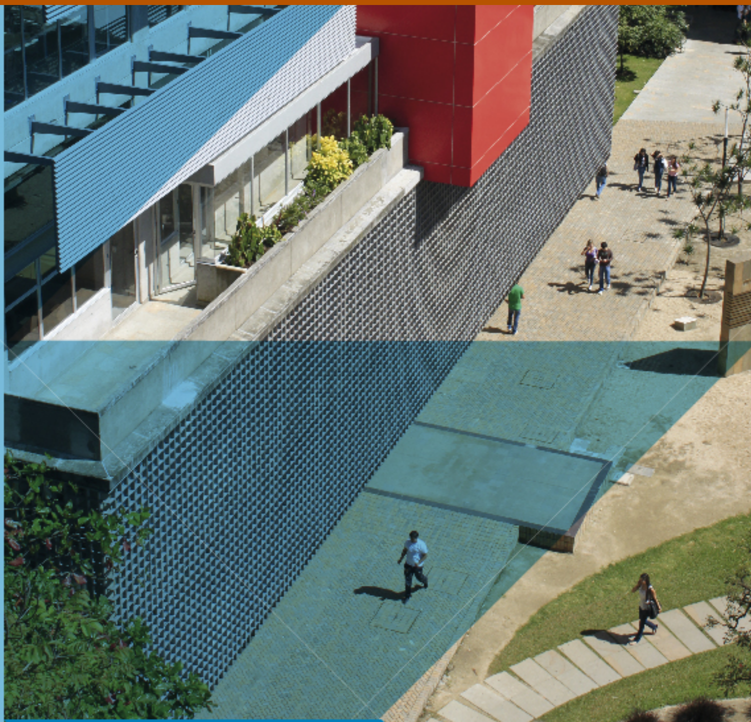

UNIVERSIDAD EAFIT
Abierta al mundo

Escuela de Economía y Finanzas

Documentos de trabajo

Economía y Finanzas

Centro de Investigación Económicas y Financieras

No. 13-12
2014

Solving the puzzle: A new measure of trade distance in the gravity equation

Mejía, Juan F.; Ramírez, Andrés

cief
Centro de Investigaciones Económicas y Financieras
Universidad EAFIT

Solving the Puzzle: A New Measure of Trade Distance In The Gravity Equation*

Juan Felipe Mejia, Ph.D in Economics†

Andrés Ramírez Hassan, Ph.D in Statistics‡

June 11, 2014

Abstract

The gravity model is a workhorse tool that has been widely used in international trade. However, one empirical question that frequently arises is related to the conceptualization and measurement of distance. To overcome this limitation, our study proposes an index of distance based on multivariate statistical analysis. Specifically, we build our index using Factorial Analysis for Mixed Data. For robustness check, we use Principal Component Analysis. Both techniques summarize in one factor information related to geographical, cultural, political and economic variables that might affect international trade between countries. We use this index as proxy of distance, and Gross Domestic Product as proxy of mass, and we run some panel data exercises between 1995 and 2000 for 10 Latin American economies. Estimations indicate that the sign of the load factors in Factor Analysis for Mixed Data are intuitively plausible, and that panel data exercises give sensible robust outcomes.

JEL Classification: F11, F14

Keywords: Factor Analysis for Mixed Data, Gravity Equation Model, Panel Data, Trade Distance.

*We thank Michael Jetter for fruitful comments and discussions. Also, we are grateful for excellent research assistance from David Alejandro Giraldo. All remaining errors are our own. Preliminary version, please do not cite without permission.

†Department of Economics, School of Economics and Finance, Universidad EAFIT. Carrera 49 7 Sur - 5 Medellín, Colombia. Phone: +57 4 2619500 Ext. 9354. Fax: +57 4 2664284. email: jfmejia@eafit.edu.co

‡Departament of Economics, School of Economics and Finance, Universidad EAFIT. Carrera 49 7 Sur - 5 Medellín, Colombia. Phone: +57 4 2619500 Ext. 9402. Fax: +57 4 2664284. email: aramir21@eafit.edu.co

Introduction

The gravity model of international trade is a workhorse tool that has been used in an ample range of empirical fields. In the literature, it has been highlighted how it has provided "some of the most robust empirical findings in economics" (Leamer and Levinsohn, 1995). Usually, when the gravity equation has been tested, the estimated effects of distance and output have shown to be economically and statistically significant and reasonably consistent across studies (Rose, 2004). Besides that, the gravity model has been able to explain most of the variation in international trade (Mejia, 2011).

Distance, a proxy for trading costs, is expected to negatively affect the flow of international trade between a pair of countries. Even though some studies based on the gravity model have applied direct measures of transport cost barriers to trade, the majority of them rely on distance as a proxy for transport costs (Brun et al., 2005). As noted by Huang (2007) there is, however, no consensus on what geographic distances are proxying for: the costs derived from distance may include varied components such as freight charges, cultural dissimilarities and other barriers which can be difficult to measure (Anderson and Marcouiller, 1999). Batra (2004), cited in Correia (2008), argues that distance can also be a proxy for the time elapsed during shipment, synchronization costs, transaction costs or cultural distance (Mejia, 2011).

In consequence, while distance has always been an important variable in gravity equations, authors have never been sure exactly what 'costs' distance represents (Baier and Bergstrand, 2001).

From an empirical perspective, two technical problems are present in a relevant number of studies. In the first place, many of the variables included in the model show correlation, leading to a loss of precision. In the second place, some of these variables are categorical, which implies that a measure of distance should consider this fact. Our main objective is to propose a measure of distance that involves geographical, cultural, social and economic aspects, and then use it in an econometric application of the gravity model for some selected Latin American countries. This measure brings statistical and economic advantages: an index of international trade distance in just one variable is defined, where some categorical variables are considered. Besides that, the precision of the parameters estimates in econometric exercises is improved, since we mitigate the multicollinearity problem.

To test the gravity model, we estimate the effects of the Gross Domestic Product and our trade distance factor on exports by using a panel data structure from 1995 to 2010. We apply Factor Analysis for Mixed Data to build a measure of distance where we involve geographical, cultural, social and cultural variables. Additionally, we use Principal Component Analysis for robustness check. Estimations indicate that the sign of the load factors in both multivariate statistical techniques are intuitively plausible, and that panel data estimations show sensible robust outcomes.

1 Literature Review

The gravity model is considered as a workhorse tool that has been used in an ample range of empirical fields, being the impact of trade agreements, exchange rate volatility, currency unions, Foreign Direct Investment (FDI) between countries, and the so-called "border effect", some of its common applications (Baldwin and Taglioni, 1999). It is a mathematical model derived from an analogy with Newton's gravitational law, used to explain aggregate human behaviors related to spatial interaction such as migration and traffic flows (World-Bank, 2002). Other investigations that have been conducted in the light of the gravity model range from the effect of foreign aid on FDI flows, to the effect of democracy, environmental regulations or corruption and insecurity on trade.

The gravity model appeared in the 1960s as an empirical specification with some theoretical foundations. As stated by Deardorff (1995), Tinbergen (1962) and Pöyhönen (1963) carried out the first econometric estimations of trade flows based on the gravity equation. Linnemann (1966) proposed a theoretical foundation based on the Walrasian general equilibrium system. He stated that the gravity model was a reduced form from a four-equation partial equilibrium model of export supply and import demand (Bergstrand, 1985). Deardorff (1995) emphasizes how "...Leamer and Stern (1970) followed Savage and Deutsch (1960) in deriving it from a probability model of transactions". Later, efforts were undertaken to derive the gravity equation from models that assumed product differentiation, starting with Anderson (1979). He made the so-called "Armington Assumption" - where products were differentiated by country of origin.

Three aspects have been crucial for the gravity model's recognition: its suitability for explaining international trade flows, the accessibility of the data needed for its estimation, and the respectability of a number of seminal papers that have established the gravity model's reputation and proposed a set of standard practices that are used to address different empirical questions.

In one of the most influential empirical papers dealing with the gravity equation, Rose (2000), notes that the gravity model of international trade "has a remarkably consistent history of success as an empirical tool". He also cites Leamer and Levinsohn (1995), who describe the gravity model as having provided "some of the clearest and most robust empirical findings in economics." It relates bilateral trade flows to GDP, distance, and other factors that affect trade barriers (Anderson and van Wincoop, 2003), modelling the flow of international trade between a pair of countries as "proportional to their economic mass and inversely proportional to the distance between them" Rose (2000, pp 7). Anderson (1979) states that the gravity equation is "probably the most successful empirical trade device of the last twenty-five years". Influential empirical studies dealing with the gravity equation include the works of Anderson and van Wincoop (2003), Feenstra et al. (2001), Evenett and Keller (2002), Eaton and Kortum (2002), Rose (2000, 2004), Soloaga and Winters (2001), and Subramanian and Wei (2007), among others.

2 Empirical Strategy and Results

We analyze the period 1995-2010, due to availability of information and more specifically, because the democracy index measured by Polity2 from the Polity IV data is only available from 1995 onwards. Our sample comprises five South American countries and five Central American countries, namely: Argentine, Brazil, Chile, Colombia, Ecuador, El Salvador, Guatemala, Honduras, Mexico and Panama. Table 5 shows the list of variables, their definition and sources.

Descriptive statistics can be found in table 6. When building the dataset, each country's export basket comprised –at least–70% of its exports. When analyzing the descriptive statistics, it is interesting to notice how the coverage for Chilean exports is of 70,48%, while in the Mexican case, it goes up to 97.62%. In our dataset, Honduras is the country with less trade partners (32), but 85.23% of its exports are represented. Mexico is the country with the highest level of exports, and Brazil is the country with the highest Gross Domestic Product in our dataset. Additionally, El Salvador is the country whose trade partners have the highest average GDPs, but also, the highest volatilities. Concerning the democracy index, it can be observed that countries like Chile and Panama obtain the maximum score (9), while Ecuador gets the minimum score (6,68).

To create our trade distance index we use geographical distance measured in kilometers, the bilateral exchange rate between trade partners, the absolute difference in the democracy index between trade partners, and three categorical variables: the existence of a common border, if the partners share a common language, and if the partner country is an island. Therefore, we try to capture geographical, economic, cultural and political aspects to build our distance index.

To calculate our distance index, we use the first dimension associated with the previous variables from a Factor Analysis for Mixed Data. This multivariate statistical technique is a principal component method to explore data with numerical and categorical variables, which can be considered a mix between Principal Component Analysis and Multiple Correspondence Analysis, that ensures to balance influence of both types of data (Escofier and Pagès, 1994; Pagès, 2004).

We perform our Factorial Analysis for Mixed Data in R package (R Development Core Team, 2013). Specifically, we use the library FactoMimeR (Husson et al., 2013). As can be observed in Table 1, in relation to numerical variables a longer distance, higher differences in real exchange rate and bigger political discrepancies imply a larger trade distance.[1] Concerning the categorical

[1]We perform Bartlett test of sphericity, and there is not statistical evidence to reject the null hypothesis that these variables are not intercorrelated.

variables, our analysis shows that sharing a border and sharing a common language imply less distance, while having an island as a trade partner implies a larger trade distance. One of the most relevant characteristics of Factor Analysis for Mixed Data is that one can obtain different load factors for each categorical variable. For instance, as can be observed in Table 1, there are asymmetric effects of these variables on our trade distance. In particular, sharing a land border has a higher effect on trade distance compared to not sharing a land border, these effects are -3.11 and 0.29 in average, respectively. The same pattern can be identified when two partners share a common language. To have an island as a trade partner implies a higher effect on trade distance compared with the opposite situation, these load factors are 1.91 and -0.29 in average, respectively.

We perform some Principal Components Analysis for each country with the same numerical variables, and treat the categorical variables as dummy variables which are equal to 1 if the condition is met, and zero otherwise. The main idea with these exercises is to perform some robustness checks of our results. As can be seen in Table 2, we obtain the same intuitive outcomes. Specifically, larger physical distance, higher exchange rate, bigger political discrepancies, not sharing a land border or language and having an island as trade partner imply a larger distance.

Once we build an index of trade distance, we use the product of the Gross Domestic Product as proxy of the mass between a host country and its trade partner.[2] Given that ignoring dynamics might lead to incorrect inference; this aspect is an important issue in international trade since the establishment of distribution and service networks between business partners implies lower barriers and sunk costs. Additionally, there is habit formation among consumers that incentives inertial demand (Bun and Klaassen, 2002). The main model is based on the dynamic panel data model proposed by Blundell and Bond (1998). This methodology implements some moment conditions that imply more efficient estimation (Cameron and Trivedi, 2005).

[2] All our panel data models were run on Stata 12.0 (StataCorp, 2011).

Table 1: Factorial Analysis for Mixed Data: An Index of Trade Distance

	Country									
Continuous Variables	Argentine	Brazil	Chile	Colombia	Ecuador	EL Salvador	Guatemala	Honduras	Mexico	Panama
Distance	0.9170	0.9050	0.9110	0.8560	0.9210	0.8810	0.8670	0.8390	0.8820	0.8330
Real Exchange Rate	0.2090	0.1160	0.1960	0.5770	0.0790	0.1890	0.2880	0.2900	0.2680	0.3990
Political Difference	0.3690	0.3630	0.3020	0.6220	0.4800	0.6250	0.5190	0.5500	0.5310	0.1320
Categorical Variables										
Border NO	0.5050	0.4790	0.2680	0.3540	0.1770	0.2150	0.3120	0.2480	0.1300	0.2150
Border YES	-3.1340	-2.6330	-3.6660	-2.2630	-3.1060	-3.4420	-3.4270	-3.7260	-2.4000	-3.3320
Island NO	-0.2650	-0.2790	-0.3170	-0.2480	-0.3220	-0.3130	-0.2380	-0.2480	-0.3020	-0.3830
Island YES	1.6400	1.8980	1.6780	2.0490	2.0600	2.3480	1.9040	1.3380	2.0550	2.1430
Common Language NO	0.8270	0.0160	0.8880	1.0420	0.9730	0.8720	0.9410	0.8960	0.8630	0.9910
Common Language YES	-2.1510	-0.6040	-2.1180	-1.9250	-1.7970	-1.8240	-1.8830	-1.9710	-1.7270	-1.7350
Proportion[a]	42.08%	31.80%	39.73%	44.79%	39.12%	40.88%	40.81%	39.45%	37.03%	38.20%

[a] Proportion of the variance explained by the first dimension.

Source: Author's estimations.

Table 2: Principal Components Analysis: An Index of Trade Distance

	Country									
Variable	Argentine	Brazil	Chile	Colombia	Ecuador	El Salvador	Guatemala	Honduras	Mexico	Panama
Distance	0.5726	0.6547	0.5890	0.5135	0.5802	0.5419	0.5524	0.5443	0.5879	0.5465
Real Exchange Rate	0.1540	0.0872	0.1300	0.3577	0.1666	0.2078	0.1848	0.1908	0.1815	0.2731
Political Difference	0.2318	0.2634	0.1990	0.4000	0.3207	0.3928	0.3376	0.3593	0.3699	0.0808
Border	-0.4976	-0.5892	-0.4149	-0.3257	-0.3178	-0.3589	-0.4215	-0.4061	-0.2475	-0.3689
Island	0.2612	0.3804	0.3066	0.2671	0.3433	0.3372	0.2734	0.2422	0.3514	0.3963
Common Language	-0.5280	-0.0507	-0.5748	-0.5205	-0.5603	-0.5161	-0.5425	-0.5608	-0.5477	-0.5719
Proportion[a]	42.28%	31.80%	39.79%	45.84%	40.07%	41.79%	40.98%	39.53%	37.36%	38.29%

[a] Proportion of the variance explained by the first component.
Source: Author's estimations.

We estimate the following equation:

$$ln\left\{Exp_{ij,t}\right\} = \beta_0 + \beta_1 ln\left\{Exp_{ij,t-1}\right\} + \beta_2 ln\left\{GDP_{ij,t}\right\} + \beta_3 EcDist_{ij,t} + \mu_{ij,t} \quad (1)$$

where

$ln\left\{Exp_{ij,t}\right\}$ is the level of exports from country i to country j in year t.

$ln\left\{GDP_{ij,t}\right\}$ is the product of the GDP of country i and country j in year t.

$EcDist_{ij,t}$ is our index of trade distance between country i and country j in year t.

$\mu_{ij,t}$ is an stochastic perturbation.

As can be observed in Table 3 a higher level in our index of trade distance implies a reduction in the level of exports from country i to country j. Additionally, the short term GDP elasticity is positive and ranges between 0.22 and 1.17 for Brazil and El Salvador, respectively. Regarding the long term GDP elasticity, which is equal to $\beta_2/(1-\beta_1)$, this fluctuates between 0.70 for Guatemala and 1.60 for Ecuador. We test the presence of panel unit roots, and this hypothesis is rejected, as well as autocorrelation of order two. Additionally, over identification restrictions are valid for each model. Finally, all our variables are statistically significant at 5%.

We estimate different models in order to perform some robustness checks related to panel data specification, their assumptions about stochastic perturbations and the multivariate statistical technique used to calculate our measure of trade distance. First, we estimate all our dynamic models using a measure of trade distance gotten by Principal Component Analysis, and as can be seen in Table 4, we obtain similar outcomes. Additionally, we use the trade distance based on Factor Analysis for Mixed Data to estimate the gravity equation using Pooled Ordinary Least Squared, Feasible Generalized Least Squared, Prais-Winsten, Random Effects and Instrumental Variables estimators. As depicted in Tables 7 and 8, the results are robust to static specification, as well as assumptions to stochastic errors. Finally, we use our measure of trade distance based on Principal Components Analysis to estimate the gravity equation using the above methods. We can observe in Tables 9 and 10 that our outcomes are robust to the multivariate statistical method to calculate the trade distance.

Conclusions

The empirical success of the gravity model is explained by different factors. Its theoretical underpinning, wide range of fields of application, the accessibility of the data needed for its estimation, and the respectability of a number of seminal papers that have established the gravity model's reputation, have been decisive.

However, a persistent problem in the literature related to the gravity model of international trade is associated with the conceptualization and measurement of distance. Even though some studies have applied direct measures of transport cost barriers to trade, the majority of them rely on distance as a proxy for transport costs. Besides that, there is no consensus on the extent of the distance concept. Certainly, geographical distance and transport costs are relevant. But besides that, synchronization costs, transaction costs or cultural distances should also be considered.

Taking this limitation into account, the main contribution of this paper is to propose a measure of trade distance based on geographical, economic, cultural and political aspects, applying it to test the gravity model in 10 Latin American economies. In our analysis, the gravity equation fits the data well. A higher level in our index of trade distance implies a reduction in trade flows between the trade partners.

This measure is built using Factor Analysis for Mixed Data, which is a multivariate statistical technique that simultaneously incorporates numerical and categorical variables. Our econometric exercises suggest that there are sensible outcomes. These results were also tested using Principal Component Analysis, and basically the same results were obtained.

Table 3: Econometric Results: Blundell and Bond (1998) Dynamic Panel Data Model using FAMD

	Country									
Variable	Argentine	Brazil	Chile	Colombia	Ecuador	El Salvador	Guatemala	Honduras	Mexico	Panama
$EcDist_{it}$	-0.7064	-0.2616	-0.3452	-0.2472	-1.5057	-1.0140	-0.3897	-1.205	-0.1915	-3.3197
	(0.0230)	(0.0208)	(0.0152)	(0.0204)	(0.0429)	(0.0962)	(0.0347)	(0.0862)	(0.0218)	(0.0993)
$ln\{GDP_{it}\}$	0.2721	0.2263	0.8514	0.4997	0.6512	1.1793	0.4662	0.9821	0.4369	0.7783
	(0.0069)	(0.0113)	(0.0096)	(0.0057)	(0.0153)	(0.0516)	(0.0181)	(0.0206)	(0.0059)	(0.0449)
$ln\{Exp_{it-1}\}$	0.6997	0.7131	0.0129	0.3307	0.5936	0.1346	0.3420	0.0601	0.5051	0.1701
	(0.0144)	(0.0151)	(0.0015)	(0.0077)	(0.0012)	(0.0066)	(0.0097)	(0.0048)	(0.0075)	(0.0079)
$Constant$	-8.2684	-6.196	-24.9836	-13.5241	-25.8109	-45.8623	-12.4084	-34.9532	-13.5999	-26.8434
	(0.3213)	(0.3071)	(0.4782)	(0.2570)	(0.7350)	(2.7437)	(0.8389)	(1.0430)	(0.2406)	(2.2429)
Long Term GDP Elasticity	0.9061	0.7888	0.8625	0.7466	1.6024	1.3627	0.7085	1.0449	0.8828	0.9378
Autocorrelation Test[a]	-1.0073^{+}	-1.6868^{+}	0.4083^{+}	0.1791^{+}	1.3157^{+}	-0.4580^{+}	1.5740^{+}	1.4487^{+}	0.0846^{+}	0.6738^{+}
OIR Test[b]	35.2578^{+}	38.2503^{+}	43.4051^{+}	35.3016^{+}	36.6280^{+}	33.1575^{+}	35.5747^{+}	31.3575^{+}	38.3757^{+}	29.1993^{+}
PUR Test[c]	-11.4508^{-}	-1.4259^{-}	-12.2991^{-}	-9.9089^{-}	-5.9416^{-}	-14.482^{-}	-4.4724^{-}	-3.1642^{-}	-3.6359^{-}	-8.2050^{-}

Standard deviation in parenthesis. All variables are statistically significant at 5%.

[a] Arellano-Bond autocorrelation test. Null hypothesis: No autocorrelation of order 2.

[b] Sargent test of Over Identification Restrictions. Null hypothesis: Overidentifying restrictions are valid.

[c] Harris and Tzavalis (1999) Panel Unit Root test applied to $ln\{Exp_{it}\}$. Null hypothesis: panel unit root.

$^{+}$ No rejection of the null hypothesis.

$^{-}$ Rejection of the null hypothesis.

Source: Authors' estimations.

Table 4: Econometric Results: Blundell and Bond (1998) Dynamic Panel Data Model using PCA

	Country									
Variable	Argentine	Brazil	Chile	Colombia	Ecuador	El Salvador	Guatemala	Honduras	Mexico	Panama
$EcDist_{it}$	-0.8116	-0.2565	-0.3670	-0.2588	-1.5484	-1.1298	-0.3874	-1.1159	-0.3220	-3.2345
	(0.02521)	(0.0370)	(0.0148)	(0.0173)	(0.0462)	(0.0777)	(0.0254)	(0.1813)	(0.0295)	(0.1024)
$ln\{GDP_{it}\}$	0.2745	0.2351	0.8512	0.4892	0.5878	1.1895	0.4605	0.9371	0.4543	0.5654
	(0.0069)	(0.0081)	(0.0087)	(0.0095)	(0.0164)	(0.0394)	(0.0177)	(0.0222)	(0.0159)	(0.0561)
$ln\{Exp_{it-1}\}$	0.6929	0.7039	0.0101	0.3229	0.5915	0.1263	0.3401	0.0621	0.4835	0.1754
	(0.0104)	(0.0113)	(0.0015)	(0.0076)	(0.0017)	(0.0051)	(0.0074)	(0.0057)	(0.0142)	(0.0089)
$Constant$	-8.2521	-6.4801	-24.9264	-12.8217	-22.5669	-46.2200	-12.0964	-32.7187	-14.1065	-16.4352
	(0.2204)	(0.2246)	(0.4435)	(.4153)	(0.8322)	(2.0225)	(0.8355)	(1.1197)	(0.6257)	(2.7790)
Long Term GDP Elasticity	0.8939	0.7940	0.8598	0.7225	1.4391	1.3615	0.6978	0.9991	0.8796	0.6856
Autocorrelation Test[a]	-1.0075^{+}	-1.6787^{+}	0.2728^{+}	0.0296^{+}	1.3459^{+}	-0.4907^{+}	1.5710^{+}	1.5608^{+}	0.0419^{+}	0.7028^{+}
OIR Test[b]	35.3612^{+}	38.5987^{+}	43.1190^{+}	36.0049^{+}	35.4587^{+}	31.3089^{+}	35.6524^{+}	30.9464^{+}	38.4743^{+}	29.1817^{+}
PUR Test[c]	-11.4508^{-}	-1.4259^{-}	-12.2991^{-}	-9.9089^{-}	-5.9416^{-}	-14.482^{-}	-4.4724^{-}	-3.1642^{-}	-3.6359^{-}	-8.2050^{-}

Standard deviation in parenthesis. All variables are statistically significant at 5%.

[a] Arellano-Bond autocorrelation test. Null hypothesis: No autocorrelation of order 2.

[b] Sargent test of Over Identification Restrictions. Null hypothesis: Overidentifying restrictions are valid.

[c] Harris and Tzavalis (1999) Panel Unit Root test applied to $ln\{Exp_{it}\}$. Null hypothesis: panel unit root.

$^{+}$ No rejection of the null hypothesis.

$^{-}$ Rejection of the null hypothesis.

Source: Authors' estimations.

Acknowledgements

We thank Michael Jetter for his help and advice in the development of this paper, and additionally, the participants of the seminary developed by the research group in Economics from Universidad EAFIT for fruitful comments and discussions. Also, we are grateful for excellent research assistance from David Alejandro Giraldo. All remaining errors are our own.

References

Anderson, J. and Marcouiller, D. (1999). Trade, location and security: an empirical investigation. Working Paper 7000, NBER.

Anderson, J. E. (1979). A theoretical foundation for the gravity equation. *American Economic Review*, 69:106–116.

Anderson, J. E. and van Wincoop, E. (2003). Gravity with gravitas: a solution to the border puzzle. *American Economic Review*, 93(1):170–192.

Baier, S. and Bergstrand, J. (2001). The growth of world trade: tariffs, transport costs, and income similarity. *Journal of International Economics*, 53:1â27.

Baldwin, R. and Taglioni, D. (1999). Gravity for dummies and dummies for gravity equations. *NBER Working Paper Series*, (Working Paper 12516):1–31.

Batra, A. (2004). Indiaâs global trade potential: The gravity model approach. Working Paper 151, Indian Council for Research on International Economic Relations.

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

Blundell, R. and Bond, S. (1998). Initials conditions and moment restrictions in dynamics panel data models. *Journal of Econometrics*, 87:115–143.

Brun, J., CarrÃ¨re, C., Guillaumont, P., and de Melo, J. (2005). Has distance died? evidence from a panel gravity model. *The World Bank Economic Review*, 19(1):99–120.

Bun, M. and Klaassen, F. (2002). The importance of dynamics in panel gravity models of trade. Technical report, University of Amsterdam.

Cameron, C. and Trivedi, P. (2005). *Microeconometrics: Methods and Applications*. Cambridge, first edition.

Correia, L. J. (2008). The determinants of colombian exports: an empirical analysis using the gravity model. *Revista Desarrollo y Sociedad, Universidad de los Andes-CEDE*.

Deardorff, A. V. (1995). Determinants of bilateral trade: does gravity work in a neoclassical world? *NBER*, (Working Paper 5377):1–28.

Eaton, J. and Kortum, S. (2002). Technology, geography and trade. *Econometrica*, 70(5):1741–1779.

Escofier, B. and Pagès, J. (1994). Multiple factor analysis. *Computational Statistics & Data Analysis*, 18:121–140.

Evenett, S. J. and Keller, W. (2002). On theories explaining the success of the gravity equation. *Journal of Political Economy*, 110(2):281–316.

Feenstra, R. C., Markusen, J. R., and Rose, A. K. (2001). Using the gravity equation to differentiate among alternative theories of trade. *Canadian Journal of Economics*, 34(2):430–447.

Harris, R. and Tzavalis, E. (1999). Inference for unit roots in dynamic panels where the time dimension is fixed. *Journal of Econometrics*, 91:201–226.

Huang, R. (2007). Distance and trade: Disentangling unfamiliarity effects and transport cost effects. *European Economic Review*, 51:161–181.

Husson, F., Josse, J., Le, S., and Mazet, J. (2013). *Multivariate Exploratory Data Analysis and Data Mining with R. R package 1.25.*

Leamer, E. and Levinsohn, J. (1995). International trade theory: the evidence. In Grossman, G. and Rogoff, K., editors, *The handbook of international economics*. North-Holland, Elsevier.

Leamer, E. and Stern, R. (1970). *Quantitative international economics*. Allyn and Bacon, Boston.

Linnemann, H. (1966). *An econometric study of international trade flows*. North-Holland, Amsterdam.

Mejia, J. F. (2011). *Export Diversification and Economic Growth: An Analysis of Colombiaâs Export Competitiveness in the European Unionâs Market*. Springer, Berlin.

Pagès, J. (2004). Analyse factorielle de donnees mixtes. *Revue Statistique Appliquee*, 4:93–111.

Pöyhönen, P. (1963). A tentative model for the volume of trade between countries. *Weltwirtschaftliches Archiv*, 90(1):93–99.

R Development Core Team (2013). *R: A Language and Environment for Statistical Computing. R. Version 3.01.* R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Rose, A. K. (2000). One money, one market: Estimating the effect of common currencies on trade. *Economic Policy*, 15(30):7–46.

Rose, A. K. (2004). So we really know that the WTO increases trade? *American Economic Review*, 94(1):98–114.

Savage, I. R. and Deutsch, K. W. (1960). A statistical model of the gross analysis of transaction flows. *Econometrica*, 28(3):551–572.

Soloaga, I. and Winters, L. A. (2001). Regionalism in the nineties: what effect on trade? *North American Journal of Economics and Finance*, 12(1):1–29.

StataCorp (2011). *Stata Statistical Software: Release 12.* College Station, TX: StataCorp LP.

Subramanian, A. and Wei, S. (2007). The WTO promotes trade, strongly but unevenly. *Journal of International Economics*, 72(1):151–175.

Tinbergen, J. (1962). *Shaping the world economy: suggestion for an international economic policy.* Twentieth Century Fund, New York.

World-Bank (2002). Trade frictions and welfare in the gravity model.

Annex

Table 5: Data sources and calculations

Variable	Definition	Source
Exports	Value of exports to trade partner in US$	IMF's Direction of Trade Statistics
GDP	Exporter country's GDP in US$	United Nations National Accounts, Main Aggregates Database
GDP Partner	Importer country's GDP in US$	United Nations National Accounts, Main Aggregates Database
Political Difference	Absolute value of the difference between Democracy Index in both countries	Democracy Index measured by Polity2 variable from the Polity IV data set
Distance	Great-circle distance between the trade partners measured in kilometres	CIA World Factbook (latitudes and longitudes, capital cities). US Federal communications Commission Website, Converter
Real Exchange Rate	Bilateral real exchange rate	IMF's International Financial Statistics Database (nominal exchange rates, producer price indexes)
Border	Binary dummy variable, unity if the country pair shares a land border	CIA World Factbook
Island	Dummy variable, unity if the partner country is an island	CIA World Factbook
Common Language	Dummy variable, takes the value of one if both countries share a common language	CIA World Factbook

Source: Author's calculations.

Table 6: Descriptive Statistics

		Country									
Variable		Argentine	Brazil	Chile	Colombia	Ecuador	El Salvador	Guatemala	Honduras	Mexico	Panama
Exports	Mean	8.29E+08	1.86E+09	6.21E+08	4.91E+08	2.10E+08	7.53E+07	1.16E+08	5.02E+07	4.57E+09	2.68E+07
	St. Dev.	1.51E+09	3.26E+09	1.17E+09	1.51E+09	7.11E+08	2.93E+08	3.89E+08	1.75E+08	2.59E+10	7.03E+07
GDP	Mean	2.51E+11	9.69E+11	1.08E+11	1.48E+11	3.24E+10	1.54E+10	2.43E+10	9.02E+09	7.14E+11	1.50E+10
	St. Dev.	7.22E+10	4.53E+11	4.42E+10	5.98E+10	1.36E+10	3.88E+09	9.19E+09	3.31E+09	2.39E+11	5.41E+09
GDP Partner	Mean	8.82E+11	8.09E+11	7.88E+11	7.29E+11	8.44E+11	9.09E+11	8.63E+11	9.57E+11	8.04E+11	9.37E+11
	St. Dev.	1.98E+12	1.92E+12	1.83E+12	1.89E+12	1.97E+12	2.04E+12	1.99E+12	2.09E+12	1.92E+12	2.06E+12
Distance	Mean	10207.6100	9715.6910	10642.2900	6381.3260	8649.7060	7807.9990	8164.9720	7446.4000	8436.4120	7580.3270
	St. Dev.	5302.1310	4769.1450	5169.5310	3582.2480	5722.9620	5125.2080	5105.7850	4880.3120	4363.4530	5398.9190
Exchange Rate	Mean	1.1199	1.3144	272.8375	1378.1040	2985.4030	4.4933	4.1457	10.2099	5.7332	0.4426
	St. Dev.	1.2862	1.8334	274.0287	1260.8280	8084.3930	4.5362	4.1006	10.2587	5.7577	0.4922
Political Difference	Mean	3.3368	4.0929	2.7798	2.3480	3.8564	2.7904	3.3472	3.3516	3.2051	1.8598
	St. Dev.	4.5627	5.3594	4.5863	1.1147	3.8458	2.6069	4.3656	3.3436	3.5251	3.1381
Polity2	Mean	7.7500	8.0000	9.0000	7.0000	6.6875	7.1250	7.6875	6.7500	7.1250	9.0000
	St. Dev.	0.4334	0.0000	0.7911	0.0000	1.5309	0.3310	1.2113	0.4334	1.4098	0.0000
Polity2 Partner	Mean	6.6250	5.7307	7.1917	9.0236	6.9257	8.0404	6.9097	7.8750	7.5384	8.1288
	St. Dev.	5.5424	6.3517	5.0040	1.6318	5.1817	3.6859	5.3205	4.5776	4.5502	3.5430
Border	Mean	0.1389	0.1538	0.0682	0.1351	0.0541	0.0588	0.0833	0.0625	0.0513	0.0606
	St. Dev.	0.3461	0.3611	0.2522	0.3422	0.2263	0.2355	0.2766	0.2423	0.2207	0.2388
Island	Mean	0.1389	0.1282	0.1591	0.1081	0.1351	0.1176	0.1111	0.1563	0.1282	0.1515
	St. Dev.	0.3461	0.3346	0.3660	0.3108	0.3422	0.3225	0.3145	0.3634	0.3346	0.3589
Common Language	Mean	0.2778	0.0256	0.2955	0.3514	0.3514	0.3235	0.3333	0.3125	0.3333	0.3636
	St. Dev.	0.4483	0.1582	0.4566	0.4778	0.4778	0.4683	0.4718	0.4640	0.4718	0.4815
Number of countries		36	39	44	37	37	34	36	32	39	33
Time period		1995-2010	1995-2010	1995-2010	1995-2010	1995-2010	1995-2010	1995-2010	1995-2010	1995-2010	1995-2010
Percentage of exports		74.84%	71.12%	70.48%	83.94%	85.68%	87.56%	86.76%	85.23%	97.62%	86.73%

Source: Authors' estimations.

Table 7: Econometric Results: Static Panel Data Models using FAMD (South America)

			Argentine		
Variable	Pooled OLS	FGLS AR(1)	Prais-Winsten	Random Effect	Instrumental Variable
$EcDist_{it}$	-0.6908	-0.7053	-0.6817	-0.6916	-0.6246
	(0.1156)	(0.1232)	(0.1682)	(0.1103)	(0.0901)
$ln\{GDP_{it}\}$	0.6361	0.6443	0.5752	0.6895	0.6033
	(0.1044)	(0.1099)	(0.1292)	(0.1103)	(0.0861)
Constant	-13.7579	-14.2361	-10.7111	-16.5664	-11.9290
	(5.5745)	(5.8810)	(6.9561)	(5.8892)	(4.5269)
			Brazil		
Variable	Pooled OLS	FGLS AR(1)	Prais-Winsten	Random Effect	Instrumental Variable
$EcDist_{it}$	-0.4506	-0.3702	-0.4072	-0.3843	-0.3798
	(0.079)	(0.0801)	(0.0392)	(0.0832)	(0.0718)
$ln\{GDP_{it}\}$	0.6216	0.5851	0.596	0.7193	0.6007
	(0.0487)	(0.0285)	(0.0277)	(0.0213)	(0.0209)
Constant	-12.8745	-10.9692	-11.537	-18.1247	-11.7063
	(2.6290)	(1.5452)	(1.5016)	(1.1712)	(1.1284)
			Chile		
Variable	Pooled OLS	FGLS AR(1)	Prais-Winsten	Random Effect	Instrumental Variable
$EcDist_{it}$	-0.5444	-0.5441	-0.5285	-0.5371	-0.5001
	(0.0754)	(0.0768)	(0.1244)	(0.0864)	(0.0818)
$ln\{GDP_{it}\}$	0.8196	0.8300	0.8359	0.8625	0.7690
	(0.0541)	(0.0573)	(0.0985)	(0.0612)	(0.0463)
Constant	-23.1168	-23.677	-24.0371	-25.3276	-20.4755
	(2.7888)	(2.9607)	(5.1971)	(3.1582)	(2.3912)
			Colombia		
Variable	Pooled OLS	FGLS AR(1)	Prais-Winsten	Random Effect	Instrumental Variable
$EcDist_{it}$	-0.8178	-0.6340	-0.6730	-0.4655	-0.4338
	(0.1453)	(0.1118)	(0.1108)	(0.0933)	(0.0912)
$ln\{GDP_{it}\}$	0.8151	0.7304	0.7451	0.7374	0.6960
	(0.1386)	(0.0970)	(0.0849)	(0.0979)	(0.0468)
Constant	-23.7682	-19.466	-20.2142	-19.7699	-17.6023
	(7.2072)	(5.1214)	(4.4427)	(5.0905)	(2.4259)
			Ecuador		
Variable	Pooled OLS	FGLS AR(1)	Prais-Winsten	Random Effect	Instrumental Variable
$EcDist_{it}$	-1.4567	-1.4791	-1.4711	-1.5113	-1.4722
	(0.1998)	(0.2143)	(0.1988)	(0.2414)	(0.1831)
$ln\{GDP_{it}\}$	1.0115	1.0459	1.0553	1.1327	1.1590
	(0.1544)	(0.1602)	(0.1172)	(0.1835)	(0.1050)
Constant	-34.2026	-36.0124	-36.5238	-40.2945	-41.5795
	(7.8687)	(8.1881)	(6.0060)	(9.3849)	(5.2957)

Standard deviation in parenthesis.
All variables are statistically significant at 5%.
Source: Authors' estimations.

Table 8: Econometric Results: Static Panel Data Models using FAMD (Central America)

			El Salvador		
Variable	Pooled OLS	FGLS AR(1)	Prais-Winsten	Random Effect	Instrumental Variable
$EcDist_{it}$	-1.4757	-1.3707	-1.345	-1.006	-1.0469
	(0.2900)	(0.2716)	(0.2773)	(0.3453)	(0.2231)
$ln\{GDP_{it}\}$	0.9482	0.903	0.8962	1.1653	1.1572
	(0.2408)	(0.2204)	(0.1671)	(0.2214)	(0.1247)
$Constant$	-32.3937	-30.0812	-29.7359	-43.1715	-42.7964
	(12.0957)	(11.1018)	(8.4513)	(11.2506)	(6.2169)
			Guatemala		
Variable	Pooled OLS	FGLS AR(1)	Prais-Winsten	Random Effect	Instrumental Variable
$EcDist_{it}$	-0.8298	-0.8637	-0.8402	-0.7572	-0.7338
	(0.2034)	(0.2055)	(0.1775)	(0.1983)	(0.1724)
$ln\{GDP_{it}\}$	0.5529	0.6325	0.6517	0.785	0.6757
	(0.1865)	(0.1773)	(0.1300)	(0.1519)	(0.0706)
$Constant$	-11.2263	-15.2449	-16.2173	-22.8501	-17.2927
	(9.3654)	(8.9658)	(6.5927)	(7.7022)	(3.5573)
			Honduras		
Variable	Pooled OLS	FGLS AR(1)	Prais-Winsten	Random Effect	Instrumental Variable
$EcDist_{it}$	-1.8806	-2.1698	-2.1394	-2.5399	-1.8706
	(0.3514)	(0.3944)	(0.2922)	(0.4845)	(0.2941)
$ln\{GDP_{it}\}$	1.597	1.8512	1.8246	2.6373	1.8779
	(0.3713)	(0.4008)	(0.2374)	(0.4605)	(0.2013)
$Constant$	-64.6342	-77.5194	-76.1764	-120.0213	-77.9149
	(18.5010)	(20.0134)	(11.7670)	(22.9292)	(9.9261)
			Mexico		
Variable	Pooled OLS	FGLS AR(1)	Prais-Winsten	Random Effect	Instrumental Variable
$EcDist_{it}$	-0.712	-0.4963	-0.6491	-0.3363	-0.462
	(0.1732)	(0.0954)	(0.0561)	(0.1261)	(0.1001)
$ln\{GDP_{it}\}$	0.7571	0.773	0.7603	0.7833	0.8413
	(0.1378)	(0.0693)	(0.0427)	(0.0735)	(0.0358)
$Constant$	-20.8834	-21.6291	-20.9914	-22.2817	-25.4095
	(7.3531)	(3.7431)	(2.2897)	(4.0213)	(1.9233)
			Panama		
Variable	Pooled OLS	FGLS AR(1)	Prais-Winsten	Random Effect	Instrumental Variable
$EcDist_{it}$	-1.4346	-1.3522	-1.3398	-1.3727	-1.7487
	(0.3389)	(0.3146)	(0.3555)	(0.3486)	(0.3459)
$ln\{GDP_{it}\}$	0.8857	0.8262	0.826	0.8537	1.2556
	(0.2336)	(0.2166)	(0.2322)	(0.2594)	(0.1829)
$Constant$	-29.5369	-26.4257	-26.3461	-27.9477	-48.2686
	(11.7124)	(10.8741)	(11.7299)	(13.0383)	(9.1072)

Standard deviation in parenthesis.
All variables are statistically significant at 5%.
Source: Authors' estimations.

Table 9: Econometric Results: Static Panel Data Models using PCA (South America)

		Argentine			
Variable	Pooled OLS	FGLS AR(1)	Prais-Winsten	Random Effect	Instrumental Variable
$EcDist_{it}$	-0.6985	-0.7156	-0.7122	-0.7142	-0.6548
	(0.1175)	(0.1262)	(0.1707)	(0.0967)	(0.0893)
$ln\{GDP_{it}\}$	0.6437	0.6543	0.5962	0.6976	0.6170
	(0.1060)	(0.1122)	(0.1326)	(0.0802)	(0.0854)
$Constant$	-14.1609	-14.7601	-11.8195	-16.9911	-12.6502
	(5.6568)	(6.0019)	(7.1346)	(4.2164)	(4.4889)
		Brazil			
Variable	Pooled OLS	FGLS AR(1)	Prais-Winsten	Random Effect	Instrumental Variable
$EcDist_{it}$	-0.4516	-0.3858	-0.4144	-0.4024	-0.3967
	(0.0791)	(0.0813)	(0.0394)	(0.0825)	(0.0721)
$ln\{GDP_{it}\}$	0.6226	0.5921	0.6013	0.7224	0.6041
	(0.0486)	(0.0290)	(0.0277)	(0.0211)	(0.0208)
$Constant$	-12.9297	-11.3463	-11.8228	-18.2942	-11.8904
	(2.6267)	(1.5710)	(1.5031)	(1.1608)	(1.1270)
		Chile			
Variable	Pooled OLS	FGLS AR(1)	Prais-Winsten	Random Effect	Instrumental Variable
$EcDist_{it}$	-0.5449	-0.5448	-0.5308	-0.5388	-0.5059
	(0.0753)	(0.0767)	(0.1248)	(0.0873)	(0.0815)
$ln\{GDP_{it}\}$	0.8193	0.8299	0.8370	0.8606	0.7692
	(0.0541)	(0.0574)	(0.0988)	(0.0617)	(0.0462)
$Constant$	-23.1008	-23.6733	-24.0950	-25.2249	-20.4854
	(2.7901)	(2.9631)	(5.2152)	(3.1828)	(2.3844)
		Colombia			
Variable	Pooled OLS	FGLS AR(1)	Prais-Winsten	Random Effect	Instrumental Variable
$EcDist_{it}$	-0.7961	-0.6662	-0.6992	-0.5273	-0.4712
	(0.1437)	(0.1134)	(0.1122)	(0.1292)	(0.0933)
$ln\{GDP_{it}\}$	0.8030	0.7425	0.7546	0.7056	0.6737
	(0.1385)	(0.0989)	(0.0850)	(0.0963)	(0.0455)
$Constant$	-23.1453	-20.0839	-20.6997	-18.1363	-16.4650
	(7.2050)	(5.2133)	(4.4457)	(5.0036)	(2.3583)
		Ecuador			
Variable	Pooled OLS	FGLS AR(1)	Prais-Winsten	Random Effect	Instrumental Variable
$EcDist_{it}$	-1.4102	-1.4515	-1.4605	-1.4678	-1.4353
	(0.2023	(0.2225)	(0.1849)	(0.2353)	(0.1857)
$ln\{GDP_{it}\}$	1.0230	1.0626	1.0734	1.1138	1.1396
	(0.1682)	(0.1745)	(0.1161)	(0.1873)	(0.1054)
$Constant$	-34.7811	-36.8605	-37.4390	-39.3421	-40.6090
	(8.5666)	(8.9145)	(5.9504)	(9.5826)	(5.3152)

Standard deviation in parenthesis.
All variables are statistically significant at 5%.
Source: Authors' estimations.

Table 10: Econometric Results: Static Panel Data Models using PCA (Central America)

			El Salvador		
Variable	Pooled OLS	FGLS AR(1)	Prais-Winsten	Random Effect	Instrumental Variable
$EcDist_{it}$	-1.4410	-1.3706	-1.3459	-1.1511	-1.1235
	(0.2802)	(0.2638)	(0.3050)	(0.3394)	(0.2259)
$ln\{GDP_{it}\}$	0.9682	0.9363	0.9309	1.2206	1.2038
	(0.2538)	(0.2312)	(0.1874)	(0.2240)	(0.1260)
$Constant$	-33.3871	-31.7367	-31.4544	-45.9148	-45.1051
	(12.7344)	(11.6302)	(9.4661)	(11.3694)	(6.2786)
			Guatemala		
Variable	Pooled OLS	FGLS AR(1)	Prais-Winsten	Random Effect	Instrumental Variable
$EcDist_{it}$	-0.8235	-0.8665	-0.8547	-0.7272	-0.7630
	(0.2025)	(0.2052)	(0.1847)	(0.2119)	(0.1714)
$ln\{GDP_{it}\}$	0.5514	0.6460	0.6730	0.7863	0.6644
	(0.1866)	(0.1773)	(0.1329)	(0.1495)	(0.0699)
$Constant$	-11.1513	-15.9459	-17.3147	-22.9154	-16.7167
	(9.3725)	(8.9702)	(6.7437)	(7.5850)	(3.5219)
			Honduras		
Variable	Pooled OLS	FGLS AR(1)	Prais-Winsten	Random Effect	Instrumental Variable
$EcDist_{it}$	-1.8609	-2.1071	-2.0854	-2.4194	-1.8213
	(0.3491)	(0.3896)	(0.2925)	(0.4539)	(0.2917)
$ln\{GDP_{it}\}$	1.5852	1.8233	1.8006	2.5851	1.8348
	(0.3677)	(0.3942)	(0.2378)	(0.4474)	(0.1993)
$Constant$	-64.0538	-76.1723	-75.0172	-113.1776	-75.7803
	(18.3233)	(19.6914)	(11.7924)	(22.2994)	(9.8279)
			Mexico		
Variable	Pooled OLS	FGLS AR(1)	Prais-Winsten	Random Effect	Instrumental Variable
$EcDist_{it}$	-0.7123	-0.5347	-0.6669	-0.4457	-0.4854
	(0.1757)	(0.1108)	(0.0600)	(0.1276)	(0.1032)
$ln\{GDP_{it}\}$	0.7659	0.8090	0.7901	0.8077	0.8416
	(0.1402)	(0.0735)	(0.0444)	(0.0766)	(0.0358)
$Constant$	-21.3567	-23.5808	-22.6035	-23.5841	-25.4120
	(7.4856)	(3.9757)	(2.3869)	(4.1809)	(1.9230)
			Panama		
Variable	Pooled OLS	FGLS AR(1)	Prais-Winsten	Random Effect	Instrumental Variable
$EcDist_{it}$	-1.4118	-1.3265	-1.3135	-1.2799	-1.7237
	(0.3358)	(0.3092)	(0.3525)	(0.3330)	(0.3371)
$ln\{GDP_{it}\}$	0.8664	0.8046	0.8034	0.7836	1.1869
	(0.2313)	(0.2134)	(0.2306)	(0.2476)	(0.1769)
$Constant$	-28.5814	-25.3561	-25.2262	-24.4737	-44.8640
	(11.6004)	(10.7121)	(11.6443)	(12.4592)	(8.8099)

Standard deviation in parenthesis.
All variables are statistically significant at 5%.
Source: Authors' estimations.