2
Abril
2021

RPC
Radar da Política Comercial

Text for discussion nº02/2021

International trade and the gravity model: recent evidence in theoretical and empirical analysis

Carolina Rodrigues Corrêa Ferreira

ufjf | CAMPUS GV

About the extension project:

The Commercial Policy Radar (RPC) is an extension project of the UFJF economics department on the Governador Valadares campus and aims to periodically monitor the adoption of non-tariff measures on international trade by member countries of the World Trade Organization (WTO). The purpose is to facilitate the knowledge and understanding of the measures by Governador Valadares exporters and importers and from all over Brazil, allowing them to adapt to them.

The project also aims to monitor and analyze the trade balance in Valadares, aiming to give greater prominence to international trade in the region, highlighting the main exported and imported products, opportunities for expansion and possible shortcomings in the sector.

Finally, the project aims to bring information and training to entrepreneurs participating in international trade, as well as to those who wish to enter this market.

URL: https://www.ufjf.br/rpcgv.

Suggestion for quote:

FERREIRA, C. R. C. **International trade and the gravity model: recent evidence in theoretical and empirical analysis**. Governador Valadares: UFJF campus GV, mar. 2021. (Text for discussion nº02/2021). Available in: <https://www.ufjf.br/rpcgv/publicacoes/>.

International trade and the gravity model: recent evidence in theoretical and empirical analysis

Carolina Rodrigues Corrêa Ferreira[1]

Abstract: Despite the solid theoretical foundations and the notable empirical success of the gravity trade model, it was/is often used without foundation and taking into account the econometric challenges that can generate biased and inconsistent estimates. Therefore, this work presents a broad theoretical review of the model and demonstrates its application in the recommended way by the specialized literature. The main recommendations for a robust, consistent and unbiased analysis of gravity models were synthesized and the following stand out: 1) Intranational trade data should be included; 2) Directional fixed effects of temporal variation (country-year or country-sector-year) must be added in the equation; 3) Country pairs fixed effects should also be adopted, and; 4) The Poisson Pseudo Maximum Likelihood (PPML) estimator ought to be used. Then, a regression was estimated – using PPMLHDFE command in Stata 16 – to exemplify, verifying the impact of nations' economic freedom on their exports. The results showed that there is a positive, significant and elastic relationship between economic freedom and international trade. Finally, it is expected that the present work will serve as an instrument for several researches in the area, generating new evidence and directing the formulation of more efficient commercial policies.

Keywords: gravitational model; international trade; theoretical foundation.
JEL: F13; F14.

[1] PhD in Applied Economics from PPGEA / DER / UFV. Professor in the economics department at the Federal University of Juiz de Fora, Governador Valadares campus. Researcher at Econúcleo - Socioeconomic Studies. E-mail: carolina.correa@ufjf.edu.br.

1. Initial considerations

The gravity model has emerged as a simple and efficient approach for analyzing and forecasting economic variables related to bilateral trade flows. Tinbergen (1962) was the first to formally use the model in an initial version, in which the bilateral trade from the origin country to the destination is explained by the nations' economic masses, represented by their income, and by the geographical distance between them. This proposition makes sense, in intuitive terms, since richer countries produce more and have more income, therefore import and export more, and a greater distance represents higher transport costs, which tends to reduce trade.

According to Deardorff (1998), although it was only intuitive at first, the gravitational model gained wide use due to the rigorous theoretical foundation it received later and its empirical success in forecasting bilateral trade flows of different commodities in different situations.

Considering that goods are differentiated by place of origin and consumer preferences are homothetic, identical between nations and approximated by a CES-like utility function, Anderson (1979) derived the first theoretical foundations of economic gravity, later polished by Anderson and van Wincoop (2003). Then came the popularly used equation, simplified here:

$$lnX_{ij} = \alpha + \delta_1 lnY_i + \delta_2 lnY_j + \delta_3 \sum_{m=1}^{M} lnZ_{mij} + \mu_{ij} \quad (1)$$

Where Xij represents the trade flow from country i to j, α is the gravitational constant, Yi and Yj are the incomes of origin and destination countries, respectively, and Zmij a set of M observable variables that affects trade costs (for example, distance between countries, tariffs and non-tariff measures, trade agreements, among others).

The model gained greater statistical quality with the inclusion of fixed effects to control multilateral resistance terms, as indicated by Anderson and van Winccop (2003). Such terms represent the effect that the position of exporting and importing countries in the global market and their economic situation has on their own bilateral trade. Piermartini and Yotov (2016) explain that multilateral resistances translate the initial effects of partial equilibrium of the conjuncture at the bilateral level into specific effects of each country on consumer and producer prices.

Works such as those by Alhassan and Payaslioglu (2020), Gopinath et al. (2020) and Kox and Rojas-Ramagosa (2020) are examples of recent studies that used the gravity model for different purposes.

Alhassan and Payaslioglu (2020) carried out a comparative analysis of the impact of political and economic institutions on bilateral trade in emerging and low-income African countries. Using the gravity model, the study revealed that such institutions are key determinants of bilateral trade in Africa. The results showed that the effects differ by type of institutions and income levels of countries, and that institutions have a greater impact on trade in emerging African economies (EEA) than for low-income countries (PBR). Relatively, economic institutions affect trade more than political institutions in PBRs, while the reverse occurs in EEA. Therefore, the authors conclude that the design of an institutional framework that aims to facilitate bilateral trade must consider the peculiarities of countries in terms of their level of revenue and the different facets of the institutions.

Gopinath et al. (2020) applied the gravity model to verify the effect of machine learning[2] (ML) techniques on trade flows. Focusing on agricultural commodities, the study used data from supervised ML (data from 2010 to 2014, projection for 2011 to 2016) and unsupervised (data from 2014, projection from 2014 to 2020) to decipher trade patterns. The results show the high efficiency of ML models combined with the gravity model to predict short and long term trade patterns in relation to traditional approaches. While supervised ML techniques showed the main economic factors that influence agricultural trade flows, unsupervised approaches provided better long-term adjustments.

Kox and Rojas-Ramagosa (2020) used the gravity model to analyze the impact of preferential trade agreements (APCs), bilateral investment treaties

[2] Term linked to artificial intelligence, a technique that encompasses the idea of machines with the ability to learn alone from large volumes of data. Supervised algorithms are those in which the human being needs to interact by controlling the output and input of data and interferes with the training of the machine. Finally, the machine applies what has been learned in its algorithm for the next analysis. In the unsupervised category, the algorithms use deep learning to process complex tasks without human training (RABELO, 2019).

(TBIs) and other policies on the flow of bilateral foreign direct investment (FDI). The results showed that, on average, the formalization of an APC increases bilateral FDI stocks by about 30%. In addition, belonging to the European Union has a great impact and increases FDI by around 135%. Finally, it was found that the adoption of a TBI has an effect comparable to the signing of an APC, causing a positive effect on the flow of FDI.

In addition to these, several other works have used the gravitational model over the years, with the most diverse approaches. Piermartini and Yotov (2016) point out that, despite the solid theoretical foundations and the remarkable empirical success, the gravity model was and still is frequently applied without theoretical basis and without taking into account the econometric challenges that can lead to biased and inconsistent estimates.

Therefore, the present work aims, first, to make a theoretical review about the model and its microeconomic, mathematical and statistical foundation and, later, to demonstrate its application in the way recommended by the most recent specialized literature. Thus, we seek to contribute to future researches in international trade field, mainly in the assessment of trade policies and integration (preferential trade agreements), raising the quality of estimates and suggestions for policy makers.

In addition to this introduction, a broad review of the theoretical and econometric aspects related to the model is the subject of the next section, followed by an application of the method with empirical data and, finally, the conclusion.

2. Theoretical aspects

The gravity model is a tool widely used in several empirical fields and has a large number of applications in international trade studies. Its popularity is based on three pillars: first, trade flows are a key element in all types of economic relations; second, the data needed to estimate it are easily accessible to everyone today; third, a large number of high-standard works brought greater respectability to the gravity model (BALDWIN; TAGLIONI, 2006).

As previously mentioned, the theoretical basis for the gravity model was developed by Anderson (1979), who formulated the equation based on preferences with constant elasticity of substitution (CES) and differentiation of goods by region of origin. Following this basis, Anderson and van Wincoop (2003) developed a more advanced form of the theoretical model, which will be presented below.

It is assumed that each region is specialized in the production of a single good and that the quantity offered is fixed. Preferences are homothetic and represented by a CES function. Thus, consumers in region j maximize consumption (c) of goods from region i:

$$\left(\sum_i \beta_i^{\frac{1}{\sigma}} {c_{ij}}^{\frac{(\sigma-1)}{\sigma}}\right)^{\frac{\sigma}{(1-\sigma)}} \tag{2}$$

Subject to:

$$\sum_i p_{ij} c_{ij} = y_{ij} \tag{3}$$

where βi is a positive distribution parameter; σ is the elasticity of substitution between all goods; yj is the region's income and p is the price of goods in region i for consumers in region j. Prices differ between regions due to trade costs (transport, tariff and non-tariff barriers).

Considering pi the price of the good for exporter i free of trade costs and tij the factor of the cost of trade between i and j, we have that $p_{ij} = t_{ij}p_i$. Each good sent from i to j is assumed to incur a brick cost of $t_{ij} - 1$. This cost is absorbed by the importer. The amount paid by j to i for imports is $x_{ij} = p_{ij}c_{ij}$ and the costs are $(t_{ij} - 1)p_{ij}c_{ij}$. Therefore, the total income obtained by exporting region i is $y_i = \sum_j x_{ij}$.

The total demand for goods from i comes from j is given by maximizing (3) subject to (4):

$$x_{ij} = \left(\frac{\beta_i p_i t_{ij}}{P_j}\right)^{(1-\sigma)} y_j \tag{4}$$

where Pj is the consumer price index in j, given by:

$$P_j = \left[\sum_i \left(\beta_i p_i t_{ij}\right)^{1-\sigma}\right]^{\frac{1}{(1-\sigma)}} \tag{5}$$

The general equilibrium structure of the model imposes a free market, which implies:

$$y_i = \sum_j x_{ij} = \sum_j \left(\frac{\beta_i p_i t_{ij}}{P_j}\right)^{(1-\sigma)} y_i, \text{ for all i.} \quad (6)$$

Assuming that trade costs are symmetric ($t_{ij} = t_{ji}$), the authors arrive at an implicit solution in which the gravity equation is:

$$x_{ij} = \frac{y_i y_j}{y^W}\left(\frac{t_{ij}}{P_i P_j}\right)^{1-\sigma} \quad (7)$$

where y^W is the world nominal income, defined by $y^W \equiv \sum_j y_j$.

The basic gravity model is the one presented in equation (8) and this is subject to:

$$P_j^{1-\sigma} = \sum_i P_i^{\sigma-1}\theta_i t_{ij}^{1-\sigma}, \text{ for all j,} \quad (8)$$

where $\theta_i \equiv \frac{y_i}{y^W}$, that is, the i-th region's world income share.

In addition, it is necessary to include the unobservable factors of transaction costs. t_{ij} s assumed to be a log-linear function of a set of m observable variables Z^m_{ij} that can represent such costs (for example, distance between countries, technical measures and trade agreements.). So, we have to:

$$t_{ij} = \prod_{m=1}^{M}\left(Z_{ij}^m\right)^{\rho_m} \quad (9)$$

where Z^m_{ij}= 1 if there are no barriers related to m and equal to 1 plus a tariff equivalent of the barrier, otherwise, and ρ is a parameter.

Finally, using what has already been presented and applying logarithm in (8), we have the basic gravity equation widely used:

$$lnx_{ij} = k + lny_i + lny_j - lny^W + (1-\sigma)\rho_m \sum_{m=1}^{M} lnZ_{ij}^m - (1-\sigma)lnP_i - (1-\sigma)lnP_j \quad (10)$$

where k is a constant and the other variables are the same as previously defined.

The econometric estimation of gravity models comes up against some obstacles, as highlighted by Piermartini and Yotov (2016). The first of them is the inclusion of the terms of multilateral resistance (TMR), since their concept is theoretical and they cannot be directly observed in a database. Baldwin and Taglioni (2006) point out that one of the solutions - which stands out for efficiency and popularity - to this issue was the inclusion of fixed effects (dummies) for each exporting country, importing country and year of the sample[3].

More recently, Olivero and Yotov (2012) demonstrated that the TMR should be included as fixed effects of the exporter's time (and the importer's time) together, allowing the dynamic severity estimate with panel data. This approach, by eliminating countries' incomes as explanatory variables due to the collinearity with the fixed country-year effects, also resolves the issue of endogeneity between exports and gross domestic product (GDP). It should be noted that the procedure will also absorb all other country-specific, observable and unobservable characteristics, including various non-discriminatory trade policies, institutional variables and exchange rates[4].

The second challenge pointed out by Piermartini and Yotov (2016) is the presence of trade values equal to zero, which makes the use of linear estimators not recommended because it generates sample selection bias when log-linearizing the function. The zero trade flow is very important for the estimates as it affects the terms of multilateral resistance. The more disaggregated the trade data, the greater the number of zeros in the sample.

It is important to note that, for a consistent and efficient severity model, it is recommended to include all possible country pairs in the sample, thus avoiding sample selection bias. That is why dealing with zero trade is so important (BACCHETTA ET AL, 2012).

Several solutions, not very efficient, to this question were presented over the years until, finally, Silva and Tenreyro (2006) offered a simple and effective resource, in addition to being the most statistically efficient[5]: Estimate the model using a nonlinear estimator adapted from the Poisson estimator, named by the referred authors of Poisson Pseudo Maximum Likelihood (PPML)[6].

The referred method still presents the solution to the third obstacle presented by Piermartini and Yotov (2016), heteroscedasticity in the data of

[3] Procedure developed by (HUMMELS, 2001) and (FEENSTRA, 2004).
[4] There is a method to avoid the absorption of specific characteristics of the country, it will be addressed in the sixth obstacle.
[5] (SANTOS SILVA; TENREYRO, 2011).
[6] For more details see (SANTOS SILVA E TENREYRO, 2006).

commercial flows. This is a serious problem because, as pointed out by Santos Silva and Tenreyro (2006), in the presence of heteroscedasticity, estimates of the effects of commercial costs are skewed and inconsistent when the severity is estimated by ordinary least squares (OLS). To this end, the answer is, once again, to use PPML which also corrects unobservable heteroscedasticity.

The Poisson estimator is known as the standard approach for modeling discrete data. However, it has been gaining popularity as a viable alternative for estimating multiplicative models where the dependent variable is non-negative. Usually, these models are estimated by linear regression applied to a dependent variable transformed into a log. But, as in OLS, the only necessary assumption for the consistency of the Poisson estimator is the correct specification of the conditional average of the dependent variable (GOURIEROUX et al., 1984). Thus, the Poisson's estimator becomes the PPML estimator.

Correia et al. (2020) state that, in the presence of non-negative data with many zeros, PPML seems to be the safest bet. This situation is very likely to occur in many areas of research, especially when working with highly disaggregated data (for example, when modeling a company's R&D expenses, patent citation counts, daily sales of products in stores and bilateral trade).

The fourth challenge, according to Piermartini and Yotov (2016), is the specification of bilateral trade costs. The standard practice in the literature to represent the bilateral trade cost term, from specification (9), is to use a series of observable variables that have become standard covariables in the empirical specification of gravity (distance, contiguity, common language and trade agreements), according to Head and Mayer (2014). The use of fixed effects from country pairs also serves this purpose.

The fifth challenge is the endogeneity of trade policy, that is, the probable existence of bicausality between trade policy and the unobservable trade costs of cross-sections. The most suitable method to solve the endogeneity problem would be the use of instrumental variables, however, given the lack of reliable instruments, it is not possible (PIERMARTINI; YOTOV, 2016). Therefore, Baier and Bergstrand (2009) advocate the use of fixed effects for country pairs or first difference to explain or eliminate, respectively, the unobservable links between trade policies and the term of error in severity estimates. It should be noted that the inclusion of fixed effects for country pairs absorbs any time-invariant bilateral variable (distance, for example). Despite this, this is the most widely used method.

The estimation of gravity models with non-discriminatory trade policies (that is, which are adopted by countries for any trading partner) is the sixth challenge highlighted by the authors. The same goes for any variables that are unilateral, that is, they vary only in the country of origin or destination over time. This is because the fixed country-year effects absorb them in the estimation. One of the simplest solutions was proposed by Heid, Larch and Yotov (2017): the inclusion of intranational commerce in the sample. As these non-discriminatory policies do not affect domestic trade, the collinearity between the variables ceases to exist, preventing absorption.

The seventh challenge is to capture the time for natural adjustment of trade flows in response to changes in trade policies. For that, researchers like Baier and Bergstrand (2009) and Anderson and Yotov (2016) estimated models with time intervals of 3 to 5 years in the trade data, obtaining more credible estimates.

Estimating the gravity model with disaggregated data, to better capture the effect of policies that are applied differently for each good or sector (for example, tariffs), is the eighth and last challenge pointed out by Piermartini and Yotov (2016). As the CES function used in the microfoundation of the gravity model has the characteristic of divisibility, the estimation can occur with sectorial data, with fixed country-sector-year effects, as recommended by the authors.

Finally, Yotov et al (2016) summarize the main recommendations for efficient, robust and unbiased analysis of gravity models: 1) Whenever available, panel data should be used to obtain structural severity estimates; 2) Panel data with intervals (2, 3 or 5 years) should be used instead of data grouped by consecutive years; 3) Intranational trade data, constructed as the difference between the raw data of the production value[7] and total exports, must be

[7] Gross domestic product (GDP) data should not be used as they are in added value, while trade data are in gross values. See some recommended databases in (YOTOV et al., 2016).

included; 4) Directional fixed effects of temporal variation (country-year or country-sector-year) must be included in the equation; 5) Fixed effects of country pairs must also be adopted, and; 6) The Poisson Pseudo Maximum Likelihood (PPML) estimator should be used[8].

3. Application

3.1 Methodology

To exemplify the use of the gravity model applied to the analysis of international trade flows, the impact of nations' economic freedom on their exports was assessed. Depken and Sonora (2005) claim that the international trade in goods naturally appears to be based on some levels of economic freedom. The lack of this tends to correlate with limited access to foreign products, probably to the benefit of those who have power.

Bilateral export data from all countries and independent economies in the world between 2005 and 2016 were used, with respect to trade in goods divided into 3 sectors (agriculture, mining and manufacturing), as well as an indicator of economic freedom.

The bilateral trade flow International Trade and Production Database for Estimation (ITPD-E), developed by Borchert et al. (2020), which contains consistent data on international and intranational trade at the industry level, covering agriculture, mining, energy, manufacturing and services. The base covers 243 countries and 170 sectors[9]. For the present analysis, the services sector was excluded because it had a very different configuration from trade in goods, the other sectors were condensed into agriculture, mining and manufacturing.

As an explanatory variable, the Index of Economic Freedom (IEF) was adopted, which seeks to analyze, in a society, how free individuals are to work,

---

[8] Correia, Guimarães and Zylkin (2020) developed a faster and more efficient command for estimating with PPML in the stata software, ppmlhdfe. The original command developed by Santos Silva and Tenreyro (2006) has problems in the presence of large fixed effects, often making it impossible to include all the necessary fixed effects. With ppmlhdef this was resolved.

[9] See the list of countries and sectors at https://usitc.gov/publications/332/working_papers/itpd-e_usitc_wp.pdf.

produce, consume and invest in any way they want. Economic freedom is measured based on 12 quantitative and qualitative factors, grouped into four broad categories: 1) Rule of Law (property rights, government integrity, judicial effectiveness); 2) Government size (government spending, tax burden, fiscal health); 3) Regulatory efficiency (commercial freedom, freedom of work, monetary freedom), and; 4) Open markets (commercial freedom, investment freedom, financial freedom). Each of the twelve economic freedoms within these categories is graded on a scale of 0 to 100. A country's overall score is derived from the average of these twelve economic freedoms, with equal weight being given to each (THE HERITAGE FOUNDATION, 2020).

The sample covers the years 2005, 2009, 2013 and 2016, using intervals of 3 years (except in the last year of the sample, as it is the last available) as recommended by Baier and Bergstrand (2009), to allow the necessary adjustment after changes in policies.

Thus, the estimated model was:

$$X_{mijt} = \alpha + \beta_{\text{imt}} + \Omega_{\text{jmt}} + \pi_{ij} + \delta_1 \ln IEF_{it} + \mu \qquad (11)$$

where Xijt are exports from country i to j, sector m (agriculture, mining and manufactures), in year t; α is the gravitational constant; β are the exporter-sector-year fixed effects; Ω are the import-sector-year fixed effects; π are the fixed effects of country pairs; IEFit is the Index of Economic Freedom for country i in year t, e; μ is the error term.

The estimation[10] was performed by Poisson Pseudo Maximum Likelihood (PPML) as recommended by Yotov et al. (2016), however, using the command developed by Correia et al. (2020) for STATA software, PPMLHDFE, which is more efficient in the presence of large fixed effects (as is the case with the gravity model, since it is recommended to use as many countries as possible to capture all the resistance multilateral). The command absorbs the fixed effects during the estimation, showing only the coefficients of the explanatory variables of interest in the outputs.

[10] Appendix A contains, step by step, the commands used to perform the estimation in the STATA 16 MP software.

3.2 Results

Table 1 below shows the model results, coefficients and standard errors, estimated by Poisson Pseudo Maximum Likehood (PPML), using the ppmlhdfe command from the Stata software.

Table 1: Estimation results

Variable	Coefficient	Standard errors
Ln IEF	1.643002***	0.433281
Constant	7.632578***	0.922237
Pseudo R2	0.9896	
Nº of observations	274986	
FE exporter-sector-year	Yes	
FE importer-sector-year	Yes	
FE country pairs	Yes	

*** represents 1% significance. Standard errors are robust. EF = fixed effects.

Source: own elaboration.

As expected, the Index of Economic Freedom (IEF) showed a positive and significant relationship to exports. An increase of 10% in the indicator results, on average, in an increase of approximately 16.4% in the bilateral trade flow. Such elastic effect shows the great relevance of the economic freedom measured by the IEF for a better export performance. This finding is in line with works such as those by Depken and Sonora (2005) and Naanwaab and Diarrassouba (2013), which also found a positive relationship between economic freedom and international trade.

The constant, as Baldwin and Taglioni (2006) explain, is not a constant as in the physical world; it is what can be called "gravitational inconstant", since it includes all the costs of bilateral trade and multilateral resistance, therefore it varies with country and time. Therefore, although significant on average, the "inconstant" coefficient varies for each fixed effect included in the estimation, so it does not make sense to analyze it here.

McFadden's R2 (pseudo R2) shows a high adjustment, but this always occurs with the presence of multilateral resistance terms, which explain most of the trade. Therefore, it should not be used to assess the explanatory power of the equation, but to compare the quality of adjustment in relation to similar equations (VEALL; ZIMMERMANN, 1996).

The fixed effects necessary to control multilateral resistance and country pairs were inserted, as instructed by Yotov et al. (2016). Their coefficients were omitted because they were not of interest to the study, fulfilling only their statistical function.

Finally, it was possible to verify the application of the specified structural gravity model according to the most recent recommendations, as well as the quality of its results. The possibilities for empirical analysis with it are numerous, especially in the evaluation of trade policies and preferential trade agreements.

4. Conclusions

The gravity model, one of the most popular in the study of international trade, is a simple and efficient approach for analyzing and forecasting economic variables related to bilateral trade flows. Its popularity is based on three main facts: international trade flows are an important element for different types of economic relations; the data needed to estimate it are easily accessible to everyone today, and; a large number of high-standard works brought greater respectability to the gravity model, as well as its consolidated theoretical and statistical development.

However, in spite of solid theoretical foundations and remarkable empirical success, the gravity model was and still is often used without foundation and without taking into account the econometric challenges that can generate biased and inconsistent estimates.

Therefore, this work sought to make a broad theoretical review of the model and its econometric basis, as well as to demonstrate its application in the manner recommended by the specialized literature.

Based on a wide review of recent work, the main recommendations for efficient, robust and unbiased bias estimation were reported: 1) Whenever available, panel data should be used to obtain structural severity estimates; 2) Panel data with intervals (2, 3 or 5 years) should be used instead of data grouped by consecutive years; 3) Intranational trade data, constructed as the difference between the raw data of the value of production and total exports, must be included; 4) Directional fixed effects of temporal variation (country-year or country-sector-year) must be included in the equation; 5) Fixed effects of country pairs must also be adopted, and; 6) The Poisson Pseudo Maximum Likelihood (PPML) estimator must be used.

Finally, a gravity equation was estimated to exemplify, verifying the impact of nations' economic freedom on their exports. The results, as expected, demonstrated that there is a positive, significant and elastic relationship between economic freedom and international trade.

It is hoped that the present work will serve as an instrument for several researches in the area, generating new evidence and directing the formulation of more efficient commercial policies.

References

ALHASSAN, A.; PAYASLIOGLU, C. Institutions and bilateral trade in Africa: an application of Poisson's estimation with high-dimensional fixed effects to structural gravity model. **Applied Economics Letters**, v. 27, n. 16, p. 1357–1361, 19 set. 2020.

ANDERSON, J. E. A Theoretical Foundation for the Gravity Equation. **American Economic Review**, v. 69, n. 1, 1979.

ANDERSON, J. E. AND VAN WINCOOP, E. Gravity with Gravitas: A Solution to the Border Puzzle. **American Economic Review**, v. 93, n. 1, 2003.

ANDERSON, J. E.; YOTOV, Y. V. Terms of trade and global efficiency effects of free trade agreements, 1990-2002. **Journal of International Economics**, v. 99, p. 279–298, 1 mar. 2016.

BACCHETTA, MARC. ET AL. **A Practical Guide to Trade Policy Analysis**. [s.l.] WTO, 2012.

BAIER, S. L.; BERGSTRAND, J. H. Estimating the effects of free trade agreements on international trade flows using matching econometrics. **Journal of International Economics**, v. 77, n. 1, p. 63–76, 2009.

BALDWIN, R.; TAGLIONI, D. **Gravity for Dummies and Dummies for Gravity Equations**: NBER WORKING PAPER SERIES. [s.l: s.n.]. Disponível em: <http://www.nber.org/papers/w12516>. Acesso em: 26 out. 2020.

BORCHERT, I., LARCH, M., SHIKHER, S., AND YOTOV, Y. **The International Trade and Production Database for Estimation (ITPD-E)**. Disponível em: <https://www.usitc.gov/data/gravity/itpde.htm>. Acesso em: 3 nov. 2020.

CORREIA, S.; GUIMARÃES, P.; ZYLKIN, T. Z. Fast Poisson estimation with high-dimensional fixed effects. **Stata Journal**, v. 20, n. 1, p. 95–115, 2020.

DEARDORFF, A. V. Determinants of Bilateral Trade: Does Gravity Work in a Neoclassical World? In: JEFFREY A. FRANKEL (Ed.). . **The Regionalization of the World Economy**. [s.l.] University of Chicago Press, 1998.

DEPKEN, C. A.; SONORA, R. J. Asymmetric Effects of Economic Freedom on International Trade Flows. **International Journal of Business and Economics**, v. 4, n. 2, p. 141–155, 2005.

FEENSTRA, R. C. **Advanced International Trade: Theory and Evidence**. [s.l.] Princeton university press, 2004.

GOPINATH, M.; BATARSEH, F. A.; BECKMAN, J. **Machine Learning in gravity models: an application to agricultural trade**: NBER WORKING PAPER SERIES. Cambridge, MA: [s.n.]. Disponível em: <http://www.nber.org/papers/w27151>. Acesso em: 28 out. 2020.

GOURIEROUX, C.; MONFORT, A.; TROGNON, A. Pseudo Maximum Likelihood Methods: Theory. **Econometrica**, v. 52, n. 3, p. 681, maio 1984.

HEAD, K.; MAYER, T. Gravity Equations: Workhorse,Toolkit, and Cookbook. In: **Handbook of International Economics**. [s.l.] Elsevier B.V., 2014. v. 4p. 131–195.

HEID, B.; LARCH, M.; YOTOV, Y. **Estimating the Effects of Non-discriminatory Trade Policies within Structural Gravity ModelsDrexel University Working Paper Series**. [s.l: s.n.].

HUMMELS, D. L. **Toward a Geography of Trade Costs**. Purdue University. 54 p. 2001.

KOX, H. L. M.; ROJAS-ROMAGOSA, H. How trade and investment agreements affect bilateral foreign direct investment: Results from a structural gravity model. **World Economy**, n. March, p. 1–40, 2020.

NAANWAAB, C.; DIARRASSOUBA, M. The impact of economic freedom on bilateral trade: A cross-country analysis. **International Journal of Business Management and Economic Research**, v. 4, n. 1, p. 668–672, 2013.

OLIVERO, MARÍA PÍA; YOTOV, Y. Dynamic Gravity: Endogenous Country Size and Asset Accumulatio. **Canadian Journal of Economics**, v. 45, n. 1, 2012.

PIERMARTINI, R.; YOTOV, Y. V. **Estimating Trade Policy Effects with Structural GravityCESIFO WORKING PAPERS**: CESIFO WORKING PAPER. [s.l: s.n.]. Disponível em: <www.RePEc.org>. Acesso em: 26 out. 2020.

RABELO, A. **O que é machine learning? Entenda essa tendência no marketing digital!** Disponível em: <https://rockcontent.com/br/blog/machine-learning/>. Acesso em: 3 nov. 2020.

SANTOS SILVA, J M C E TENREYRO, S. The log of Gravity. **Review of Economics and Statistics**, v. 88, n. November, p. 641–658, 2006.

SANTOS SILVA, J. M. C.; TENREYRO, S. Further simulation evidence on the performance of the Poisson pseudo-maximum likelihood estimator ☆. **Economics Letters**, v. 112, p. 220–222, 2011.

THE HERITAGE FOUNDATION. **Index of Economic Freedom**. Disponível em: <https://www.heritage.org/index/>. Acesso em: 3 nov. 2020.

TINBERGEN, J. **Shaping the world economy: suggestions for an international economic policy**. [s.l.] New York: Twentieth Century Fund, 1962.

VEALL, M. R.; ZIMMERMANN, K. F. Pseudo-R2 measures for some common limited dependent variable models. **Journal of Economic Surveys**, v. 10, n. 3, p. 241–259, 1 set. 1996.

YOTOV, Y. V. et al. **An Advanced Guide to Trade Policy Analysis: The Structural Gravity Model**. [s.l.] UNCTAD, 2016.

Appendix

Appendix 1A: List of commands for estimation in STATA 16

1) Install the ppml command (if using)
ssc install ppml

2) Install the ppmlhdfe command
ssc install ppmlhdfe (you will be asked to install other commands when using)

3) Complete all country pairs for trade flow data
Open the file with the bilateral trade data.
Four columns are required: exporter (iso3 code), importer (iso3 code), year and trade flow (trade).
fillin exporter year importer
replace trade = 0 if trade ==.

4) Include intranational trade in bilateral trade data (data go in the trade column, along with international flow data)
For example, exporter = BRA, importer = BRA, year and trade = domestic production in gross values - gross exports (by sector / merchandise).
It is not recommended to use GDP data as it is measured in added value, and exports are gross.

5) Joining databases: the merge command

Attention to the label of the variables, it must be exactly the same in all bases. Country names too, so it's best to use the 3-letter iso code.

Open the trade flow file.

Use the direct merge command in the options menu (suggestion): *Data> combine datasets> merge two datasets*. Select the common variables of the bases. For example, in the case of the IEF, they were exporter and year.

6) Generate control dummies (if using the ppml command. For ppmlhdfe it is not necessary)

6.1 Country-year fixed effects

egen exporano = group (exporter year)

tabulate exporano, generate (exporano_)

Same for importer.

6.2 Fixed effects of country pairs

egen pair_id = group (importer exporter)

tabulate pair_id, generate (pairfe_)

7) Generate logarithmic variables

gen "new variable name" = ln ("original variable")

Do for any quantitative explanatory variable.

8) For intranational trade, make a dummy that must be multiplied by the explanatory variables (already in ln):

gen nointratrade = 1

replace nointratrade = 0 if exporter == importer

Then just generate new explanatory variables that will be the variable already in ln multiplied by the dummy nointratrade (for example, gen lnIEFnew = lnIEF * nointratrade).

9) If possible, use time intervals between the years of the sample.

10) Estimation by ppml

ppml trade pairfe_* exporano_* imporano_* "explanatory variables", cluster (pair_id)

For many countries, it takes a long time to complete and exclude a large part of the fixed effects due to collinearity.

The cluster () command corrects standard errors calculated based on country pairs. PPML estimation is already robust.

11) Estimation by ppmlhdfe

The estimator does not recognize the iso3 codes, so it is necessary to create a numerical label for each exporting and importing country.

egen exportadorid = group (exporter)

egen importadorid = group (importer)

ppmlhdfe *trade "explanatory variables", vce(cluster exportadorid#importadorid) absorb (exportadorid#year importadorid#year exportadorid#importadorid)*

Yotov et al. (2016) explains that, as the data used is a panel data set with repeated observations from pairs of countries over time, common observable and unobservable effects can arise naturally. Despite the control of fixed effects, some pattern of correlation between pairs of countries over time may still be present in the error term. This correlation pattern can be captured by grouping errors by pairs of countries, with the vce (cluster).