variables determine the behavior of partner countries as a result of changes in bilateral trade costs or a form of regional integration. Multilateral resistance also makes it possible to examine the effects of international trade on the other partner countries.

P6: *A decomposition of the incidence of trade cost can be obtained from the multilateral resistances.*

Anderson (2011) describes multilateral resistance as a measure of the average incidence of trade cost of the importing country (buyer) and the exporting country (seller). The external multilateral resistance $P_i^{1-\sigma}$ *or* $\Pi_i^{1-\sigma}$ is defined as an increase of a shipment that the exporting country uses on the world market. The internal multilateral resistance $P_j^{1-\sigma}$ *or* $\Pi_j^{1-\sigma}$ is defined as the incidence of the cost of trade: this is the cost that the importing country needs to pay to acquire goods on the world market.

Anderson and Yotov (2008) found that the Outward (or external) multilateral resistance (OMR) or the incidence of foreign trade cost is 5 times higher than the Internal (or domestic) multilateral resistance (IMR). According to these authors, there is no specific industry that faces a high or low multilateral resistance because the incidence of cost of trade varies according to the classes of goods. They have developed an index named after the Constructed Home Bias(CHB) that determines the predicted internal trade under border barriers in relation to internal trade with global friction. This bias is given the variable CHB:

$$CHB_i = \left(\frac{t_{ii}}{\Pi_i P_i}\right)^{1-\sigma} \qquad (1.30)$$

It is a determinant of intra-regional trade that is an extension of the gravitational equation of Anderson and Van Wincoop (2004). A comparison of intra-regional trade costs is inaccurate when regions are distinguished by different resistances and often trade barriers vary if the regional trade costs are equal $t_{ii} = t_{jj}$ and CHB_i are not necessarily identical because $\Pi_i P_i \neq \Pi_j P_j$.

1.3.2. Distance measurement

Distance is a very important variable in the gravity equation and negatively influences bilateral trade between two partner countries. Indeed, the estimation of the gravity equation poses enormous difficulties related to distance. Some authors have approximated the bilateral distance by the cost of trade which represents a bilateral resistance.

Distance is a variable that is difficult to measure. It has been the subject of several studies in the analysis of trade between different countries or groups of countries. In reality, distance is supposed to capture all information related to the trade barriers between trading partners. That is, it must inform us about the quantitative characteristics of barriers such as trade costs, tariffs and exchange rates, and qualitative cultural, institutional, and colonial characteristics.

There are different measures of distance according to Xubei (2001) in his article that studies the problem of distance in the gravitational equation applied to trade between the Chinese provinces and Japan. The contribution of Xubei (2001) is the determination of the actual adjusted distance and its approximation by the transport costs. These measures include:

- ✓ *Bird fly distance :*

It is a measure of the distance obtained from the arc distance between economic centers (often capitals) of two countries. This measure has very particular advantages over the fact that the distance is easily determined with the condition of all things being equal. Indeed, this condition neglects other factors that may influence bilateral trade. Because if we assume three countries i, j and k that have identical distances to the birds, we find that the transport costs resulting from trade between these countries are different. This is the reason why Hummels (1999) substituted the distance as the crow flies by the actual distance which yields much more information on the other factors determining the cost of a trade. Some authors have associated these factors with the quality of the infrastructures used for the transport of marketed goods: Limas and Venables (2000), Obstfeld and Rogoff (2000) are some examples.

- ✓ *The real distance :*

It is a measure that applies according to the route taken to transport the goods marketed. There are several transport routes: the land route via the road or the railway, the seaway, and the airway. These actual distances are also correlated with travel time. The longer the journey time, the higher the actual distance. For a similar real distance large enough, the trade associated with the actual distance is less expensive than that associated with the actual terrestrial distance.

- ✓ *Adjusted real distance :*

The actual distance adjusted a measure of the distance that takes into account the quality of the infrastructures or the ease of the transports, of more than the real maritime distance or the real terrestrial distance. Xubei (2001) in his article proceeds by constructing transport network density indices as an approximation of the quality of infrastructure to study trade between Chinese provinces and Japan. According to Xubei (2001), this measure remains a good approximation of the distance and constitutes a determining variable of the transport costs. However, this measure also opens up other research perspectives on the measurement of distance.

1.4. Application Fields of the Gravity Model

The gravity model has become the most successful model in international economics. Its fields of application are quite wide. Indeed, the gravitational equation can be used to determine bilateral trade between two partner countries, the effect of trade policy in regional integration agreements or in a monetary union, foreign direct investments flow[11], and migratory flows [12]. However, its fields of application are not limited to the international economy. The gravitational model can be used to evaluate financial asset flows, international bank flows, and bank credit stocks.

1.4.1. Economic integration agreement

By definition, an economic integration agreement is an agreement between and among several countries to promote trade and thereby facilitate the movement of goods, services, and people. According to Bela Balassa (1961), five forms of economic integration exist:

- *Free trade area:* It is a merchandise exchange area in a liberal way. That is, tariffs are zero between member countries and there is no quota. In addition, each country has its own trade policy vis-à-vis to other non-member countries of the Free Trade Area;
- *Custom union:* Trade policy is common to all member countries in a free trade area. In other words, trade policy implies the establishment of a common external tariff compared to other non-member countries;

[11] Several studies are based on the gravity equation to study FDI flows. In the various economic specifications, a delayed variable of FDI flows is added in addition to gravity variables in international trade.
[12] The study of migration flows implies an introduction of the number of populations of the country of origin and of the host country, the GDP of each country to assess the economic attractiveness of the host country and the distance that separates them. These are the fundamental gravity variables.

- *Common market:* It is a custom union characterized by openness in all markets, particularly the labor and capital markets. In this form of economic integration, there is free movement of natural persons and capital;
- *Economic union:* The economic union is a common market with the harmonization of economic policies. According to Delors (1989), it is characterized by a common market, a competition policy, common policies and the coordination of macroeconomic policies.
- *Economic and Monetary Union:* This is the deepest form of economic integration. Economic and Monetary Union is an economic union in which member countries have a single currency managed by a single institution which is the central bank.

The application of the gravitational equation on economic integration agreements has been the subject of several theoretical and empirical studies. Indeed, the most dominant trade agreement is that of the World Trade Organization called before GATT[13], which has 164 member countries with 21 observer countries. Studies have been conducted in this context to evaluate the effect of multilateral trade agreements on international trade flows based on the gravity model. Rose (2004) made an empirical study of 175 countries over a 50-year period by considering a qualitative variable that determines whether a country is a member of the World Trade Organization. In addition, a variable reflecting the most-favored-nation clause in which a member country reduces its tariffs vis-à-vis a country i member of the organization is added to the standard gravity model. It must apply this rate of reduction immediately and unconditionally to all member countries. However, this clause can be violated if both countries are members of regional integration. According to the empirical results, being a member does not have a significant effect on trade flows, but the variable reflecting the most-favored-nation clause has a positive and significant effect.

Other authors such as Carrer (2006) and Ries (2010) have applied the gravity equation to determine the effect of regional integration agreements. For example, Ries (2010) studied the effect of colonization on international trade flows between formerly colonizing and formerly colonized countries.

Rose and Van Wincoop (2001) have empirically estimated the effectiveness of a monetary union on trade flows. According to their results, the creation of a monetary union often increases the trade between the member countries of a small fraction and diminishes the multilateral resistance and the commercial barriers.

[13] GATT (General Agreement on Tariffs and Trade) is a treaty signed in 1947 in Geneva by 23 countries and is the former name of the WTO (World Trade Organization). This international organization counted in 2016, 164 members.

In summary, the application of the gravitational equation to economic integration agreements implies an introduction of a qualitative variable that is meant to take into account these forms of integration in the specification of the gravity model.

1.4.2. Foreign direct investments

Globalization has led to a significant increase in FDI flows around the world. By definition, foreign direct investment is an international direct investment which is characterized by a long-term relationship and a significant level of influence on the management of the firm in the host country. FDI is a mode of investment that a firm undertakes when it wants to expand internationally in order to gain market share. Indeed, there are two types of FDI:

- Greenfield investment which involves the creation of a new subsidiary;
- Brownfield investment which involves partial or total ownership of an existing firm by acquisition or merger.

FDI is seen as a transfer of technology, idea, and knowledge. The specification of the severity equation applied to FDI is identical to that applied to trade. FDI flows positively depend on the gross domestic products of the host country and the country of origin and negatively on the distance between them. However, the relationship between trade and FDI has been the subject of several studies. This relationship depends on the nature of the investment that can be vertical or horizontal.

The theoretical foundations of the gravitational equation have been associated with different theories of international trade. The models that most revolutionized these theoretical foundations are those of Eaton and Kortum (2002) and Anderson and Van Wincoop (2003) with the introduction of multilateral resistance and the determination of the level of supply of exporting countries in case of heterogeneity in the production.

Thus all these models departing from different hypotheses and theories lead to a similar specification of the gravitational equation. That is to say, bilateral trade between two partner countries depends in addition to the gravity variables on two types of resistance: the bilateral resistance approximated by the bilateral trade costs and the external and internal multilateral resistances approximated by the remoteness index.

However, the determination of multilateral resistance poses enormous difficulties in estimating the gravity model. Moreover, the remoteness of some countries compared to others has led to an absence of trade between these countries. Therefore, the specification and the estimation of the equation are supposed to take into account all these effects in order to have very satisfactory and unbiased empirical results.

CHAPTER TWO

2. SPECIFICATION AND ESTIMATION OF THE GRAVITY MODEL

In the specification of the gravity equation applied to international trade, trade flows are functions of gravity variables and multilateral resistance factors. Indeed, the specification of the gravity equation can be multiplicative or log-linear. Given the disadvantages of the method of estimating the multiplicative form, traditionally, the gravity model is estimated in its log-linear cross-sectional form by the OLS assuming that the errors are homoscedastic.

Estimation of the gravitational equation requires very thorough techniques to avoid all kinds of bias in the empirical analysis. In the specification of the gravity equation, the distance is approximated by the time-dependent transport costs. Then the estimation in panel data is more than necessary and gives a new breath in the empirical analysis because the panel data make it possible to capture all the sources of heterogeneity in the gravitational equation thanks to its temporal and individual dimensions. The panel gravity model can be a fixed-effects model or a random-effects model, the choice of which is made by a test of the existence of effects.

In addition, the log-linearization of the gravity equation raises some technical difficulties. Thus, a log-linear gravitational model causes a source of heterogeneity of the variance of the error terms, so there is a presence of heteroscedasticity whose estimation is done by the Generalized Least Square (GLS) if we know its form. However, the distance from one country to another usually leads to zero trade between these countries. Therefore, the log-linear specification is out of date for the estimation of the gravitational equation because the logarithm of zero is indeterminate. Several studies have eliminated commercial zeros and estimated the gravity model by OLS, but the resulting estimators are biased. So, we need more efficient techniques for dealing with zeros trading flows.

The gravity equation can be used as part of a study of the trade of economic integration and the estimate can, therefore, be biased. This bias is especially caused by a problem of endogeneity linked to an omitted variable, simultaneity in the context of the Foreign Trade Agreements (FTA), or self-selection in the context of a monetary union.

In this second chapter, we will first determine the different specifications of the gravitational equation and its estimation problems and in a second step, we will examine the sources of endogeneity of the variables used in the gravity equation in the framework economic integration.

2.1. Different Form of Specification

The gravity model is originally a multiplicative model whose estimation method is that of non-linear least squares. There is also another specification in logarithm in order to linearize the gravity equation. However, the interpretations of the parameters differ according to the specification used.

2.1.1. The multiplicative form

The gravitational equation applied to international trade is a model that explores bilateral trade flows between different partner countries, whether or not members of an economic integration. The standard gravity equation applied to international trade is written in the following form:

$$X_{ij} = \beta_0 \frac{Y_i^{\beta_1} Y_j^{\beta_2}}{D_{ij}^{\gamma}} \qquad (2.1)$$

This equation is given by the following multiplicative form with $\beta_3 = -\gamma$:

$$X_{ij} = \beta_0 Y_i^{\beta_1} Y_j^{\beta_2} D_{ij}^{\beta_3} \mathcal{E}_{ij} \qquad (2.2)$$

The volume of bilateral trade is given by X_{ij}. The variables $Y_i\ et\ Y_j$ represent the gross domestic products of the partner countries. $\beta_k, k = 1,2,3$ is a parameter the multiplicative model and $\mathcal{E}_{ij}$ the error term.

This specification was used by Tinbergen (1962) and Pöyhönen (1963) and was augmented by other variables such as the population of the partner countries in the studies by Linnemann (1966) and Leamer and Stern (1970):

$$X_{ij} = \beta_0 Y_i^{\beta_1} Y_j^{\beta_2} N_i^{\beta_3} N_j^{\beta_4} D_{ij}^{\beta_5} \mathcal{E}_{ij} \qquad (2.3)$$

Population sizes are given in this multiplicative equation by $N_i\ and\ N_j$. This multiplicative form of the gravitational equation is atheoretical.

The multiplicative gravitational equation augmented by multilateral resistances is as follows:

$$X_{ij} = \beta_0 Y_i^{\beta_1} Y_j^{\beta_2} D_{ij}^{\beta_3} e^{\beta_4 P_i + \beta_5 P_j + \beta_6 b_{ij}} \qquad (2.4)$$

Multilateral resistances are equivalent to fixed specific effects. The distance is approximated to the cost of transportation. From this equation, we can increase qualitative and quantitative variables to neutralize for example the effect of a common border and/or the effect of economic integration. This specification is less used today in the empirical analysis of the gravitational equation. The linear specification is more practical in the empirical analysis but it often poses technical difficulties related to its estimation.

2.1.2. The logarithmic form

The logarithmic form of the gravity equation is obtained from its multiplicative form. The logarithmic gravitational equation is the linear gravity equation and is given by the following relationship obtained by applying the log operator to the equation (2.2) :

$$\log(X_{ij}) = \beta_o + \beta_1 \log(Y_i) + \beta_2 \log(Y_j) + \beta_3 \log(D_{ij}) + \eta_{ij} \quad (2.5)$$

This relation results from the standard or traditional gravity equation and the distance is approximated to the cost of transport. The linear specification of the augmented gravity model of multilateral resistances is obtained from the relation (2.4) and is written in the following form:

$$\log(X_{ij}) = \beta_0 + \beta_1 \log(Y_i) + \beta_2 \log(Y_j) + \beta_3 \log(t_{ij}) + \beta_4 \log(P_i) + \beta_5 \log(P_j) + \eta_{ij} \quad (2.6)$$

Since the variable is a function of distance and a binary variable such as:

$$t_{ij} = b_{ij} D_{ij} \qquad \begin{cases} b_{ij} = 1 & if\ i\ and\ j\ \in to\ the\ same\ country \\ b_{ij} = 1 + c_{ij}\ or\ b_{ij} = e^{\sum_k \theta_k I_{kij}} & if\ i\ and\ j\ \notin to\ the\ same\ country \end{cases}$$

I_{kij} is an indicator variable and c_{ij} is the price of transportation of the goods. By replacing the bilateral resistance by its expression we obtain the following linear gravity equation:

$$\log(X_{ij}) = \beta_0 + \beta_1 \log(Y_i) + \beta_2 \log(Y_j) + \beta_3 \log(D_{ij}) + \beta_4 \log(b_{ij}) + \beta_5 \log(P_i) + \beta_6 \log(P_j) + \eta_{ij}$$

Note by STR_{ij}, the sum of the terms of resistance and is equal to the following relation:

$$STR_{ij} = \beta_4 \log(b_{ij}) + \beta_5 \log(P_i) + \beta_6 \log(P_j) \quad (2.7)$$

The linear equation becomes in this form:

$$\log(X_{ij}) = \beta_0 + \beta_1 \log(Y_i) + \beta_2 \log(Y_j) + \beta_3 \log(D_{ij}) + STR_{ij} + \eta_{ij} \quad (2.8)$$

This linear model is derived from the gravitational equation of Anderson and Van Wincoop (2003) $(\beta_1 = \beta_2 = 1\ et\ \beta_3 = \beta_4 = \beta_5 = 1 - \sigma\ , \sigma\ is\ the\ elasticity\ of\ substitution)$. The parameters of the linear gravity model are interpreted as elasticities. The variable STR_{ij} has been replaced by a qualitative variable in McCallum's (1995) estimates. That was the reason he found biased results. The work of McCallum (1995) did not take into account multilateral resistance, hence the problem of the border effect solved by Anderson and Van Wincoop (2003). However, the indices of multilateral resistance pose enormous difficulties in estimating the linear form. This problem is absolutely related to the measurement of multilateral resistance. Indeed, we have several solutions to overcome these technical difficulties. One of its solutions has been presented in Chapter 1. However, the most accepted solution in the empirical analysis of the gravity equation is to replace multilateral resistances with unilateral specific effects:

$$\log(X_{ij}) = \beta_0 + \alpha_i + \delta_j + \beta_1 \log(Y_i) + \beta_2 \log(Y_j) + \beta_3 \log(D_{ij}) + \eta_{ij} \quad (2.9)$$

Individual specific effects can also be combined into a bilateral specific effect in a cross-sectional estimation:

$$\log(X_{ij}) = \beta_0 + (\alpha\delta)_{ij} + \beta_1 \log(Y_i) + \beta_2 \log(Y_j) + \beta_3 \log(D_{ij}) + \eta_{ij} \quad (2.10)$$

The bilateral specific effect is better suited to the sum of the terms of multilateral resistance than the individual-specific effects. Thus, the linear model with a specific bilateral effect is more advantageous in the context of empirical analysis. Moreover, it should not be forgotten that these multilateral resistances can be approached at the variable remoteness. Multilateral resistance can also be replaced by specific effects on all three dimensions in the context of estimation on panel data:

$$\log(X_{ijt}) = \beta_0 + \alpha_i + \delta_j + \gamma_t + (\alpha\delta)_{ij} + (\alpha\gamma)_{it} + (\delta\gamma)_{jt} + \beta_1 \log(Y_i) + \beta_2 \log(Y_j) + \beta_3 \log(D_{ij}) + \eta_{ijt} \quad (2.11)$$

This specification was proposed by Baltagi et al. (2003). It is supposed to capture all sources of heterogeneity in the gravity equation. The parameters $\alpha_i, \delta_j\ et\ \gamma_t$ are called direct effects or main effects and the parameters $(\alpha\gamma)_{it}\ et\ (\delta\gamma)_{jt}$ are interpreted as interaction terms between the individual dimension and the temporal dimension. From the linear specification, one can increase qualitative variables such as the sharing of a language, a border, culture, the colonial past, the agreement of economic integration. In short, all the variables that can affect bilateral trade between two partner countries.

2.2. Estimation of the Gravity Equation

The estimation of the gravity model can be done on a cross-section or on panel data. This estimation requires the introduction of qualitative variables in place of country-specific fixed effects and time in a panel data estimate. Indeed, the specification of the gravity model can be fixed effects or random effects. The choice of the specification is determined by an endogeneity test. The fixed-effects model is most often used because it is suitable for an estimate in which multilateral resistances are replaced by dummy variables.

It is clear that estimating the fixed-effects model requires a resolution of the problem of identifying the parameters when the specification is based in addition to the gravity variables, on indicator variables specific to the exporting country and the importing country. These qualitative variables are supposed to capture all sources of unobservable heterogeneity in the specification of the severity equation.

2.2.1. Fixed effects model estimation

The estimation of the fixed-effects model is done on a cross-section or on panel data. The specific fixed effects are replaced by qualitative or indicator variables specific to partner countries and times in the context of panel data. The following table summarizes the fixed effects models on cross-section and panel data.

The introduction of dummy variables into the linear fixed effects model specific to exporting and importing countries will result in some changes in the specification. Let $E_i \ and \ M_j$ be two dummy variables associated respectively with exporting and importing countries such as:

$$E_i = \begin{cases} 1 & If\ i\ is\ an\ exporter \\ 0 & otherwise \end{cases} \qquad et \quad M_j = \begin{cases} 1 & If\ j\ is\ an\ importer \\ 0 & otherwise \end{cases}$$

These dummies must capture all sources of unobservable heterogeneity of the exporting country and the importing country. In the case of an estimation on N country, it is necessary to introduce into the model with fixed effects 2N dummy variables. As a result, the coefficients associated with the one-sided dimension of the model become unidentifiable. Moreover, the variables of the model are collinear if they vary in a single unilateral dimension. In other words, the variables $Y_i \ and \ Y_j$ vary respectively in the size of the exporting country and the importing country. The estimation of this model by the OLS will not make it possible to identify the effects of the variables $Y_i \ et \ Y_j$. In this case, the elimination of these variables from the model that only vary in the direction of the exporter or importer is more than necessary. The new model to estimate is in the linear form:

$$\log(X_{ij}) = \beta_0 + \beta_3 \log(D_{ij}) + \sum_i \alpha_i E_i + \sum_j \rho_j M_j + \eta_{ij} \qquad (2.12)$$

The estimation procedure differs in the case of the bilateral fixed-effect model. The bilateral dummy variable corresponding to the exporting country and the importing country is defined:

$$I_{ij} = \begin{cases} 1 & If\ i\ and\ j\ are\ partners \\ 0 & otherwise \end{cases}$$

In the case of an estimate on N countries, we will have in total $\frac{N(N-1)}{2}$ dummy variables. The introduction of this dummy variable implies a problem of identifying the coefficients associated with variables that vary bilaterally. In other words, the distance is collinear with the bilateral dummy variable. So the estimation is done without the variable distance and the model to estimate the following:

$$\log(X_{ij}) = \beta_0 + \beta_1 \log(Y_i) + \beta_2 \log(Y_j) + \sum_{i \neq j} \sum_j \theta_{ij} I_{ij} + \eta_{ij} \qquad (2.13)$$

Table 2.1. The gravity equation specified in cross-section and panel data

	Multiplicative Model	Linear model
Cross-sectional estimation	• The fixed-effects model specific to exporting and importing countries is given by the following equation: $X_{ij} = AY_i^{\beta_1} Y_j^{\beta_2} D_{ij}^{\beta_3} e^{\alpha_i + \delta_j} \varepsilon_{ij}$ $\alpha_i = \sum_i \theta_i E_i\ et\ \delta_j = \sum_j \rho_j M_j$ Then we get the following multiplicative model: $X_{ij} = AY_i^{\beta_1} Y_j^{\beta_2} D_{ij}^{\beta_3} e^{\sum_i \theta_i E_i + \sum_j \rho_j M_J} \varepsilon_{ij}$ • The bilateral specific effect model is written: $X_{ij} = AY_i^{\beta_1} Y_j^{\beta_2} D_{ij}^{\beta_3} e^{(\alpha\delta)_{ij}} \varepsilon_{ij}$ $(\alpha\delta)_{ij} = \sum_{i \neq j} \sum_j \theta_{ij} I_{ij}$ We obtain a new multiplicative equation: $X_{ij} = AY_i^{\beta_1} Y_j^{\beta_2} D_{ij}^{\beta_3} e^{\sum_{i \neq j} \sum_j \theta_{ij} I_{ij}} \varepsilon_{ij}$	• The linear model is written as : $\log(X_{ij}) = \alpha_i + \delta_j + \beta_0 + \beta_1 \log(Y_i) + \beta_2 \log(Y_j) + \beta_3 \log(D_{ij}) + \eta_{ij}$ Expressions of specific fixed effects are replaced: $\log(X_{ij}) = \beta_0 + \beta_1 \log(Y_i) + \beta_2 \log(Y_j) + \beta_3 \log(D_{ij}) + \sum_i \theta_i E_i + \sum_j \rho_j M_j + \eta_{ij}$ • The linear model with a specific bilateral effect is equal to: $\log(X_{ij}) = \beta_0 + (\alpha\delta)_{ij} + \beta_1 \log(Y_i) + \beta_2 \log(Y_j) + \beta_3 \log(D_{ij}) + \eta_{ij}$ In this equation, the expression of the bilateral fixed specific effect is replaced:

		$\log(X_{ij}) = \beta_0 + \beta_1 \log(Y_i) + \beta_2 \log(Y_j) + \beta_3 \log(D_{ij}) + \sum_{i \neq j} \sum_j \theta_{ij} I_{ij} + \eta_{ij}$

Anyway, we can have a much more general model that contains all the unilateral and bilateral fixed effects. In this case, the OLS method does not make it possible to identify the coefficients associated with gravitational variables Y_i, Y_j *and* D_{ij}. Then, the solution to the problem of identification is to estimate the model by considering that the effects are not fixed but random or to estimate the model with fixed effects on data of panel.

	Multiplicative Model	Linear model
Estimation of panel data	The multiplicative model on panel data is as follows: $X_{ijt} = AY_{it}^{\beta_1} Y_{jt}^{\beta_2} D_{ij}^{\beta_3} e^{u} \mathcal{E}_{ijt}$ with $u = \alpha_i + \delta_j + \gamma_t + (\alpha\delta)_{ij} + (\alpha\gamma)_{it} + (\delta\gamma)_{jt}$ $\alpha_i = \sum_i \theta_i^1 E_i$, $\delta_j = \sum_j \theta_j^2 M_j, \gamma_t = \sum_t \theta_t^3 T_t, (\alpha\delta)_{ij} = \sum_{i \neq j} \sum_j \lambda_{ij}^1 (EM)_{ij}$, $(\alpha\gamma)_{it} = \sum_i \sum_t \lambda_{it}^2 (ET)_{it}$ *et* $(\delta\gamma)_{jt} = \sum_j \sum_t \lambda_{jt}^3 (MT)_{jt}$ Let $v = \sum_i \theta_i^1 E_i + \sum_j \theta_j^2 M_j + \sum_t \theta_t^3 T_t + \sum_{i \neq j} \sum_j \lambda_{ij}^1 (EM)_{ij} + \sum_i \sum_t \lambda_{it}^2 (ET)_{it} + \sum_j \sum_t \lambda_{jt}^3 (MT)_{jt}$ Then, the multiplicative model becomes in this form: $X_{ijt} = AY_{it}^{\beta_1} Y_{jt}^{\beta_2} D_{ij}^{\beta_3} e^{v} \mathcal{E}_{ijt}$	The linear model on panel data is equal to: $\log(X_{ijt}) = \beta_0 + \alpha_i + \delta_j + \gamma_t + (\alpha\delta)_{ij} + (\alpha\gamma)_{it} + (\delta\gamma)_{jt} + \beta_1 \log(Y_{it}) + \beta_2 \log(Y_{jt}) + \beta_3 \log(D_{ij}) + \eta_{ijt}$ The linear model is written as follows: $\log(X_{ijt}) = \beta_0 + \beta_1 \log(Y_{it}) + \beta_2 \log(Y_{jt}) + \beta_3 \log(D_{ij}) + \sum_i \theta_i^1 E_i + \sum_j \theta_j^2 M_j + \sum_t \theta_t^3 T_t + \sum_{i \neq j} \sum_j \lambda_{ij}^1 (EM)_{ij} + \sum_i \sum_t \lambda_{it}^2 (ET)_{it} + \sum_j \sum_t \lambda_{jt}^3 (MT)_{jt} + \eta_{ijt}$

The panel fixed effects model captures all sources of heterogeneity in the gravity model. Indicator variables such as E_i, M_j, *and* T_t are introduced into this model and are respectively the dummies corresponding to the exporting and importing countries and time. In addition to these variables, we add the variables of interaction between the individual dimensions and the temporal ones that are $(EM)_{ij}, (ET)_{it}$, and $(MT)_{jt}$. The last two variables make it possible to capture cyclical effects such as economic fluctuations, commercial policy or the state of infrastructure. The coefficient θ_t^3 associated with the temporal dummy variable indicates the effect of global shocks on trade.

The panel model is very advantageous because it makes it possible to neutralize any heterogeneity and to take into account the temporal dimension of bilateral and multilateral resistance. However, this model has limitations in its estimation. Indeed, the introduction of the dummy variables specific to the exporting and importing countries causes a problem of estimating the effects of all the variables that are collinear with them and of all the variables that are invariant over time whose main cause is the multiplication of parameters to estimate. Thus, different possible specifications are proposed with the number of parameters to be estimated. In this case, we assume a panel composed of N countries and T periods (N = 100 and T = 10), we determine the number of the parameter to be estimated and the observations that result. In this case, we have a three-dimensional panel, so the number of observations is equal to NT (N-1). In the context of our example, we have 2940 observations.

Table 2.2. Specifications considered in panel data

Fixed specific effects	Number of variables	Interpretations
γ_t	10	Absence of multilateral resistance
$(\alpha\delta)_{ij}$; γ_t	$100 \times 10 + 10 = 9910$	Classic panel model with constant multilateral resistance over time
$(\alpha\delta)_{ij} = (\alpha\delta)_{ji}$; γ_t	$\frac{100 \times 10}{2} + 10 = 4960$	Conventional panel model with constant symmetric multilateral resistance over time
α_i; δ_j; γ_t	$100 \times 2 + 10 = 210$	Multilateral resistances constant over time
$(\alpha\gamma)_{it}$; $(\delta\gamma)_{jt}$; γ_t	$100 \times 2 \times 10 + 10 = 2010$	Multilateral resistances are specified correctly and depend on time. All the time, there is an absence of individual effects in the model.
$(\alpha\delta)_{ij}$; $(\alpha\gamma)_{it}$; $(\delta\gamma)_{jt}$; γ_t	$100(2 \times 10 + 10) + 10 = 6970$	This specification eliminates the maximum bias associated with the omission of

		relevant variables. But technically, the estimate is unachievable for large samples.

So according to the above table, we have a different specification of the panel gravity equation. All the time, the choice is not obvious because one can quite estimate all the versions of this model. Therefore, it is necessary to estimate all these specifications in order to analyze the robustness of the results and to identify the best specification. When the effects specific to the exporting and importing country are separated in a specification then there is a problem of multicollinearity with the variables that vary in the same dimension of the exporter or the importer. Therefore, questions arise about the nature of these effects in the gravity equation.

2.2.2. Fixed effects or random effects

Generally, the terms of multilateral resistance are approached by the variable remoteness which has become a less efficient approach. Thus, several researchers have tried to estimate the gravity equation by considering the terms of resistance as sources of heterogeneity. The fixed-effects model is very suitable in this practical case. Indeed, in the fixed-effects model, multilateral resistances are considered as fixed specific effects that will, in turn, be approximated by indicator variables. However, the introduction of dummy variables into the fixed effects model poses a problem of identifying the parameters. In this case, it is necessary to eliminate any collinear variable in the sense of the size of the exporting or importing country. Otherwise, another more effective possibility is to estimate the random effects model on a cross-section.

Fixed effect models[14] admit a correlation between specific effects and gravity variables, while random effects models accept zero correlation and specific effects are correlated with the error term. Thus, the individual effects and the explanatory or gravitational variables are orthogonal in the random-effects models. This orthogonality results in the existence of unobservable and country-specific factors that are not correlated with the gravity variables but are correlated with the error term. On a cross-section, the estimation of the random effects model is a solution to the problem of identification. In panel data, the estimation of the fixed effects model requires the adoption of several types of specification and one can also consider that the effects are random. In this case, it is necessary to proceed by specification tests.

[14] You can refer to Baltagi's work on fixed and random effects panel models to better understand endogeneity in panel theory.

Park (2010) suggests a scheme based on the choice of model to estimate.

- The test of the existence of fixed effects which is the Wald test;
- The test of the existence of random effects which is the LM test of Breush-Pagan.

Indeed, in a model one can find the existence or the absence of the two fixed and random effects. In the case of the existence of fixed effects and random effects, the model is chosen from the Hausman test whose assumptions are as follows:

$$H_0 : Cov(\alpha_i; x_{it}) = 0, \qquad \forall\, t$$

$$H_a : Cov(\alpha_i; x_{it}) \neq 0, \qquad \forall\, t$$

The acceptance of H_0 is choosing the model with random effects and in the case of a rejection, one chooses the model with fixed effects. The following table summarizes the specification tests and the corresponding decision rules.

Table 2.3. Fixed effects or random effects tests

	Fixed effects (Wald test)	Random effects (Breush-Pagan LM test)	Selected model
Rejection of H_0? (Hypothesis of the absence of effects)	No (No fixed effects)	No (No random effects)	Aggregated model and estimation by OLS
	Yes (Presence of fixed effects)	No (No random effects)	Fixed effects model
	No (No fixed effects)	Yes (Presence of random effects)	Random effects model
	Yes (Presence of fixed effects)	Yes (Presence of random effects)	The fixed or random-effects model is chosen by the Hausman test.

Source: Park (2010)

The choice of the model depends theoretically on the results of the tests carried out. If the data often show a significant presence of the types of effects then the Hausman test is performed to see which of the fixed effects model or the random-effects model should be selected. The estimation of fixed or random-effects models is also done by the Least Square Dummy Variable (LSDV) method. However, the estimation method often poses a problem in the linear model because the errors are no longer homoscedastic following the log-linear transformation and moreover it may be that there are no trade flows bound by a very large distance and the linear model is defined in zero. Therefore, other alternative methods are needed to estimate the linear gravitational equation. Moreover, gravity variables can be endogenous if the gravity model is not correctly specified. In this case, the estimation of the gravity equation will be biased.

2.3. Endogeneity and Gravity Variable

The endogeneity of gravity variables is a problem that is not common in estimating the gravity equation. The endogeneity of explanatory variable results in a non-zero correlation of the explanatory variable with the error term of the model to be estimated. Indeed, the problem of endogeneity in the gravity equation can be caused by a measurement error of the gravity variables, the omission of relevant variables in the model specification or the simultaneity of the gravity variables.

2.3.1. The omission of relevant variables

The omission of explanatory variables in the specification of the gravity equation is a source of unavoidable bias in the estimation. Indeed, the standard gravity equation is a function of gravity variables and other institutional, political and cultural variables. These variables have a considerable impact on bilateral trade between two partner countries. For example, if we consider McCallum's (1995) estimates, multilateral resistances were omitted from the specification and are associated with the error term of the model:

$$\eta_{ij} = \beta_4 \log(P_i) + \beta_5 \log(P_j) + \mu_{ij}$$

It is for this reason that he has obtained biased estimators. In the case of omission of relevant variables, the OLS estimator is no longer effective, is biased and non-convergent. So, it is more than necessary to choose the explanatory variables. For example, in the context of economic integration agreements, all variables related to these agreements that can influence bilateral trade between two partner countries must be included in the gravity equation specification.

2.3.2. Simultaneity

The gravity equation is composed of gravity variables and dummy variables intended to capture institutional, political, economic and cultural effects. The problem of endogeneity linked by simultaneity is caused by the size of the countries represented most often by their gross domestic product (GDP). But also, indicator variables reflecting a free trade agreement, a monetary union are sources of simultaneity bias.

Indeed, the determination of the country's GDP requires special information regarding the level of its exports and imports. But GDP is a determining factor of trade because the relationship between these two variables is quite positive. In other words, there is a simultaneous effect in determining the GDP and the level of trade. This level of trade is an integral part of the construction of GDP and GDP is a very important gravity variable in determining the level of trade. So GDP is an endogenous gravity variable. Cyrus (2002) studied the endogeneity of GDP and estimated the gravity equation by OLS and instrumental variables or double least squares with instruments: human capital, physical capital and the rate of labor accumulation. According to its results, the endogeneity of GDP does not have a significant impact on the estimation of the parameters. The OLS estimators are almost equal to those of the instrumental variables and the explanatory power of the model is identical regardless of the estimation method. This is because GDP is a function of the net, not gross, exports. In addition, gross exports constitute a small proportion of GDP. This means that the gravity equation is not a representation of the overall level of trade as a function of the GDP product of the partner countries, but on the contrary, the gravity equation establishes a positive relationship between the level of bilateral trade and GDP from these partner countries. Several works have confirmed the results of Frankel (1997) and Cyrus (2002). Therefore, the endogeneity of the GDP can be ignored in the estimation of gravity and estimate the model by the method of the OLS.

Another source of simultaneity bias is the introduction of an indicator variable associated with a free trade agreement. Indeed, assuming that the FTA variable is exogenous in a specification then our estimates will be biased. In reality, there is a simultaneous relationship between the FTA variable and the level of trade. Tinbergen (1962) was the first to introduce this variable considering it as exogenous in the estimation of the gravity equation. Since then, specialists have not found a solution to the nature of FTAs. For example, Baier and Bergstrand (2004) provided a theoretical and empirical model of the economic determinants of FTAs. Depending on their outcome, several factors explain the existence of an FTA such as the level of trade that can itself be explained by an FTA.

Rose (2000) worked on the impact of money on trade for the European union (EU). His results have raised particular attention to the dummy variable associated with a monetary union. Rose (2004) estimates were non-convergent and biased due to the endogeneity of the dummy variable. In fact, apart from the single currency, there are several other factors that have a considerable impact on the trade of the member countries. In other words, member countries of a monetary union could trade with each other not because they share a single currency, but there are other factors that can promote bilateral trade. It is these unobservable factors, such as the strong diversification of economies according to Kenen (1969), the intensification of trade and the correlation of economic cycles between member countries according to Frankel and Rose (1998) or the neighborhood according to Alesina and Baro (2002) who lead them to create or integrate a monetary union. The indicator variable reflecting membership in a monetary union or not is an endogenous decision and the resulting bias is self-selection bias. Only countries with specific characteristics that self-select to form a monetary union.

2.4. Estimation Issues of the Log-linear Model

By linearizing the multiplicative model, some properties of the variables of the model changed later. The log-linearization of the gravity equation has led to a change in the error term whose variance is no longer constant. Therefore, any estimation of the log-linear model by OLS will be biased.

In addition, another problem concerns the data used to estimate the log-linear model. Indeed, distance one of the control variables of the gravitational equation has a negative impact on trade flows and sometimes leads if it is very high, to zero trade flows between countries. In data with the level of trade of several countries, there may be zero trade flows between some countries because of the very large distance or other unobservable factors that do not facilitate bilateral trade between these countries. Since the log operator is indeterminate at zero then the linear model is no longer practical for estimating from this kind of data. Then, we need other alternative methods that we will present later, in order to estimate the model in a successful way with satisfactory results.

2.4.1. Heteroscedasticity and the log-linear model

The standard multiplicative gravitational equation is written as:

$$X_{ij} = \beta_0 Y_i^{\beta_1} Y_j^{\beta_2} D_{ij}^{\beta_3} \mathcal{E}_{ij} \qquad (2.14)$$

It is assumed that in this model the errors are heteroscedastic. The estimation is done by the GLS method. However, the OLS estimator remains unbiased and convergent but is not effective. The independent and identically distributed errors and the convergence condition of the OLS estimators are given by the following conditional expectation:

$$E(\mathcal{E}_{ij}/Y_i\,, Y_j, D_{ij}) = 1 \qquad (2.15)$$

The gravity equation of Anderson and Van Wincoop(2003) is the multiplicative equation augmented by the terms of bilateral and multilateral resistances:

$$X_{ij} = \beta_0 Y_i^{\beta_1} Y_j^{\beta_2} t_{ij}^{\beta_3} e^{\beta_4 P_i + \beta_5 P_j} \mathcal{E}_{ij} \qquad (2.16)$$

On the other hand, the errors are assumed to be iid and of hope equal to:

$$E(\mathcal{E}_{ij}/Y_i, Y_j, t_{ij}, P_i, P_j) = 1 \quad (2.17)$$

The linear form of the gravity model is obtained by applying the log operator to the multiplicative model. In practice, this is the most commonly used form for making estimates. The log-linearization implies the non-convergence of the OLS estimators. The linear gravitational equation is given by the following equation:

$$\log(X_{ij}) = \beta_0 + \beta_1 \log(Y_i) + \beta_2 \log(Y_j) + \beta_3 \log(t_{ij}) + \beta_4 \log(P_i) + \beta_5 \log(P_j) + \log(\mathcal{E}_{ij}) \quad (2.18)$$

The new term is in logarithm and is equal to $\eta_{ij} = \log(\mathcal{E}_{ij})$. This error term is assumed to be iid and is not correlated with gravity variables. To ensure the convergence of OLS estimates, it is necessary that the condition at which the expectation of the error term is equal to 1. Now, according to Jensen's inequality which states that:

$$E(\log(\mathcal{E}_{ij})) \neq logE(\mathcal{E}_{ij})$$

To further exploit this inequality, it is further assumed that the variance of the error term is a function of the gravity variables and is given by:

$$\sigma_{ij}^2 = f(Y_i, Y_j, t_{ij}, P_i, P_j)$$

The new error term is then distributed according to the log-normal law and an average which is a function of the variance of the error term $\mathcal{E}_{ij}$. This average is equal to the expectation of the new error term:

$$E(\log(\mathcal{E}_{ij}) \,/\, Y_i, Y_j, t_{ij}, P_i, P_j) = E(\eta_{ij}/Y_i, Y_j, t_{ij}, P_i, P_j) = \frac{1}{2}\log(1 + \sigma_{ij}^2) \quad (2.19)$$

It is clear that the convergence condition of the estimators is not respected. Moreover, the condition of orthogonality is not verified because the expectation of the term of the log-linear model is a function of the gravity variables. So the estimation of the linear model by the MCO is not convergent. Several authors have worked on this question and have tried to determine the cause related to the problem of heteroscedasticity. We can cite Silva Santos and Tenreyro (2006) justified by the fact that the error term is a function of the gravity variables and the log-linearization of the gravity equation in the presence of heteroscedasticity implies non-consistent estimators because the expectation of the logarithm of the random variable is a function of the moments of its distribution. Moreover, Kalirajan (2008) shows that the presence of heteroskedasticity is linked to an omission of relevant variables, which is justified by the fact that commercial costs are generally approximated by geographical distance.

2.4.2. Dealing with the zero trade flows

A very common problem in empirical trade studies is the presence of zero trade flows in the data from which estimates are made. If such data are present, the estimation of the linear model by the OLS is no longer possible because the log operator is not at zero and remains indeterminate. Indeed, the presence of zero trade flows can be interpreted according to three origins:

- Trade flows may be related to trade barriers such as the very remote geographical distance or the very high trade costs which consequently result in an absence of trade between certain countries;
- The low level of trade between some countries may lead to a rounding of the corresponding value zero in a database on trade flows. As a result, this low level of trade must be defined as being below a predetermined threshold. In general, the countries that trade little, are those of small sizes or those distant because of the geographical distance which separates them;
- A very important problem in an empirical study is the problem of the lack of data. In such situations, zeros are missing data.

However, approaches have been proposed to properly conduct empirical studies using such data. Historically, the estimate was made without the null values. Positive and non-zero values are considered and the model is estimated by OLS. This is a very simplistic approach of the reality because to suppress the null flows of the estimate would be a loss of information. As a result, the estimation results remain very critical. Another approach would be to estimate by OLS by replacing zero flows with approximately low values; which is very simple because such data do not correspond to actual levels of trade. In this case, any estimation of the linear model by the OLS is biased.

Moreover, in the presence of zero trade flows, one could not log-linearize the gravity equation. These amounts to estimating the multiplicative gravitational equation not by the OLS but by the pseudo-maximum likelihood method. It is clear that the estimation of the multiplicative model is non-convergent but constitutes a solution to the problem of zero flow. As a result, the most practical approaches for dealing with zeros are the Heckman approach and the generalized linear model approach.

2.4.2.1. *Heckman's selection model*

Heckman's selection model is a solution to the problem of zero trade flows. Heckman (1979) generalized Tobin's approach (1958). To carry out its application, we use the linear specification of the gravitational equation. In this case, we use the specification to Anderson and Van Wincoop (2003):

$$\log(X_{ij}) = \beta_0 + \beta_1 \log(Y_i) + \beta_2 \log(Y_j) + \beta_3 \log(t_{ij}) + \beta_4 \log(P_i) + \beta_5 \log(P_j) + \eta_{ij} \quad (2.20)$$

Let :

$$Z = \log(X_{ij}); \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{bmatrix}; \quad W\ is\ a\ matrix\ containing\ explanatory\ variables.$$

The linear gravity model is equivalent to the following equation:

$$Z = W'\beta + \mathcal{E}$$

Heckman's approach to the problem of trade flow zeros comes in the context of two equations:

$$\begin{cases} Z^* = W'\beta + \mathcal{E}_1 = f(W_{ij}, \beta) + \mathcal{E}_{1ij} & (1) \\ p_{ij} = V'\delta + \mathcal{E}_2 = f(V_{ij}, \delta) + \mathcal{E}_{2ij} & (2) \end{cases}$$

We suppose that :

$$\begin{pmatrix} \mathcal{E}_{1ij} \\ \mathcal{E}_{2ij} \end{pmatrix} \backsim \boldsymbol{\mathcal{N}} \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho\sigma_1 \\ \rho\sigma_1 & 1 \end{pmatrix} \right]$$

(1) represents a behavioral equation that determines the expected level of trade between countries. (2) is the selection equation of the estimation sample that determines whether observations in equation (1) are observed or not. From these equations, this system results:

$$Z = \begin{cases} Z^* & if\ p_{ij}^* > 0 \\ is\ not\ bservable & if\ p_{ij}^* \leq 0 \end{cases}$$

This selection equation relates the latent variable to the observed explanatory variables. Moreover, this set of explanatory variables V must not only include all the gravity variables, but also V must preferably include at least one variable that affects the probability that two countries commit to exchange but not the volume of trade. As an additional variable, Helpman et al. (2008) in their extension of the Heckman selection model in the case of heterogeneity of firms, chose the cost of entry into the exporter and importer market, noted C_{ij}. However, we define the selection probability p^*_{ij} and the dummy variable p_{ij} which is equal to unity if the observations are in the sample and zero otherwise. In this case, we can write this selection equation as follows:

$$p^*_{ij} = V'\delta + \mathcal{E}_2 \; avec \; p_{ij} = \begin{cases} 1 \;\; si \; p^*_{ij} > 0 \\ 0 \;\; si \; p^*_{ij} \leq 0 \end{cases}$$

The selection equation can be written as follows:

$$p^*_{ij} = \delta_0 + \delta_1 \log(Y_i) + \delta_2 \log(Y_j) + \delta_3 \log(t_{ij}) + \delta_4 \log(P_i) + \delta_5 \log(P_j) + \delta_6 \log(C_{ij}) + \mathcal{E}_{2ij}$$

The Heckman procedure is done in two steps:

First step :

The probability must be estimated for the observations to be in the gravity model sample using the Probit qualitative model. The following probability is often estimated:

$$Prob(p_{ij} = 1) = \Phi(\delta_0 + \delta_1 \log(Y_i) + \delta_2 \log(Y_j) + \delta_3 \log(t_{ij}) + \delta_4 \log(P_i) + \delta_5 \log(P_j) + \delta_6 \log(C_{ij})) = \Phi(V'\delta)$$

$$\Phi = Distribution\ function\ of\ the\ normal\ centered\ reduced\ law$$

$$\phi = Density\ of\ probability\ of\ the\ reduced\ normal\ centered\ law$$

The inverse of Mills ratio[15] noted by IMR is determined from the Probit estimate which is equal to:

$$IMR = \widehat{\lambda_{ij}} = \frac{\phi(V'\hat{\delta})}{\Phi(V'\hat{\delta})}$$

This ratio corresponds to the probability of selecting variables omitted in equation (2) and the introduction of an additional variable corrects the bias of omitted variables; which implies a satisfactory estimate in the presence of a nonrandom selection sample.

Second step:

[15] Under Mills, John P. (1926) name. Heckman applied it for correction of selection bias

We estimate the inverse equation (1) of the inverse of the Mills ratio:

$$Z = W'\beta + \rho\sigma_1\lambda_{ij} + e_{ij}$$

The error term is given e_{ij} and its expectation and conditional variance are:

$$E\left[e_{ij} / p_{ij}^* > 0\right] = 0 \quad and \quad V\left[e_{ij} / p_{ij}^* > 0\right] = \sigma_1(1 - \rho^2\gamma_{ij})$$

$$\text{With } \gamma_{ij} = \lambda_{ij}\left(\lambda_{ij} + V'\delta\right) \ and\ \lambda_{ij} = \frac{\phi(V'\delta)}{\Phi(V'\delta)}$$

The augmented equation of the IMR variable is rewritten as:

$$Z = W'\beta + \beta_\lambda\widehat{\lambda_{ij}} + e_{ij} + \beta_\lambda\left(\widehat{\lambda_{ij}} - \lambda_{ij}\right)$$

$$= W'\beta + \beta_\lambda\widehat{\lambda_{ij}} + e_{ij}^*$$

$$= \begin{bmatrix} W' & \widehat{\lambda_{ij}} \end{bmatrix} \begin{bmatrix} \beta \\ \beta_\lambda \end{bmatrix} + e_{ij}^*$$

$$= X^*\beta^* + e_{ij}^*$$

The estimation of this equation is done by the OLS. This estimate, also called the Heckman estimator is equal to:

$$\widehat{\beta^*} = (X^{*\prime}X^*)^{-1}X^{*\prime}Z$$

2.4.2.2. Generalized linear models (GLM)

Generalized linear models are an answer to the problem of trade flow zeros in estimating the linear gravity equation. These models are very general and are used when the dependent variable is a linear function of explanatory variables. In fact, the GLMs have a somewhat specific formulation and make it possible to study the dependent variable which is a function of explanatory variables.

Thus we give the following multiplicative gravity model:

$$X_{ij} = \beta_0 Y_i^{\beta_1} Y_j^{\beta_2} D_{ij}^{\beta_3} e^{\beta_4 P_i + \beta_5 P_J} \mathcal{E}_{ij} \qquad (2.21)$$

As a result of this multiplicative equation, the linear model given by the equation below:

$$\log(X_{ij}) = \beta_0 + \beta_1\log(Y_i) + \beta_2\log(Y_j) + \beta_3\log(D_{ij}) + \beta_4\log(P_i) + \beta_5\log(P_j) + \eta_{ij}$$

$$= W_{ij}{}'\beta + \eta_{ij}$$

Let's apply to the linear model the mathematical property $e^{\log(a)} = a$ and we will have the following equality:

$$X_{ij} = \exp(W_{ij}{}'\beta + \eta_{ij}) \qquad (2.22)$$

We suppose that the variable X_{ij} follows a law belonging to the exponential family of average determined from the logarithmic link function:

$$\log(X_{ij}) = W_{ij}'\beta + \eta_{ij} \quad alors \qquad E[X_{ij} \,/\, W_{ij}] = \exp(W_{ij}'\beta)$$

Thus, the dependent variable can follow any law belonging to the exponential family within the GLM framework. Therefore, the volume of exchanges can follow a discrete law (Poisson, Binomial or Negative Binomial) or a continuous law (Gamma, Normal, Normal inverse).

The estimation method is then done either by the Maximum Likelihood method if the errors are normal, or by the Pseudo Maximum Likelihood method if the errors follow the Poisson's Law (PPML), Gamma (GPML), Binomial (BPML) or Negative Binomial (NBPML).

2.4.2.2.1. Pseudo Maximum likelihood method

We suppose that the variable X_{ij} follows a law belonging to the exponential family. Its probability density is given by the function below:

$$f(X_{ij}) = \begin{cases} \Pr(X_{ij} = x_{ij}) & if\ X_{ij}\ follows\ a\ discrete\ law \\ \Pr(X_{ij} \leq x_{ij}) & if\ X_{ij}\ follows\ a\ continuous\ law \end{cases}$$

The likelihood function, which is equal to the product of the probability densities, is thus determined:

$$L(.) = \prod_{i,j} f(X_{ij})$$

Then, we will linearize the likelihood function; which leads to the log of the likelihood function:

$$\log L(.) = \log \prod_{i,j} f(X_{ij}) = \sum_i \sum_j \log f(X_{ij})$$

To determine the Maximum Likelihood Pseudo estimators, we maximize the log of likelihood:

$$Max_{\beta} \log L(.) = \sum_{i} \sum_{j} \log f(X_{ij})$$

The first-order conditions associated with this maximization program give us the estimators of the gravity equation:

$$\frac{\partial \log L(.)}{\partial \beta} = 0$$

The estimation by the Pseudo Maximum Probability method has several advantages and disadvantages. However, the PPML method from the Poisson law or the GPML from the Gamma law is the most used in practice.

2.4.2.2.2. Panorama of the estimation methods

The estimation of the gravity model is done from the specification form. The estimation method differs from the multiplicative form to the linear form. Thus, we summarize in the following table some of the possible estimation methods applied to the multiplicative and linear gravity model.

Table 2.4. Alternative estimation methods in the presence of zero trade flows

Estimation method	Advantages	Disadvantages	References
Truncated OLS	-Simple	- Loss of information related to the elimination of trade flows; - The coefficients are biased.	Linders and de Groot (2006); Westerlund and Wilhelmsson (2009); Martin and Pham (2008)
OLS (1+X_{ij})	-Simple - Eliminates the problem of zero trade flows.	- The coefficients are biased.	Linneman (1966) ; Bergeijk and Oldersma (1990) ; Wang and Winters (1991); Balwin and DiNino (2006)
	-Simple - Controls the sources of unobservable heterogeneity.	- Loss of information related to the abandonment of the	Màtyàs (1998); Egger (2000); Glick and Rose (2002); Egger and Pfaffermayr

Fixed effects panel		constant terms of the regression. - Eliminates the problem of zeros flow of trade. - The bias of the sample selection	(2003); Micco et al. (2003); Andrew et al. (2006); Henderson and Millimet (2008)
Heckman Selection Model	- Different sets of variables and coefficients to determine the probability of having observations in the sample and the value of the dependent variable. - Absence of multicollinearity - Provides rationality to flow trade zeros.	- It can be difficult to find an identification restriction. - The exclusion of variables is required.	Bikker and de Vos (1992) ; Linders and de Groot (2006) ; Martin and Pham (2008)
Non-linear least squares (NLS)	- Eliminates the problem of zero trade flows.	- Assigns more weight to observations with greater variance. - Not robust in case of heteroscedasticity. - The bias of the sample selection	Silva Santos and Tenreyro (2006)

Pseudo Poisson Maximum Likelihood (PPML)	- Eliminates the problem of zero trade flows. - Provides unbiased estimators in the presence of heteroscedasticity. - All observations have equal weights. - Positive Mean	- May exhibit limited bias of the dependent variable when a significant portion of the observations is critical.	Westerlund and Wilhelmsson (2009); Siliverstovs and Schumacher (2009); Liu (2009); Shepherd and Willson (2009); Martinez-Zarzosa and al. (2007); Silva Santos and Tenreyro (2006); An and Puttitanun (2009)
Gamma Pseudo Maximum Likelihood (GPML)	- Eliminates the problem of zero trade flows. - Robust in case of heteroscedasticity.	- Low weight for observations with high conditional expectation	Martinez-Zarzosa et al. (2007)

Source: Gomez Herrera, comparing alternative methods to estimate the gravity model, march 2016

2.4.2.2.3. Park's test (1966)

The estimation of the gravity model with zero trade flows is very large and involves a selection of the probability law belonging to the family of exponential laws if the estimation method chosen is that of the Pseudo Maximum Probability.

Indeed, the Pseudo Maximum Probability method has excellent statistical properties if the laws that are associated are of the family of exponential laws. In empirical studies, Poisson's law and Gamma's law are used. In general, there is no rule for choosing one law over another to estimate the gravitational equation. However, we can identify an adequate law thanks to its moments. This identification is based on a test called Park's test (1966), which takes place in three stages:

i) Do the regression of the dependent variable X_{ij} on the explanatory variables and estimate from the OLS;

ii) Determine the estimated value of the dependent variable $\widehat{X_{ij}}$ and take back the residuals $\widehat{\eta_{ij}}$;

iii) Make a second regression of the $\log(\widehat{\eta_{ij}})$ on the variable $\log(\widehat{X_{ij}})$ and determine the relationship between the expectation and the variance of X_{ij} which is written in the following way:

$$V(X_{ij}) = c\left[E(X_{ij})\right]^{\theta} \ ou \ \log\left[V(X_{ij})\right] = \log(c) + \theta\log\left[E(X_{ij})\right]$$

Decision rule :

- If θ = 1 then the variance is exactly proportional to the expectation. This is a characteristic of Poisson's law. So we choose the Poisson's law and we estimate the gravity model by the PPML method;
- If θ = 2 then the variance is proportional to the square of expectation, which implies the Gamma law. Then we estimate the gravity model by the GPML method.

2.5. Estimation of the Gravity Model by Panel Data Cointegration Technics

In many studies, researchers have always considered in panel data estimation of the gravity equation that gravitational variables are stationary. More recently, several studies have proved the importance of taking into account the trend of corresponding data to gravity variables and their possible non-stationarity. Thus, in the case of non-stationarity of the variables, a different statistical configuration has been highlighted to make estimates of the gravity model. If the non-stationary nature of the series is not taken into account then the gravitational regression appears as a fallacious regression. However, the problem of fallacious regression is less important than in panel than in time series. However, according to the Kao and Chiang (2000) studies, as the fixed effects estimators are asymptotically normal, the results will be biased.

Panel cointegration techniques are used for various problems of estimating the gravity equation, in particular, the endogeneity problem of gravity variables. To do this, we must proceed in three steps that concern the application of cointegration panel techniques on the gravity model:

i) Determine the order of integration of the gravity variables through panel unit root tests;
ii) Cointegration test among integrated gravity variables through panel cointegration tests;
iii) Use panel cointegration estimators to provide reliable estimates for the gravity model.

The basic specification of the gravitational equation reflects a positive relationship between the volume of trade and the Gross Domestic Product (GDP). In fact, a country's trade flows and GDP are more likely to be linked to a non-stationarity problem. If trade flows and individual GDP are non-stationary then we have to address them with more adequate statistical techniques such as the study of cointegration. The study of non-stationary panels in the gravity equation is very recent and depends on the heterogeneity and homogeneity of the panels. All the time the estimation results are quite surprising; because the heterogeneity of the panels is admitted in case of non-stationarity of the variables and moreover the coherent estimates result from the fallacious regression and the cointegration.

In addition, if the panels are heterogeneous, the dummy regression and the cointegration model are not distinguishable empirically and the coefficients can be estimated from the aggregated data. If the panels are homogeneous, the coefficients eliminate the individual effects and this homogeneity can be tested in a general or more significant way in the cointegration framework. In addition, cointegration tests in a homogeneous panel imply a reduction in the non-stationarity of the variables to be made stationary. Cointegration tests are therefore interpreted as tests for panel homogeneity or a non-parametric test approach to identify homogeneous regions or countries in trade. However, we can perform studies on homogeneous as well as heterogeneous panels. It all depends on the research goals we aim to achieve.

In the context of the gravitational model, heterogeneous and homogeneous panel cointegration techniques (with respect to parameters) are used in the empirical analyzes. Indeed, the heterogeneity in the gravity equation can be caused by certain qualitative variables such as culture, language, economic integration agreements, innovations, research, and development, etc. We had shown in section 2 of this chapter how to account for the unobservable heterogeneity of panels in the gravity equation that can come from fixed effects and random effects.

2.5.1. Determination of the non-stationary gravitational process

We start from the gravitational specification of Anderson and Van Wincoop (2003), which determines the volume of trade as a function of partner countries' GDP and bilateral and multilateral resistance:

$$X_{ij} = \frac{Y_i Y_j}{Y^W}\left(\frac{t_{ij}}{P_i P_j}\right)^{1-\sigma} \varepsilon_{ij} \qquad (2.23)$$

This specification can be rewritten as follows:

$$X_{ij} = \frac{(Y_i Y_j)^{\beta_1}}{Y^W}\left(\frac{t_{ij}}{P_i P_j}\right)^{\beta_2} \varepsilon_{ij} \quad with\ \beta_1 = 1\ et\ \beta_2 = 1-\sigma \qquad (2.24)$$

Moreover, bilateral resistance is often approximated by a geographical distance that does not evolve over time. In our case, this variable is a function of distance as defined by Anderson and Van Wincoop (2003) and is given by $t_{ij} = b_{ij}D_{ij}$. The variable b_{ij} depends on the countries that conduct the exchange by their heterogeneity as for the economic integration agreements that reduce or completely eliminate the customs duties that may affect the cost of bilateral trade. So the resulting linear model is as follows:

$$\log(X_{ij}) = \beta_0 + \beta_1(\log(Y_i) + \log(Y_j)) + \beta_2\log(D_{ij}) + \beta_2(\log(b_{ij}) - \log(P_i) - \log(P_j)) + \log(\mathcal{E}_{ij})$$

$$= \beta_0 + \beta_1 y_{ij} + S_{ij} + \eta_{ij}$$

If we suppose that $\log(X_{ij}) = x_{ij}$, we can see that the gravity model is written as a bilateral fixed-effect model determined by S_{ij}. The estimation of such a model requires the introduction of indicator variables, which pose a problem of identification of the parameters. Thus, the use of the panel model was more than necessary. Since geographic distance does not depend on time, we consider it a fixed bilateral specific effect in the panel gravity model. In addition, because of the estimation problems related to the panel specification, we use the specification given in the table in section 2 of this chapter:

$$x_{ijt} = \beta_0 + \delta_{ij} + \gamma_t + \beta_1 y_{ijt} + \eta_{ijt} \quad (2.25)$$

From this specification, one can increase time-dependent qualitative variables such as Regional Economic Integration Agreements (REIAs) if they exist according to the empirical application of selected country groups:

$$x_{ijt} = \beta_0 + \beta_1 y_{ijt} + \sum_{k=1}^{N} \theta_k (REIA)_{kijt} + \gamma_t + \delta_{ij} + \eta_{ijt} \quad (2.26)$$

$$REIA_{kijt} = \begin{cases} 1 & If\ i\ and\ j\ are\ members \\ 0 & if\ i\ and\ j\ are\ not\ members \end{cases}$$

Other time-dependent variables can be added to further capture temporal heterogeneity in the gravity model. Thus, authors like Gomez and Tamarit (2011)[16] have increased in this specification variables such as GDP per capita and a deterministic trend in their article on the study of the effect of the Euro on trade by cointegration techniques. Trade between countries is getting better and better

[16] In their work, Gomez and Tamarit (2011) used Baldwin's critique in estimating a non-stationary gravity model with and without dummy variables and a trend.

with the advancement of transport technologies, so innovation is a time-dependent factor. So the gravity model is written in the form of a TS (Trend Stationary) process:

$$x_{ijt} = \beta_0 + \beta_1 y_{ijt} + \sum_{k=1}^{N} \beta_k (REIA)_{kijt} + \delta_{ij} + \gamma_t + \alpha_{ij}.t + \eta_{ijt} \quad (2.27)$$

The variable y_{ijt} is the sum of the GDP logs of the exporter and the importer. This variable can be determined from the basis of nominal or real values.

δ_{ij} is the country-specific fixed effect and captures all individual fixed factors including all the unobservable characteristics associated with a given country pair that historically affects bilateral trade flows. This bilateral effect is invariant in time, so it also includes the geographical distance, the region, the common language, the common border, etc. This bilateral effect makes it possible to control the bias of omitted variables.

γ_t is a temporal effect common to the partner countries and $\alpha_{ij}.t$ is a deterministic trend specific to a couple of countries (i, j). These temporal effects capture the relative development of bilateral trade across all partner countries over time. This development can be materialized by technical progress, innovation in transport or in the field of telecommunication, the reduction of tariffs and bilateral trade costs, the liberalization of commercial processes, etc.

The parameters θ_k determine the effect of REIA on trade through trade policies that impact intraregional and extra-regional trade.

The panel stationarity study of the gravity equation is extensive and applies to several methods for testing the presence of unit roots and cointegration. In addition, the estimation of the gravity model can be done according to different methods. However, these tests differ from panel analysis to time series analysis.

2.5.2. Unit root tests in panel data

To study the stationarity of a panel variable, it is necessary to determine the relationship between the temporal dimension and the individual dimension, since unitary root and stationarity tests depend on it. In other words, first, one has to perform a statistical test, which determines whether the time series are independent of the cross-section. It is well known that the non-stationary panel tests have limitations related to their robustness defect or the size distortion when there is a correlation between the error terms of the cross-section. In particular, panel unit root tests are biased in the case of a conclusion in favor of stationary variance when individuals represent a dependent cross-section.

2.5.2.1. Cross-section dependence test or Pesaran's test (2004)

The dependence of the cross-section has been largely proven in practice, particularly when the variables are macroeconomic time series. Nevertheless, as part of the gravity model, we have variables for a couple of countries. Moreover, the construction of these variables follows the economic integration of a group of countries.

Gomez and Tamarit (2011) therefore suggest a two-step procedure. First, the Pesaran (2004) test statistic must be calculated to determine if the time series are independent of the cross-section. Indeed, the Pesaran test (2004) makes it possible to test the null hypothesis of the independence of the cross-section. Under this assumption, the Pesaran CD statistic (2004) converges to a normal distribution. This test remains valid for panels with temporal and individual dimensions approaching infinity, and in particular for a large individual dimension and a small temporal dimension.

2.5.2.2. Generations of panel unit root testing

In the literature, the panel stationary tests assuming that individual is independent, while some heterogeneity across i is admitted is named as "the first generations of panel unit root tests". The most used tests in the empirical study of the gravity equation by cointegration techniques are the Levin and Lin (1993) test which has been improved and replaced by the Levin test, Lin and Chu (2002), the IPS unitary root test of Im, Pesaran and Shin (1995) and the Hadri LM test (2000) which allows testing the null hypothesis of stationarity. In the case of dependence among individuals, the second generation of panel unit tests come to play. Among this later generation, Pesaran's CADF test (Cross-Section Augmented Dickey-Fuller) is the simplest.

2.5.3. Cointegration tests and Estimation technics

Cointegration tests have been developed and applied for time series. In a very practical way, these tests are adapted to the panel data. The cointegration techniques applied to the gravity equation are very advantageous and make it possible to overcome the problem of non-stationarity of trade flows and GDP. These cointegration techniques make it possible to analyze and estimate the long-term relationship between these two gravity variables.

The estimation of the severity equation by cointegration techniques can be done according to the homogeneity or heterogeneity of its estimation parameters. Indeed, the ADF test of Kao (1999) makes it possible to test the null hypothesis of non-cointegration in the framework of homogeneity or heterogeneity of the parameters. On the other hand, the McCoskey and Kao (2001) test based on the LM series in time series of Harris and Inder (1994) makes it possible to test the null hypothesis of non-cointegration in the context of heterogeneity of the parameters of the long run relationship. Kao and Chiang (2000) recommend the estimation method (FMOLS) of Phillips and Hansen (1990) which determines the fully modified estimators and the Saikkonen (1991) and Stock and Watson dynamic least squares (DOLS) method. (1993). These methods have good statistical properties and solve the problem of autocorrelation and endogeneity. Moreover, these methods show that the estimators are asymptotically and normally distributed of zero means.

The estimation of the gravity model is quite complex and involves several challenges whose estimation method requires in-depth econometric techniques. This estimate is done on a cross-section or on Panel data. Indeed, the gravitational equation is traditionally estimated by its multiplicative form and its log-linear form. The most commonly used specification is linear and is estimated by the OLS method. However, the estimation of the linear model requires an analysis of all the gravity variables and must take into account the problems of multilateral resistance that constitutes the first challenge. Because trade flows are a function of the multilateral resistances that are represented by the index of remoteness or by unilateral fixed specific effects or a specific bilateral effect in the framework of a transversal framework or by individual heterogeneities, temporal and interaction between the temporal and individual dimension in the context of panel data. Nevertheless, these procedures have limitations related to the problem of identifying model parameters.

Estimation of the linear model by OLS is not always allowed especially when one is confronted with particular data containing zero trade flows. Another challenge is that of the treatment of zeros of trade flows. Therefore, the estimation is done by the Heckman selection procedure or by the GLMs, the estimation of which is done by the PMV method. The estimation of the gravity model has become very large. More recently, cointegration techniques have been used in the context of non-stationarity of gravity variables.

In addition, the specification used in this chapter is in the context of the homogeneity of firms. Other much more recent studies deal with estimation problems in case of heterogeneity of firms with very thorough literature of the gravitational equation.

CHAPTER THREE

3. EMPIRICAL ANALYSIS OF THE GRAVITY MODEL APPLIED ON INTERNATIONAL TRADE BY DIFFERENT ECONOMETRIC APPROACHES

This chapter consists of studying trade by an empirical analysis of the gravity model while basing it on its theoretical foundations as well as its linear specification. On the one hand, we will address several issues of the gravitational equation, which we have studied extensively in the theoretical part. These problems include the estimation of the gravity model in the presence of zero trade flows and the gravitational specification according to the form of multilateral resistances. In addition, we will perform the estimation of the gravity equation by several estimation methods in order to have a much broader vision and thus be able to test the sensitivity of the estimation results. In addition, the estimation of the gravity model will be based on panel data as part of a fixed and random effects model to better capture the heterogeneity between the different partner countries. On the other hand, we will perform a series of cointegration estimation of a fixed-effects panel model to compare the results of the same model in the context where the panel data is assumed to be stationary. This new gravitational approach will allow us to analyze the robustness of the results of the different series of estimates made and to study the impact of non-stationary panel series in estimating the gravity model.

Our empirical analysis will focus on a sample of Economic Community of West African States (ECOWAS) countries whose data are observed over a period from 2001 to 2014 and relate to the bilateral trade flows between the different member countries.

3.1. Data Sources and Variable choices

The empirical study of the gravity model in this thesis concerns panel data from ECOWAS countries observed over the period from 2001 to 2014. ECOWAS is made up of 15 countries including 8 French-speaking countries (Benin, Burkina Faso, Cote ivory, Guinea, Mali, Niger, Senegal, Togo), 5 English-speaking countries (Ghana, Gambia, Liberia, Nigeria, Sierra Leone) and 2 Lusophone countries (Cape Verde, Guinea Bissau). Among the ECOWAS member countries, there are 8 countries that form the West African Economic and Monetary Union (WAEMU). Our panel consists of three dimensions including the two individual dimensions and the temporal dimension. So the number of observations in our database is 15 * 14 * 14 = 2940. These panel data are accessible and have fewer missing observations. Moreover, the ECOWAS countries are homogeneous and are part of the developing countries with relatively high economic growth, sometimes reaching double figures in some countries. The choice of variables is quite obvious because the gravity model is a function of the gravity variables (GDP and distance). Indeed, distance is more than fundamentally in the empirical study of the gravity model. This gravity variable determines the set of quantifiable international trade barriers and frictions. As we have analyzed well in the theoretical part, these barriers are none other than the bilateral resistance approximated by the bilateral distance which symbolizes the cost of transport, the cost of customs clearance or disembarkation, and the multilateral resistance that is represented by fixed effects in a panel model. In addition to the gravity variables, we added dummy variables in order to optimally analyze the applied gravitational equation. The following table summarizes all the variables as well as the source of the data.

Table 3.1. Selected variables

Variables	Interpretations	Sources
$\boldsymbol{X_{ij}}$	Annual export flow in current dollars (C $) from country i to country j	2015 Chelem Database
$\boldsymbol{GDP_i, GDP_j}$	Gross domestic product in C $ of countries i and j	WB
$\boldsymbol{POP_i, POP_j}$	The population of country i and country j	WB
$\boldsymbol{DIST_{ij}}$	Distance in Km between the capitals of countries i and j	2015 Chelem Database
$\boldsymbol{COM\_BORD_{ij}}$	Dummy Variable that is 1 if i and j share a common border and 0 otherwise	Chelem Database
$\mathbf{WAEMU}_{\boldsymbol{ij}}$	Dummy Variable which is worth 1 if i and j belong to WAEMU and 0 otherwise	Made by the author
$\boldsymbol{COM\_LANG_{ij}}$	Dummy Variable that is worth 1 if i and i share an official language and 0 otherwise	Chelem Database

The gravity model that concerns our empirical analysis focuses on several variables in addition to the gravity variables of GDP and distance. We have largely explained the importance of these variables in the explanation of trade of any economic integration.

The variables $\boldsymbol{POP_i}$ and $\boldsymbol{POP_j}$ determine the level of the population of the ECOWAS countries. According to several studies such as that of Xubei (2010), the introduction of the population variable is identical to that of the GDP per capita. In addition, the introduction of this variable into the gravitational equation will allow us to determine the types of goods exchanged between countries. Indeed, most ECOWAS countries have the same level of development. However, there are remarkable economic disparities within the region. Nigeria, Ghana and Cote d'Ivoire remain the most advanced countries in the sub-region and hold some of the largest populations. According to the theory of comparative advantage in the sense of HO, the ECOWAS countries export goods of which they have a comparative advantage. Moreover, according to Linder (1961) countries trade more if their level of development is similar. The analysis of the data shows the existence of zero and sometimes weak flows between some countries even though the distance is not important. This can be explained by the lack of infrastructure and lack of industry in the area. Therefore, the exchange of agricultural goods is more important in the ECOWAS zone than that of manufactured, finished or semi-finished goods. The sign for this variable is positive if the development levels are the same and are negative if the development levels are different. So the expected sign can be positive or negative.

The variable $\boldsymbol{COM\_BORD_{ij}}$ is an indicator variable that indicates the existence of a common border or not. In the literature, countries that share a common border exchange more. So this variable has a positive impact on trade. So the expected sign is positive.

The variable $\boldsymbol{COM_LANG_{ij}}$implies both the sharing of an official language and the same colonizer. It is worth unity if the partner countries have the same colonizer and zero if not. This variable has an impact on trade on a social and political level. In addition, ethical languages play a very important role in trade.

The variable $\mathbf{WAEMU}_{\boldsymbol{ij}}$ represents the monetary union and the only one in ECOWAS. This union brings together 8 ECOWAS countries, all of which are French-speaking except Guinea Bissau. Member countries hold a single currency that is pegged to the euro. Since the euro is an overvalued currency then WEAMU's commercial goods are expensive and can sometimes weaken its exports to other non-member countries. This empirical study will further enable us to study the impact of this monetary union in ECOWAS.

3.2. Hypothesis Test Violation

Like any empirical study, the study of statistical properties is more than necessary in order to avoid some consequences of the results of estimations. The most common problems in the gravity model are the multicollinearity of the explanatory variables and the distribution of trade flows and the stationarity of the variables. This step makes it possible to give a broader view of the nature of the specification to choose, therefore, that of a particular estimation method.

3.2.1. Multicollinearity of explanatory variables

Multicollinearity can have significant consequences for estimation results, identification and statistical properties of estimators. The collinearity between the explanatory variables can be strong or weak. On the other hand, strong collinearity between the variables implies the non-convergence of the estimators. Multicollinearity is explained by one or more redundant relationships. There are several criteria to check the presence of collinearity between variables. One can perform multicollinearity tests or calculate for each explanatory variable X_k its Variance Inflation Factor called VIF which is equal to:

$$VIF_k = \frac{1}{1-R_k}$$

With R_k equal to the coefficient of determination of the regression of X_k on all the other explanatory variables. The inverse of VIF is called tolerance. Thus the decision rule is given by a VIF strictly greater than 10. In other words, there is a risk of collinearity between the explanatory variables if the tolerance is less than 0.1.

Table 3.2. Correlation matrix

	X_{ij}	GDP_i	GDP	POP_i	POP_j	D_{ij}	CB_{ij}	CL_{ij}	MU_{ij}
X_{ij}	1,000								
GDP_i	0,372	1,000							
GDP_j	0,090	-0,032	1,000						
POP_i	0,263	0,849	-0049,	1,000					
POP_j	0,149	-0,049	0,849	-0,067	1,000				
D_{ij}	-0,062	0,117	0,121	0,135	0,142	1,000			
CB_{ij}	0,052	-0,036	-0,036	-0,024	-0,024	-0,547	1,000		
CL_{ij}	0,027	-0,027	-0,027	-0,129	-0,129	-0,093	0,218	1,000	
MU_{ij}	-0,014	-0,118	-0,118	-0,128	-0,118	-0,125	0,219	0,473	1,000

Source: Author's calculation on STATA 12

$$CB_{ij} = COM\_BORD_{ij}, CL_{ij} = COM\_LANG_{ij} et\ MU_{ij} = \text{WAEMU}_{ij}$$

The analysis of the table above shows a negative correlation between ECOWAS trade flows and distance and between trade flows and the WAEMU variable. On the other hand, the correlation of trade flows with other variables is positive. This shows that a significant distance leads to a decline in trade flows.

Table 3.3. Collinearity between explanatory variables

	VIF	$Tolerance$
gdp_i	7,94	0,125902
gdp_j	6,91	0,144649
pop_i	7,35	0,136027
pop_j	6,42	0,155765
d_{ij}	1,80	0,554390
COM_BORD_{ij}	1,83	0,545234
COM_LANG_{ij}	1,51	0,661056
$WEAMU_{ij}$	1,46	0,685742
Mean	4,40	

Source: Author's calculation on STATA 12

The variables in the table above are the logarithms of the explanatory variables. The analysis in Table 3.3 shows that all VIFs are strictly less than 10 and that the resulting tolerances are greater than 0.1. So we can deduce that there is no risk of collinearity between the explanatory variables.

3.2.2. Distribution of trade flows

The hypothesis of normality of errors is rejected in several empirical studies of the gravity model. Indeed, the fairly important presence of trade flows is one of the causes. In addition, an asymmetric distribution and spread to the right with thick distribution tails maybe another. The presence of these characteristics has very negative consequences for OLS estimation results and forecasts.

Thus, to test the normality of the dependent variable, we use the Jarque Bera test based on the calculation of statistics such as Skewness (S) or asymmetry and Kurtosis (K) or flattening. The (S) is a moment of order 3 and the (K) is a moment of order 4. For a given distribution of mean m and standard deviation σ, the two statistics are written formally:

$$S = \frac{E[(x-m)^3]}{\sigma^3} \; et \; K = \frac{E[(x-m)^4]}{\sigma^4}$$

The Skewness statistic has the following properties:

- If $S > 0$, the distribution is said to be spread to the right;
- If $S = 0$, the distribution is symmetric ;
- If $S < 0$, the distribution is said to be spread to the left.

A very high Kurtosis value implies a sharp distribution whereas a low value indicates a flattened or crushed distribution.

For a standard normal distribution of law, Skewness S = 0 and Kurtosis K = 3. The assumptions on which the Jarque Bera test is based are:

$$H_0^1: S = 0; H_0^2 K = 3; H_0^3 S = K - 3 = 0$$

The J statistic of Jarque Bera follows a Khi2 and for a p-value less than 0.05, the hypothesis H_0 is rejected. Table 3.4 summarizes the results of the test of normality of trade flows.

Table 3.4. Jarque Bera normality test of the trade flows

$$\begin{cases} H_0: \quad trade\ flows\ are\ normally\ distributed \\ H_a: trade\ flows\ are\ not\ normally\ distributed \end{cases}$$

	Pr(Skewness)	**Pr(Kurtosis)**	**Khi2(2)**	**Pr (Jarque Bera)**
Trade flows in level	0,0000	0,0000	4593,93	0,0000
Trade flows in logarithmic	0,0000	0,0000	53,23	0,0000

Source: Author's calculation on STATA 12

The results in Table 3.4 show that the trade level or log distribution is not symmetrical because the probability of Skewness is well below the 5% threshold. Since this distribution is not symmetrical, it does not matter the result of the Kurtosis test because the law can not be a normal law. The following graphs are the distribution of trade flows in level and logarithm.

The analysis of Jarque Bera's statistics shows that the probability associated with it is well below the set threshold. Then, the distribution of trade flows in level or logarithm is not a normal distribution. However, the comparison of Fig.1 and Fig.2 with the graphical representation of the distribution of a normal distribution, tells us that the logarithmic transformation of trade flows partly corrects asymmetry and flattening.

Figure 3.1. *Distribution of trade flows in level*

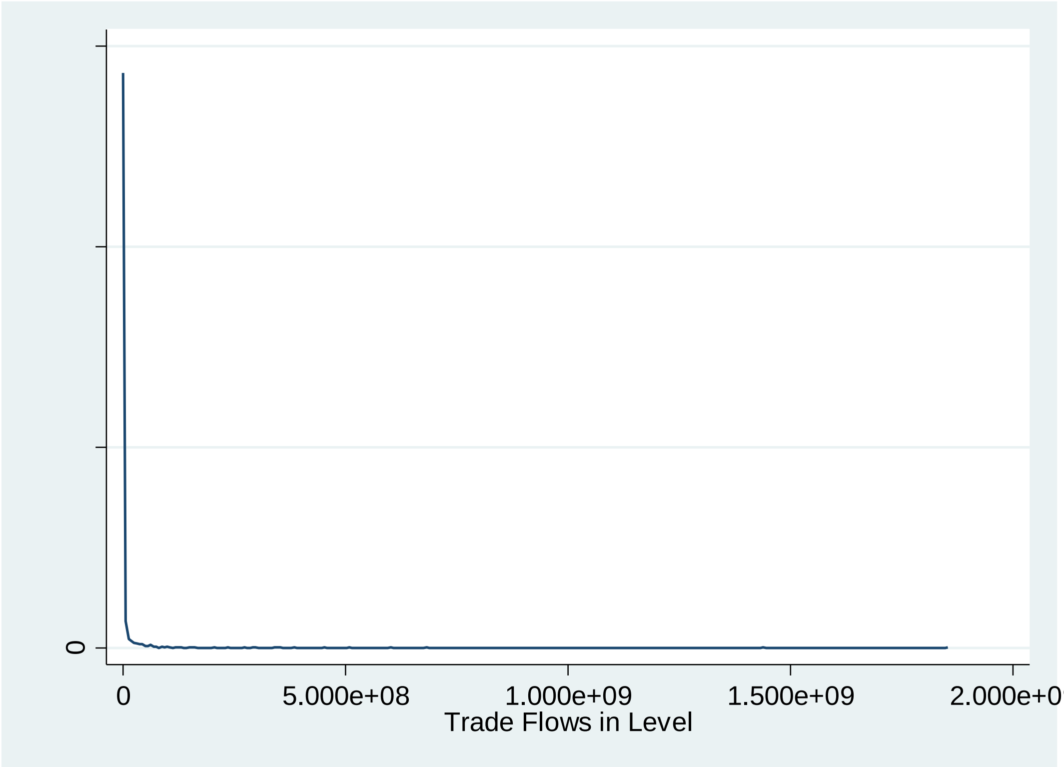
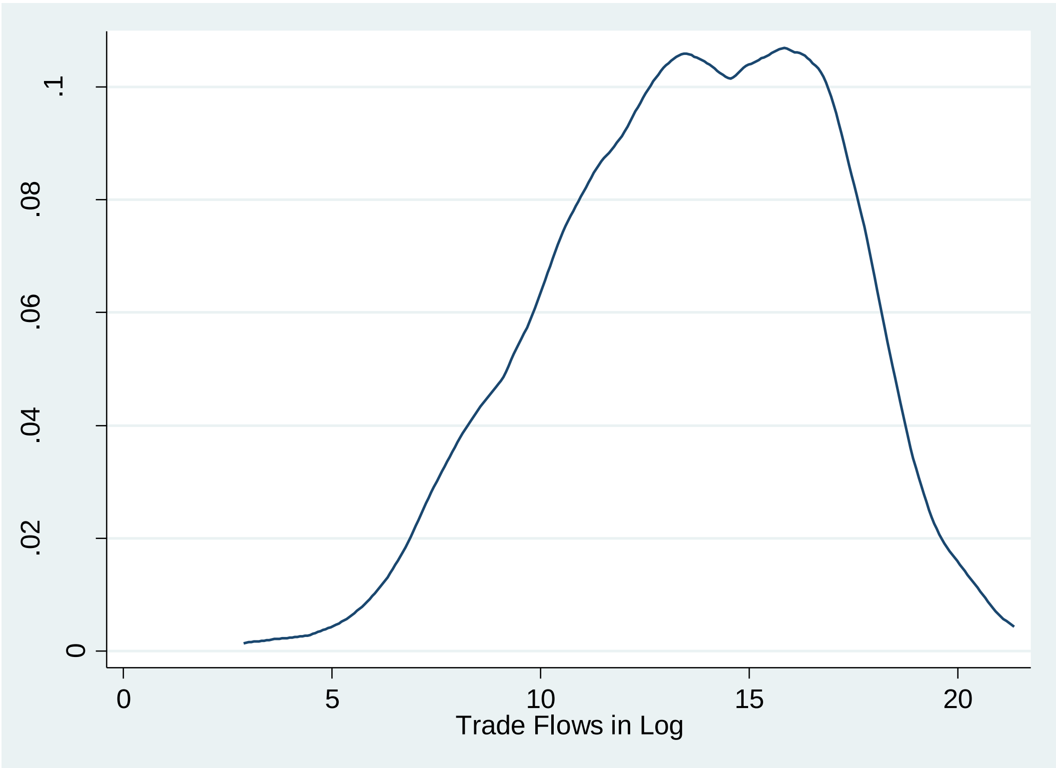


Figure 3.2. *Distribution of trade flows in logarithm*ic



3.3. Gravity Equation Estimation

Our empirical analysis of the gravity model is based on a logarithmic gravitational specification. The specification chosen here is that of Anderson and Van Wincoop (2003) in which internal and external multilateral resistances are considered to be fixed or random effects. The objective of our study is to give an explanation of the trade flows within ECOWAS based on the variables we have just presented. This empirical study will be carried out by series of estimates with which multilateral resistance will be determined. All this in order to have robust results to bring an optimal analysis to our empirical study.

The estimate will be done initially with techniques that take into account zero trade flows. Because the estimation by the MCO is no longer valid in the presence of these types of data. Nevertheless, we will apply the solutions provided to estimate the gravitational equation in the presence of zero trade flows. In a second step, we will estimate the severity model by cointegration techniques in order to make a broader analysis of the results obtained.

The following gravitational relation gives the specification of Anderson and Van Wincoop (2003):

$$X_{ij} = \frac{GDP_i GDP_j}{GDP^M}\left(\frac{t_{ij}}{P_i P_j}\right)^{1-\sigma}$$

The linear gravity model is obtained by a logarithmic transformation. Multilateral resistance can be represented by fixed effects that depend on time or by indices of remoteness. So our gravitational panel specification is:

$$\log(X)_{ijt} = \beta_0 + \beta_1 \log(GDP_{it} \times GDP_{jt}) + \beta_2 \log(t)_{ijt} + \beta_3 \log(P)_{it} + \beta_4 \log(P)_{jt} + \varepsilon_{ijt}$$

The transport cost t_{ij} is approximated by the distance variable. From this model, we increase the parameter γ_t, a temporal effect, which determines the evolution of economic integration within ECOWAS. This time effect captures the relative development of bilateral trade across all partner countries over time. This development can be materialized by technical progress, innovation in transport or in the field of telecommunication, the reduction of tariffs and bilateral trade costs, and the liberalization of commercial processes...

For reasons of simplification, we suppose:

$$\log(X)_{ijt} = x_{ijt}, \log(DIST)_{ij} = dist_{ij}, \log(GDP_{it} \times GDP_{jt}) = gdp_{ijt},\ \log(POP_{it} \times POP_{jt}) = pop_{ijt} \text{ et } \log(REMOTE_{it} \times REMOTE_{jt}) = remote_{ijt}$$

The gravitational equations considered to solve the problem of multilateral resistances the treatment of the zero trade flows are the following ones:

$$(1) \qquad x_{ijt} = \beta_1 gdp_{ijt} + \beta_2 pop_{ijt} + \beta_3 dist_{ij} + \beta_4 WEAMU_{ijt} + \beta_5 COMBORD_{ij} + \beta_6 COMlANG_{ij} + \alpha_i + \delta_j + \gamma_t + \tau_{ij} + \varepsilon_{ijt}$$

$$(2) \qquad x_{ijt} = \beta_0 + \beta_1 gdp_{ijt} + \beta_2 pop_{ijt} + \beta_3 remote_{ijt} + \beta_4 WEAMU_{ijt} + \beta_5 dist_{ij} + \beta_6 COMBORD_{ij} + \beta_7 COMLANG_{ij} + DUM\_EXP + DUM\_IMP + DUM\_TEM + \varepsilon_{ijt}$$

$$(3) \qquad X_{ijt} = exp\big(\beta_1 gdp_{ijt} + \beta_2 pop_{ijt} + \beta_3 remote_{ijt} + \beta_4 WEAMU_{ijt} + \beta_5 dist_{ij} + \beta_6 COMBORD_{ij} + \beta_7 COMLANG_{it} + DUM\_EXP + DUM\_IMP + DUM\_TEM + \varepsilon_{ijt}\big)$$

(1) is a gravitational equation where the terms of multilateral resistances are time-independent fixed or random individual effects of time and the bilateral fixed effect t_{ij} represents the common variables of the partners and symmetric like the distance, the common border, the shared languages (official and ethnic) and any other variable that may have an impact on trade. It also controls the zeros of trade flows. We will perform the Hausman test to determine the choice between fixed effects and random effects.

(2) represents the gravity model with an approximation of multilateral resistances by the remoteness index and the variables DUM_EXP, DUM_IMP and DUM_TEM are respective effects to exporting, importing countries and time. This equation will make it possible to apply the Heckman selection method in two stages because there are zero trade flows.

(3) is the gravity equation used for estimation by the PML method. The most commonly used estimation method is PPML.

3.4. Stationarity of the Gravity Variables

The study of the stationarity of gravity variables is a new concept in the empirical analysis of the gravity model. Macroeconomic variables such as GDP and trade flows do not change in the same way. So the study of their stationarity is more than necessary in order to determine a solid estimate of the gravitational equation.

Zwinkels and Beugeldjik (2010) have shown in their study that gravity variables tend to be cointegrated. Therefore, there is often the risk of fallacious regression. However, this problem is less important in the panel than in time series because the estimators of a fixed-effects model with nonstationary data are asymptotically normal. Moreover, the application of the standard methods of the gravity model does not usually consider a possible endogeneity of the long-term relationship between trade flows and explanatory variables. However, Fidrmuc (2009) shows that the estimation of a fixed-effects gravity model by standard techniques is as robust as the estimation by panel cointegration techniques. Other studies show that the use of panel cointegration techniques is quite similar to that of the fixed-effects model. This implies that the bias resulting from the non-stationarity of the gravity model is rather weak. The estimation of the gravity model by cointegration techniques will be done from equation (1) which can be rewritten in this form:

$$x_{ijt} = \beta_1 pib_{ijt} + \beta_2 pop_{ijt} + \alpha_i + \delta_j + \gamma_t + \tau_{ij} + \varepsilon_{ijt}$$

For this, one must determine before any study the cross-section dependence test.

3.4.1. Pesaran's CD (2004)

Gomez and Tamarit (2011) therefore suggest a two-step procedure. First, the Pesaran (2004) test statistic must be calculated to determine if the time series are independent of the cross-section. Indeed, the Pesaran test (2004) makes it possible to test the null hypothesis of the independence of the cross-section. Under this assumption, the Pesaran CD statistic (2004) converges to a standard normal distribution. The results of the test are given in Table 3.5:

Table 3.5. CD test before estimation

$$\begin{cases} H_0: Independence\ of\ the\ cross\ section \\ H_a: \quad Dependence\ of\ the\ cross\ section \end{cases}$$

Pesaran CD Test (2004)				
Variables in Level	GDP_{it}	GDP_{jt}	POP_{it}	POP_{jt}
CD	517,90	517,90	552,49	552,49
P-value	0,000	0,000	0,000	0,000

Variables in Log	$\boldsymbol{gdp_{it}}$	$\boldsymbol{gdp_{jt}}$	$\boldsymbol{pop_{it}}$	$\boldsymbol{pop_{jt}}$
CD	521,87	521,87	552,45	552,45
P-value	0,000	0,000	0,000	0,000

Source : Autor's calculation on STATA 12 $\boldsymbol{CD \sim (0, 1)}$

The results in this table clearly indicate that the p-value is zero for all the variables, so we reject the hypothesis H0. So the cross-section is dependent on the time dimension. However, the Pesaran

CD test before estimation of the fixed effects model does not allow to test the trade flows variable either in level or logarithm due to its non-convergence or non-normality and to the indeterminable values of the logarithmic transformation of zero flows.

We then carried out the dependence tests of the post-estimation cross-section. The test results are given in the table below:

Table 3.6. CD test post-estimation

Cross-section dependence test

$$\begin{cases} H_0: Independence\ of\ the\ cross\ section \\ H_a: \quad Dependence\ of\ the\ cross\ section \end{cases}$$

Gravity model with fixed effects		
Tests	**Statistics**	**Probabilities**
Breush-Pagan (Khi2)	29945,50	0,0000
Pesaran (LM)	67,9737	0,0000
Pesaran (CD)	7,0861	0,0000
Friedman (Khi2)	115,6313	0,0000
Gravity model with random effects		
Tests	**Statistics**	**Probabilities**
Breush-Pagan (Khi2)	91530,62	0,0000
Pesaran (LM)	360,3620	0,0000
Pesaran (CD)	1,8783	0,0603
Friedman (Khi2)	65,5299	0,0000

Source: Autor's calculation on STATA 12

The results in Table 3.6 indicate probabilities below 1%, 5% and 10% thresholds for both fixed and random effects. Therefore we reject the null hypothesis and the results are in favor of the alternative hypothesis which implies a cross-sectional dependence.

3.4.2. Panel data unit root test

In this part of our work, the Pesaran CD test indicates us to use the second generation of panel unit root test. Since Pesaran's CADF test (2007) is the simplest, we apply it to test the null hypothesis of nonstationarity. The test results are grouped in the following tables:

Table 3.7. Pesaran's CADF test (2007)

Pesaran's CADF test (2007) $\begin{cases} H_0: Panels\ have\ unitary\ roots \\ H_a: \quad Panels\ are\ stationary \end{cases}$				
Variables	**Statistic**	**P-value**	**Number of observations**	**Year**
$\boldsymbol{X_{ijt}}$	8.965	1,0000	2310	14
$\boldsymbol{GDP_{it}}$	-3.987	1,0000	2310	14
$\boldsymbol{GDP_{jt}}$	-3.987	1,0000	2310	14
$\boldsymbol{POP_{it}}$	-32.993	0,0000	2310	14
$\boldsymbol{POP_{jt}}$	-32.993	0,0000	2310	14
Pesaran's CADF Test of the first differences				
$\boldsymbol{D(X_{ijt})}$	-8.272	0,0000	2310	14
$\boldsymbol{D(GDP_{it})}$	-10.947	0,0000	2310	14
$\boldsymbol{D(GDP_{jt})}$	-10.947	0,0000	2310	14
$\boldsymbol{D(POP_{it})}$	-28.996	0,0000	2310	14
$\boldsymbol{D(POP_{jt})}$	-28.996	0,0000	2310	14

Source: Author's calculation on STATA 12

Table 3.8. Pesaran's CADF test (2007) in log

Pesaran's CADF test (2007) $\begin{cases} H_0: Panels\ have\ unitary\ roots \\ H_a: \quad Panels\ are\ stationary \end{cases}$				
Variables	**Statistics**	**P-value**	**Nber of obs**	**Year**
$\boldsymbol{gdp_{it}}$	4.760	1,0000	2310	14
$\boldsymbol{gdp_{jt}}$	4.760	1,0000	2310	14
$\boldsymbol{pop_{it}}$	-15.286	0,0000	2520	14
$\boldsymbol{pop_{jt}}$	-15.286	0,0000	2520	14
Pesaran's CADF Test of the first differences				
$\boldsymbol{D(gdp_{it})}$	-17.883	0,0000	2310	14
$\boldsymbol{D(gdp_{jt})}$	-17.883	0,0000	2310	14
$\boldsymbol{D(pop_{jt})}$	19.724	1,0000	2310	14

Source: Author's calculation on STATA 12

Table 3.9. Pesaran's CADF test (2003) of country i and j common variables

Pesaran's CADF test (2003) $\begin{cases} H_0: \text{Panels have unitary roots} \\ H_a: \text{Panels are stationary} \end{cases}$				
Variables	**Statistics**	**P-value**	**Nber of obs**	**Year**
$\boldsymbol{gdp_{ijt}}$	2.003	0.977	2310	14
$\boldsymbol{pop_{ijt}}$	-12.372	1,0000	2520	14
Pesaran's CADF Test of the first differences				
$\boldsymbol{D(gdp_{ijt})}$	-17.308	0,0000	2310	14
$\boldsymbol{D(pop_{ijt})}$	14.797	1,0000	2310	14

Source: Author's calculation on STATA 12

The analysis in Table 3.7 shows that the population variable is stationary at level, but trade flows and GDP are stationary at the first differentiation. As a result, the trade flows and GDP variables are integrated of order I (1) and the population is integrated of order I(1). The study of the stationarity of the variables in logarithm indicates the same result in Table 3.8 for variables other than the populations ones. The analysis in Table 3.9 gives identical analyzes. In summary, trade flows and GDP in logarithm are integrated of order I (1) ,but the log population is stationary at log level.

3.4.3. Panel data cointegration test

The Kao test (1999) tests the null hypothesis of non-cointegration. The results of these tests are grouped in the following table:

Table 3.10. Cointégration test in panel data

$$\begin{cases} H_0: \quad \text{Absence of cointegration} \\ H_a: \quad \text{Variables are cointegrated} \end{cases}$$

Test de Kao (1999)		
Dickey-Fuller	**Statistics**	**Probability**
DF	-19,6554	0,0000
DF*	-8,7405	0,0000
ADF	73,4652	0,0000

Source: Author's calculation on STATA 12

The Kao test (1999) indicates that the probability or the p-value is below the threshold of 1% or 5% so we reject the hypothesis H0. So the gravitational variables x_{ijt}, gdp_{ijt} and pop_{ijt} are cointegrated. Then the estimation of the gravity model can be done by the method DOLS or FMOLS.

3.5. Gravity Model Estimations

Different scenarios are used to estimate of the gravity model. We will perform various methods listed in this chapter to analyze the robustness of each. For that, we will take into account several problems of the gravitational equation as the problems of measurement of multilateral resistances, zeros flow of commerce and stationarity.

Model (1) solves the problem of measuring multilateral resistance through unilateral individual effects and neutralizing zeros of trade flows by a bilateral individual effect. This model will be the subject of a study in the context of the cointegration of gravity variables. Model (3) is a transformation of model (2) to overcome the estimation problem in the presence of zero trade flows.

3.5.1. Fixed effects or random effects

Before we estimate our gravity model, we will perform two tests to determine the nature of the effects. Then, we will apply the Park procedure (1966) in Table 2.3 which corresponds to the fixed effects and random effects tests.

Table 3.11. Wald's fixed effects test

$$\begin{cases} H_0: \quad No\ fixed\ effects \\ H_a: Presence\ of\ fixed\ effects \end{cases}$$

Fixed effects test		
Wald test	**F Statistic**	**Probability**
	2,16	0,0092

Source: Author's calculation on STATA 12

Table 3.12. Breusch Pagan's random effects test

$$\begin{cases} H_0: Presence\ of\ random\ effects \\ H_a: \quad No\ random\ effects \end{cases}$$

Random effects test		
Breusch Pagan test	**Khi-2 Statistic**	**Probability**
	0,00	1.000

Source: Author's calculation on STATA 12

The results of the tables clearly indicate that there is both the presence of fixed effects and random effects in our gravitational model. We will estimate the gravity equation according to the fixed and random effects and we will determine the Hausman test in order to make the choice of the model.

3.5.2. Gravity model estimation applied to trade flows

The fixed and random effects, PPML, Heckman's Selection, DOLS, and FMOLS, were employed for estimating the gravity model. We grouped the estimation results together to better analyze their robustness.

Table 3.13. Estimation results of the fixed effects model applied to the trade flows of ECOWAS countries

Variables	FE	FE PPML	Remote FE	Remote PPML FE	Heckman's selection	FE DOLS	FE FMOLS
gdp_{ijt}	0,082 (0,400)	0,297 (0,000)	0,129 (0,269)	0,319 (0,000)	0,268 (0,000)	1,099 (0,001)	2,092 (0,036)
pop_{ijt}	2,166 (0,014)	-1,052 (0,000)	2,227 (0,012)	-1,347 (0,000)	0,005 (0,807)	1,740 (0,017)	1,580 (0,000)
$dist_{ij}$	-0,168 (0,917)	-0,234 (0,000)	-0,537 (0,743)	-0,013 (0,001)	-0,322 (0,000)		
$WEAMU_{ijt}$					0,336 (0,000)		
$COMBORD_{ij}$					0,679 (0,000)		
$COMLANG_{ij}$	-0,188 (0,879)	0,120 (0,00)	-0,331 (0,789)	-0,031 (0,261)	0,411 (0,000)		
$remote_{it}$			-1,391 (0,077)	-0,727 (0,000)	-0,436 (0,000)		
$remote_{jt}$			0,555 (0,438)	-0,468 (0,000)	-0,074 (0,252)		
$Constante$	-58,533 (0,048)	-30,800 (0,000)	-48,176 (0,119)	-10,306 (0,000)	-1,924 (0,000)		
exporter Individual effect	**Yes (dropped)**	**Yes (dropped)**	**Yes (dropped)**	**Yes**	**Yes**	**Yes**	**Yes**
importer Individual effect	**Yes (dropped)**	**Yes (dropped)**	**Yes (dropped)**	**Yes**	**Yes**	**Yes**	**Yes**
temporal effect	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**
bilateral effect	**Yes (dropped)**	**No**	**Yes (dropped)**	**No**	**No**	**Yes**	**Yes**
Observations	2220	2940	2220	2940	2940	1883	2057
R^2_{within}	0,203		0,205				
$R^2_{between}$	0,387		0,390				
$R^2_{overall}$	0,304		0,306				
R^2		0,562		0,656		0,946	0,886

Source: Author's estimation on STATA 12

The numbers in parentheses are the p-values of the coefficients. They make it possible to interpret the significance of the parameters or elasticities in the gravity model.

Table 3.14. Estimation results of the random effects model applied to the trade flows of ECOWAS countries

Variables	RE	RE PPML	Remote RE	Remote PPML RE	Heckman's selection	RE DOLS	RE FMOLS
gdp_{ijt}	0,082	1,190	0,109	1,127	0,268	0,197	0,225
	(0,400)	(0,000)	(0,443)	(0,000)	(0,000)	(0,011)	(0,036)
pop_{ijt}	2,166	-0,382	2,955	-0,233	0,005	1,740	1,580
	(0,014)	(0,000)	(0,024)	(0,045)	(0,007)	(0,017)	(0,000)
$dist_{ij}$	-0,168	-0,686	-1,074	-1,094	-0,322		
	(0,917)	(0,000)	(0,000)	(0,000)	(0,000)		
$WEAMU_{ijt}$	2,964	1,274	-0,126	1,105	0,336		
	(0,061)	(0,000)	(0,498)	(0,000)	(0,000)		
$COMBORD_{ij}$	-0,716	0,287	1,330	0,170	0,679		
	(0,689)	(0,065)	(0,000)	(0,254)	(0,000)		
$COMLANG_{ij}$	-0,188	-0,097	0,910	-0,357	0,411		
	(0,879)	(0,546)	(0,000)	(0,782)	(0,000)		
$remote_{it}$			-2,365	-0,287	-0,436		
			(0,046)	(0,055)	(0,000)		
$remote_{jt}$			1,166	-1,229	-0,074		
			(0,276)	(0,000)	(0,252)		
$Constante$	-59,717	-22,882	-61,334	-4,396	-1,924		
	(0,040)	(0,000)	(0,156)	(0,073)	(0,000)		
exporter Individual effect	**Yes (dropped)**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**
importer Individual effect	**Yes (dropped)**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**
temporal effect	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**
bilateral effect	**Yes**	**No**	**Yes**	**No**	**No**	**Yes**	**Yes**
Observations	2220	2940	2220	2940	2940	1883	2057
R^2_{within}	0,203		0,586				
$R^2_{between}$	0,981		0,989				
$R^2_{overall}$	0,883		0,887				
R^2		0,587		0,773		0,946	0,886

Source: Author's estimation on STATA 12

The numbers in parentheses are the p-values of the coefficients. They make it possible to interpret the significance of the parameters or elasticities in the gravity model.

Table 3.15. Hausman's test

$$\begin{cases} H_0: & Random\ effects \\ H_a: & Fixed\ Effects \end{cases}$$

	Statistic Khi2 (n)	Probability
Panel with individual effects	0,002 (n=11)	1,000
Panel with Remote	0,009 (n=13)	1,000

Source: Author's calculation on STATA 12

The choice of the model is determined by the probability. It is clear that it is well above the threshold of 5%. So we do not reject the null hypothesis of random effects. The gravity model chosen is that of a random-effects panel.

3.5.3. Discussion of the estimation results

The ultimate goal of our empirical study was to estimate the gravitational equation in the presence of zero trade flows and a cointegration relationship. However, the application of these estimation techniques required some manipulation from the point of view of the specification of the gravity model. We can cite the problem of multilateral resistance and the collinearity of certain variables in the estimation of the fixed-effects model.

The estimation of fixed and random effects gravity models (Tables 3.13 and 3.14) indicates that only the population variable has a significant effect on trade flows. What is not the case, according to theoretical analysis, GDP and distance have a significant effect on trade flows. In addition, the expected sign of the population is negative or positive. So this problem of estimation can be at two levels:

- The zero trade flows have been replaced by a bilateral fixed effect to control the missing log trade logarithmic observations;
- Multilateral outward and inward resistance has been approached by unilateral individual effects independent of time.

For this, an estimation of the gravity model was carried out using the PPML method which allows integrating the zero trade flows while controlling the multilateral resistances by unilateral individual effects. The analysis of the estimation results shows that GDP and WEAMU have a significant positive effect on ECOWAS trade. Distance has a negative and significant effect on trade flows. The expected negative sign of the population is verified and is significant. This confirms what Linder (1961) said, and the ECOWAS countries have similar levels of development.

Concerning multilateral resistances, an estimation of the gravity model was made by approximating them by the indices of remoteness. The estimation results are identical to those of the fixed and random effects model estimates. The estimation of the gravity model with the remoteness indices of the PPML method has the same conclusions as that of the gravity model with unilateral individual effects. The variable remoteness has a negative and significant impact, according to the gravitational theory of Anderson and Van Wincoop (2004).

In conclusion, the unilateral individual effects represent the same level as the variables remoteness the multilateral resistances. In addition, replacing trade-flow zeros with bilateral effects is not always the best way, as in the case of our study, because this leads to a huge loss of information, and therefore to completely biased estimators.

There are other alternatives for estimating a gravitational equation in the presence of trade flow zeros. Heckman's selection method is quite effective for studying with such data. The estimation of our gravity model by the Heckman selection procedure indicates that all gravity variables are significant and only the distance has a negative effect on trade flows.

Another question we are asked about is the cointegration of the gravitational relation. The study of stationarity has shown that trade flows and GDP are not stationary, despite the fact that the period of our study panel is rather small and does not make it possible to detect a significant change in these two macroeconomic variables. These variables are stationary at the first differentiation. In fact, since the beginning of the 21st century, the ECOWAS countries have experienced a remarkable increase in their population. For some countries, the population has more than doubled during the panel period. The study of the stationarity shows that the population variable is not stationary, even at the first differentiation. However, it becomes a variable in an integrated logarithm of an order 2. We are in the presence of a gravity relation with two integrated variables of order 1 and an integrated variable of order 2.

In this part of our study, we have seen that gravity variables are cointegrated. We draw attention to the distance variable that is time-independent and stationary, was neutralized by an independent bilateral effect as well as other variables like common border, common language and any other variable having an impact on trade. In this relation of gravitational cointegration, it is a question of studying the relation between the commercial flows and the GDP and the population.

The estimation of the gravity model indicates results identical to our first estimates. GDP and population have significant effects. The population has a positive effect, this implies that the

ECOWAS countries have different levels of development. This is justified by the fact that in ECOWAS there are two types of economies, namely weak economies, and intermediate economies. These results show nevertheless that there is an economic disparity in the sub-region and moreover the majority is in the category of weak economies.

The empirical estimation of the gravity model is quite complex, especially when we are dealing with certain types of data as in our study. In this chapter, it has been proposed to make estimates of the gravity equation according to the problems we have raised in the theoretical part of this work.

One of the most common problems is multilateral resistance, which can be replaced by unilateral individual effects or remoteness indices. Our estimation results proved the effectiveness of these indices in the analysis of the gravity equation. The solution of the trade flow problem that we used for the empirical estimation of the model are the bilateral effects (fixed or random), the PPML method, and the Heckman selection procedure. However, the PPML method and the Heckman selection procedure remain the most efficient because they give us unbiased estimators and relatively high explanatory power. From the point of view of estimation by cointegration techniques, the results are similar to PPML estimation and the Heckman selection procedure.

However, it should be noted that in the context of a panel with a relatively short period, the non-stationarity of the variables does not have a very large impact on the estimates of the gravity model, so that the problem of stationarity in a panel is real and is less important than time series.

CONCLUSION

In this dissertation on the gravity model, we had to examine the theoretical foundations of the gravitational equation, the specification, and the different econometric approaches. On the theoretical side, the gravity model has solid foundations that are mostly based on the theory of international trade. Then our main objective in the framework of this thesis was to study the gravity model by different econometric approaches in order to bring a robust analysis of our estimation results. However, we made different estimates from a specification in the sense of Anderson and Van Wincoop (2003). Special attention has been given to the problems of measuring multilateral resistance, zeros of trade flows, stationarity, and cointegration.

We had to consider several issues on the gravitational equation. First, in the chapter, we presented the theoretical foundations of Anderson (1979), Eaton and Kortum (2002) and Anderson and Van Wincoop (2003). The work of Anderson and Van Wincoop has revolutionized the theoretical underpinnings of the gravity equation and has extended the model of Eaton and Kortum (2002) to evaluate trade policies and determine partial and general equilibrium. Our study on this thesis was based on the homogeneity of the firms. This is why we used a specification in the sense of Anderson and Van Wincoop. However, we had to present a brief overview of the gravity model with the heterogeneity of firms in order to make a theoretical comparison of the gravity model and the prospects that we can achieve.

Then, we determined the set of econometric problems associated with the specification of the gravity model. The OLS method remains valid for estimating the gravity model if and only if there are no null trade flows in the data and heterogeneity problem. Otherwise, the GLS method overcomes the heteroscedasticity problem and the Maximum Likelihood Pseudo method based on an exponential family law such as the Poisson law that is most commonly used or the Heckman selection procedure remains. Very effective for the treatment of zero trade flows. In the case of non-stationarity of the gravitational and cointegration variables, the standard methods that we have just mentioned remain valid if and only if our study panel is centered on a short period in which we can not evaluate a considerable evolution of trade flows and GDP. If necessary, the gravity equation can be estimated by cointegration techniques such as DOLS and FMOLS.

Finally, in the third chapter, we performed a series of estimation of the gravity equation. Overall the results were just in line with our expectations. Multilateral resistance can be replaced by fixed unilateral individual effects or by the index of remoteness. The treatment of zero trade flows by a two-way fixed effect is not the best solution because the estimates may be biased due to omissions of some variables and loss of information. The use of a short period in our panel allowed us to compare the results with standard techniques and those with cointegration techniques. Virtually the effects of the variables remain identical and significant.

REFERENCES

Anderson. J. E (1979), a Theoretical Foundation for the Gravity Equation, American Economic Review, 69.

Anderson. J. E (2011), the Gravity Model, Annual Review of Economics, Annual Reviews, vol. 3(1), pages 133-160.

Anderson. J. E (2016), the Gravity Model of Economic Interaction, Boston College and NBER, August 17, 2016.

Anderson. J. E and Van Wincoop. E (2003), Gravity with Gravitas: A Solution to the Border Puzzle, College of Boston, 1-37.

Baldwin. R (2005), Heterogeneous Firm and Trade: Testable and Untestable Proprieties of the Melitz Model, NBER Working Papers 11471, National Bureau of Economic Research, Inc.

Baltagi. B. H, Egger. P and Pfaffermayr. M (2003), A Generalized Design for Bilateral Trade Flow Models, Economic Letters, Vol.80 (3), pp.391-397.

Bergstrand. J. H (1989), the Generalized Gravity Equation, Monopolistic Competition, and the Factor Proportions, Theory in International Trade, the Review of Economics and Statistics, 71.

Deardorff. A (1998), Determinants of Bilateral Trade: Does Gravity work in a neoclassical world? , NBER working paper, w5377.

Evenett. S. J and Keller. W (2002), on Theories explaining the Success of the Gravity Equation, Journal of political economy, University of Chicago Press, vol 110(2), page 281, page 316 April.

Frankel. J. and Wei. S. (1993): Trade Blocs and Currency Blocs, NBER Working Paper, 4335, National Bureau of Economics Research.

Gomez. H. E (2013), Comparing Alternative methods to estimate gravity models of bilateral trade, University of Canada, Campus de la Cartuja s/n, 18071 Granada, Spain, pp. 1-24.

Gomez. H. E and Tamarit. C (2010), the Euro effect on trade: evidence in gravity equations using panel cointégration techniques, pp.1-58.

Gomez. H. E, Tamarit. C and Camarero. M (2013), the euro impact on trade. Long-run evidence with structural breaks, pp. 1-30.

Haveman. J and Hummels. D (2004), Alternative Hypothesis and the Exten of Specializations, Canadian Journal of Economics, 37 (1), 199-218.

Head and Mayer (2000), Non-Europe: The magnitude and causes of market fragmentation in the E.U, Weltwirtschaftliches Archiv, vol 136, n°2, pp.109-129.

Helpman. E (1987), Imperfect Competition and International Trade: Evidence from Fourteen Industrial Countries, Journal of the Japanese and International Economics, vol.1 (1), pp.62-81.

Helpman. E and Krugman. P (1985): Market Structure and Foreign Trade, Increasing Returns, Imperfect Competition, and International Economy, the MIT Press, Cambridge, MA.

Helpman. E, Melitz. M. and Rubinstein. Y (2008), Estimating trade flows: trading partners and trading volumes, Quarterly Journal of Economics, vol.123, pp.441-487.

Hurlin. C et Mignon. V (2006), Une synthèse des tests de Cointégration sur données de panel, pp.1-33.

Linder. G. J. M and De Groot. H (2006), Estimation Gravity Equation in the presence of zero flows, Tinbergen Institute Discussion papers 06-072/3.

Markusen, James. R. and Randall. M. Wigle (1990), Explaining the Volume of North-South Trade, Economic Journal 100, 1206-1215.

McCallum. J (1995), National Borders Matter: Canada-USA, Regional Trade Patterns, American Economic Review, (3), 615-623.

Melitz. J (2007), North, South, and Distance in the Gravity Model, European Economic Review, 51, 971-991.

Mundell. R. A (1957): International Trade and Factor Mobility, the American Economic Review, June, pp.321-335.

Novy. D (2011), Gravity Redux: Measuring International Trade Costs with Panel Data, University of Warwick, July 2011, pp.1-27.

Novy. D (2012), International Trade without CES: Estimating Translog Gravity, University of Warwick, Department of Economics, November 2012.pp.1-34.

Novy. D (2015), International trade, Monopolistic Competition and Gravity, University of Warwick, Department of Economics, April 2012.

Obstfeld. M. and Rogoff. K (2000): The six major puzzles in international macroeconomics is there a common cause? , NBER Working Paper 7777.

Park. R (1966): Estimation with heteroscedastic error terms, Econometrica, 34(4).

Pesaran. H (2004), General Diagnostic Tests for Cross Section Dependence in Panels, University of Cambridge, USC and IZA Bonn, August 2004, Discussion Paper N°1240.

Pöyhönen. P (1963): a Tentative Model for the flow Trade between countries, Weltwirtschaftliches Archiv, Band 90, Heft I.

Rose. A (2000): One Money one Market, Estimating the effect of common currency Trade Economics Policy, 30.

Shawn. T (2012), Structural Estimation of a flexible Translog Gravity Model, University of Melbourne, Job Market Paper, 7 November, pp.1-42.

Silva. J. M. C and Tenreyro. S (2006), the log of Gravity, the Review of Economics and Statistics, November 2006, 88(4): 641-658.

Tinbergen. J (1962): Shaping the World Economy, suggestions for an international Economic Policy, the Twentieth-century Fund, New York.

Wei. S. H (1996): Intra-National versus International trade, how stubborn are nations in global integration, NBER w.p 5531.

Wooldridge. J. M (2002): Econometric Analysis of cross-section and panel data, Cambridge and London, MIT Press.

Yotov. V. Y, Piermartini. R, Monteiro. J. A and Larch. M (2016), an advanced Guide to Trade Policy Analysis: The structural Gravity Model, World Trade Organization, United Nations

APPENDIX

A. Pesaran Cross-Section Dependence Test (2004)

The Pesaran CD test (2004) is a cross-section dependence test, which is applicable to the panel model with homogeneity and heterogeneity of the parameters. This test can be done before or after the estimation of the panel model.

Suppose the following panel model:

$$y_{it} = \alpha_i + \beta_i' x_{it} + \varepsilon_{it}, pour\ i = 1,2, \dots, N\ et\ t = 1,2, \dots, T$$

i denotes the individual dimension and t denotes the temporal dimension. This test is correctly on a number of observations N and a fixed period T. This regression may contain delayed variables. It is further assumed that the error terms are centered, independent and identically distributed:

$$\varepsilon_{it} \sim IID(0, \sigma_\varepsilon^2)$$

The Pesaran test is no longer necessary to determine the relationship between the time dimension and cross-section and has simple properties that do not depend on a particular matrix weight; That is, the size of the individual dimension is large and that of the temporal dimension is small. So if N is big, the Breush-Pagan LM test can be adapted to this kind of problem. All the same, Pesaran (2004) proposes a simple alternative to the LM test, based on the correlation coefficient between two individuals i and j instead of using the square of the same correlation coefficient as in the LM test. The assumptions of this test are as follows:

$$\begin{cases} \boldsymbol{H_0}: Cross-section\ independence \\ \boldsymbol{H_a}: \quad Cross-section\ dependence \end{cases}$$

The Pesaran CD test (2004) is given by the following relation:

$$CD = \sqrt{\frac{2T}{N(N-1)}} \left(\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \widehat{\rho_{ij}} \right)$$

With :

$$\widehat{\varepsilon_{it}} = y_{it} - \widehat{\beta_i}' x_{it} = e_{it}\ et\ \widehat{\rho_{ij}} = \widehat{\rho_{ji}} = \frac{\sum_{t=1}^{T} e_{it} e_{jt}}{\left(\sum_{t=1}^{T} e_{it}^2\right)^{1/2} \left(\sum_{t=1}^{T} e_{jt}^2\right)^{1/2}}$$

Unlike the LM statistic, the CD statistic has exactly zero mean for fixed N and T values. To prove this statement, this author has developed the following hypotheses as part of a standard panel model:

H1: For all i, the ε_{it} errors are serially independent with a mean equal to 0 and a variance equal to σ_i^2, such that $0 < \sigma_i < \infty$;

H2: Under the null hypothesis also defined by:

$H_0: u_{it} = \sigma_i \varepsilon_{it}\ avec\ \varepsilon_{it} \sim IID(0,1)\ for\ all\ i\ and\ t$, the error ε_{it} are symmetrically distributed around 0;

H3: The explanatory variables are strictly exogenous and the matrix containing all these variables is positive definite;

H4: $T > k + 1$ and the residuals of the OLS are equal to e_{it}

Theorem: For every $N \to \infty\ and\ T > k + 1$ and under the hypothesis 1 to 4 we have :

$$\begin{cases} E(\widehat{\rho_{ij}}) = 0 \\ E(CD) = 0 \end{cases}$$

$$CD = \sqrt{\frac{2T}{N(N-1)}} \left(\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \widehat{\rho_{ij}} \right) \sim N(0,1)$$

B. Levin, Lin et Chu' Unit Root Test in Panel Data (LLC, 1999)

Panel unit root tests were developed through unit time root tests in time series. The major difference is that in a panel, we consider an asymptotic behavior of the temporal dimension T and the individual dimension N.

The Levin, Lin, and Chu test noted by LLC is one of the unit root tests in panel. This LLC test is based on the following assumptions:

$$\begin{cases} \boldsymbol{H_0}: Panel\ data\ have\ unit\ root \\ \boldsymbol{H_a}: Panel\ data\ are\ stationary \end{cases}$$

Suppose we have the following panel model:

$$\Delta y_{it} = \alpha + \theta y_{it-1} + \sum_{j=1}^{J} \gamma_{ij} \Delta y_{it-j} + \sum_{k=1}^{K} \beta_{kit} x_{kit} + \varepsilon_{it}, pour\ i = 1,2,\dots,N\ et\ t = 1,2,\dots,T$$

The LLC test is performed in three steps:

1st step: Perform estimation of the ADF statistic and determine the following auxiliary equations:

$$\widehat{u_{it}} = \Delta y_{it} - \sum_{j=1}^{J} \hat{\gamma}_{ij} \Delta y_{it-j} - \sum_{k=1}^{K} \hat{\beta}_{kit} x_{kit}\ ;\ \widehat{v_{it-1}} = y_{it-1} - \sum_{j=1}^{J} \tilde{\gamma}_{ij} \Delta y_{it-j} - \sum_{k=1}^{K} \tilde{\beta}_{kit} x_{kit}$$

2nd step: Standardize residues obtained by the standard deviation of ADF regression noted φ_i.

$$\Delta \tilde{y}_{it} = \frac{\hat{u}_{it}}{\varphi_i}\ et\ \tilde{y}_{it-1} = \frac{\hat{v}_{it-1}}{\varphi_i}$$

3 rd step: Estimate the following regression and calculate the LLC test statistic :

$$\Delta \tilde{y}_{it} = \theta \tilde{y}_{it-1} + \varepsilon_{it}$$

The assumptions of the LLC test can be rewritten as follows:

$$\begin{cases} \boldsymbol{H_0}: \theta = 0 \\ \boldsymbol{H_a}: \theta < 0 \end{cases}$$

The value obtained must be compared to those of the LLC table. The number of delays in the ADF test must be equivalent for any individual on the panel. Exogenous variables must also be specified or the fixed effects must be included in the regression model.

C. Kao's Panel Cointegration test (1999)

Kao's (1999) tests are cointegration tests on panel data. These tests are based on the DF and ADF tests. The assumptions associated with these tests are:

$$\begin{cases} \boldsymbol{H_0}: \quad NO\ cointegration \\ \boldsymbol{H_a}: variables\ are\ coint\acute{e}gr\acute{e}es \end{cases}$$

The first test of Kao (1999) is the DF test. It uses the following individual effects panel model:

$$y_{it} = \alpha_i + \beta x_{it} + \varepsilon_{it}, i = 1,\dots,N\ et\ t = 1,\dots,T$$

Subsequently, he determines a dynamic relationship of the residuals of the panel model:

$$\hat{\varepsilon}_{it} = \rho \hat{\varepsilon}_{it-1} + u_{it}$$

The null hypothesis ρ = 1 is tested and the corresponding test statistic t_{ρ}. Kao (1999) proposes four DF statistic functions of $\hat{\rho}$ $and\ of$ t_{ρ} :

- $\boldsymbol{DF}_{\rho}^{*}$ and $\boldsymbol{DF}_{t}^{*}$ take into account an endogeneity between the innovations of the integrated first-order dependent variables and the error terms;

- $\boldsymbol{DF}_{\rho}$ and $\boldsymbol{DF}_{t}$ are based on strong exogeneity between dependent variables and errors.

The second test of Kao (1999) is an ADF test with a single statistic and whose basic regression is as follows:

$$\hat{\varepsilon}_{it} = \rho\hat{\varepsilon}_{it-1} + \sum_{j=1}^{J} \theta_j \Delta\hat{\varepsilon}_{it-j} + u_{itJ}$$