International Trade and Economic Theory: Testing the Gravity Model

Charlotte Cuccia, Zack Humphries, Gina Piazza

Abstract

This paper offers an addition to the extensive body of research of the gravity model, which predicts that the volume of trade between two countries depends on their relative size and the distance between them. Using both simple and multiple linear regressions, the effects of the various factors of volume of trade, including distance, GNP, regulation, and service sector infrastructure are tested. Thus, relationships are drawn between these variables in a technologically driven age.

I. Introduction

The gravity model is a theory in International Economics which attempts to describe the effects of country size (measured in GNP) and distance on the volume of trade between two countries. This model was used as a starting point to examine variables that potentially affected bilateral trade, which is especially important when considering potential trade policies to enact. As some countries emphasize the importance of reducing trade barriers and encouraging liberalization, others bend more towards protectionism, limiting imports and in some cases people from crossing the border. By investigating which independent variables have a significant effect on volume of trade and in what direction that effect goes, one can utilize more effective trade policy by focusing only on those things which actually affect trade. On the other hand, this also implies a country could intentionally verbally hawk protectionism and undergo seemingly protectionist policies without actually decreasing the level of trade, if the policy is enacted on a variable that isn't significant. Especially in an increasingly connected global economy, the effects of various variables on the level of trade is very important for businesses and governments alike.

The hypothesis for the simple regression model, the effects of distance on bilateral trade, was that as distance increases, trade would decrease. This aligns with the gravity model as well as economic intuition. It was also predicted that as GNP increased, trade would increase, which goes along with the gravity model as well. In addition, it was predicted that more tariffs would cause trade to decrease and that more transparency would cause trade to increase. Concerning the service sector, the hypothesis was that if a country had a more developed service industry, there would be more trade.

II. Literature Review

Before examining the effect of the first and most obvious variable, distance, on the volume of trade between two countries, other bodies of research necessarily had to be consulted in order to gauge the results garnered through previous studies. Since the gravity model is one of the most robust economic theories, several articles offered potentially interesting analysis, but Cheney (2013) specifically discusses the implications of distance not as a physical barrier, but as a networking barrier to trade. He begins the paper by asking how, over the last fifty years, distance has held as inversely proportional to trade flow despite increasing technology in transportation. Cheney's model develops the idea that as a firm grows, it develops a network full of contacts that it can not only export its products to, but also import its supplies from. He uses the model to predict that over time, these contacts become located further from the firm's location, and finds that firms who export many goods are the same firms that export further away. By combining the traditional gravity model, which speaks about country-level exports, with Zipf's law, which focuses on how much different firms can sell, Cheney's result states that, on an aggregate level, trade flows are affected by the amount of direct communication each firm possesses with its network of suppliers and consumers, thus representing not a physical barrier to trade but rather an informational one. As firms reach more contacts, their exports and imports can increase, but this ability to network is more difficult with larger distances. The impact of these findings can explain in part, then, not only the role of the individual in international trade, but also how the variable distance encompasses more than previously assumed.

The next variable, regulation, also seems to have an obvious relationship with the level of trade between two countries, namely a negative one. However, the relationship between regulations such as tariffs and economic growth isn't completely clear-cut, and Lampe and Sharp (2012) take steps to promote use of a country-by-country analysis for tariff usage not simply a general, average coefficient on a global scale. They examine the relationship between tariffs and income in twenty-four countries since before World War I to potentially give these countries a better way to make future policy decisions. They stay in line with the general economic theory that open trade is preferable to autarky, but investigate claims that temporary protection of infant industries could play a role in economic growth, and therefore play a role in trade levels later down the road. Lampe and Sharp also implore the reader to remember that even if trade liberalization and economic growth have a positive relationship, one can't assume that fewer tariffs lead to higher growth; some economists have pointed out that as countries increase income they could then reduce tariffs. For their results, Lampe and Sharp do see that for the world as a whole, tariffs reduce economic growth, but that when the data is divided into time periods and countries, this doesn't necessarily hold. Unfortunately, they couldn't make any generalizations for which types of country-to-country interactions resulted in increased welfare when tariffs were applied, but it does indicate that perhaps the type of industry being protected can make a difference along with the countries implementing these tariffs. This article primarily makes a case in general for looking at specific country interactions as well as to the world as a whole, thus giving future studies an additional angle in which to investigate the effects of regulatory policies.

Another potentially interesting variable was the existence of a common language between two trading partners, and how that would affect the amount of trade. Intuition suggests that the presence of a common language, denoted as a dummy variable 1 or 0, would increase bilateral trade flows. However, the results showed that common language was not statistically significant, which went against the prediction made, thus suggesting further research would be helpful for further understanding. Mack, E. Martinez-Garcia, and M. Martinez-Garcia (2014) focus on the prevalence of English as a lingua franca which would reduce the importance of a country's primary language as well as the prevalence of multilingual people and the role of acquiring even a second language in reducing trade barriers. They included cases such as Japan's, where no other country possesses Japanese as a first, official language. However, trade is extremely important to Japan, thus the Japanese must invest time and resources into the acquisition of second and third languages or into the infrastructure of nonverbal communication. Their extensive analysis showed how a common official, or first language doesn't necessarily dictate levels of trade, which aided in the decision to drop it as a variable in the final model since it wasn't statistically significant.

This paper adds the to the extensive body of research on the gravity model by considering the effects of distance, regulation, and tariffs, but from a different lens than the papers above. For example, rather than simply looking at tariffs as the sole way to measure regulatory barriers to trade, the paper allows the regulation variable to (separately from tariffs) measure levels of transparency and corruption from the exporting country. This adds another dimension to the understanding of what factors could potentially limit trade flows between two nations. Furthermore, since international trade is an extremely multi-faceted dependent variable, simply examining the independent variables selected increases the overall knowledge and understanding of how the gravity model could be used in conjunction with other policies to increase bilateral trade as a political goal.

III. Data

While it is clear that there exist numerous factors that impact the volume of trade between two countries, the two factors encompassed within the framework of the gravity model are the size of the countries' economies and the distance between the two countries. Because the body of literature discussing the impact of the size of a country's economy on volume of trade is considerably more robust than that discussing the effects of distance on trade, our simple linear regression model attempts to uncover the impact of distance between countries on the volume of trade between them. Our dependent variable is trade, measured as the volume of manufactured exports traded from the exporting country to the importing country in one year. The independent variable is distance. Distance is measure as the geographic distance between the two capitals of the countries in kilometers, as the capitals of the majority of countries in the data set are also their economic centers, and thus the site of great volumes of trade. The Gravity Model indicates that as the distance between two countries decreases, the amount of trade between them increases, and this simple linear regression was used to test this correlation.

In the multiple linear regression model, additional independent variables are added to capture the impact of different factors on the dependent variable, the volume of trade from one given country to a second given country, and limit omitted variable bias. The first additional independent variable is Gross National Product per capita of the importing country, which is included to control for the effects of the size of the economy on trade, which takes into account the other factor of the Gravity Model. Tariffs imposed by the importing country, which are expected to have a negative relationship with volume of trade, was the third included independent variable. The final two independent variables included in the multiple regression model are the regulatory environment of the exporting country and the service sector infrastructure of the exporting country, which help to paint a picture in a broader sense that indicates the economic situation in a country in terms of how it may impact international trade.

The data for this research was gathered from the "World Bank Trade Costs and Facilitation 2000-2001" study. The major sources of data for the World Bank study are the Commodity and Trade Database (COMTRADE) of the United Nations Statistics Division for trade data, the Trade Analysis and Information System (TRAINS) of the United Nations Conference on Trade and Development (UNCTAD) for tariff data, the World Development Indicators published by the World Bank for the data of gross national product (GNP) and per capita GNP, and 3 country surveys for trade facilitation indicator. The data was edited to exclude the trade reports from 2000 to get a more modern interpretation of trade, and the trade values were multiplied by 1000 to better represent the value of manufactured trade, as the original dataset had the values in thousands of United States Dollars. Figure 1 displays the variable names as found in this data set, the interpretation of each variable, and the source for that variable.

Fig. 1

Variable Name	Interpretation	Data Source
trade	The value of manufactured exports from Country J to Country I in year t.	Commodity and Trade Database (COMTRADE) of the United Nations Statistics Division
distance	Geographical distance between capitals of Country I and Country J, measured in kilometers.	World Bank
gnppc_im	GNP per capita of the importing country (Country I) in thousands of 1995 US Dollars.	World Development Indicators of the World Bank
tariff	Tariff rate in the percent ad valorem term that is specific to the trading partners I and J and year t.	United Nations Conference on Trade and Development (UNCTAD)
regulation_ex	Regulatory environment for exporting country (Country J) in year t, measured as an average of transparency and corruption indices.	World Competitiveness Yearbook and Kauffman, Kray, and Zoido-Lobaton Study
services_ex	Service sector infrastructure for exporting country (Country J) in year t, measured as an	World Economic Forum Global Competitiveness Report

	average of internet cost, speed, and access indices.	

Also included is Figure 2, which provides descriptive statistics for each of the variables used in the regression model.

Fig. 2

Variable	Mean	Standard Deviation	Minimum	Maximum
trade	7.99e08	4.68e09	0	1.20e11
distance	8148.503	4892.503	99.726	19945.91
gnppc_im	7.956238	9.023947	0.2521024	31.20666
tariff	8.37888	10.63606	0	294.41
regulation_ex	0.6888278	0.1384331	0.3528432	1
services_ex	0.6740431	0.120041	0.4817447	1

There were 4,558 observations of trade in 2001 between different countries in this dataset. The mode of the trade variable is $0 because some countries do not trade with one another due to either political reasons or the fact that it is not economically advantageous to trade with that country. The average value of trade between countries was $799,057,000, but we see a relatively large standard deviation, meaning that the data is spread out around this mean. Furthermore, this mean may not be a strong measure of central tendency because there were a few very large outliers like the $120 billion total trade from Canada to the United States that skew the average upwards.

For the variable distance, 5,550 observations were included in the data set. The furthest distance was found to be between Taiwan and Paraguay at 19945.91 kilometers, and the shortest was between Finland and Estonia at 99.73 kilometers. The average distance between two countries' capitals is 8148.17 kilometers and is close to the median at 8628.423 kilometers. The distribution of distances between capitals is bimodal, almost trimodal. The reason for the distribution is due to the oceans that separate countries at large distances and the increase of travel needed to go across the oceans.

These are the two variables we will include in the simple linear regression model, so it is worth looking at a scatter plot in which the dependent variable is trade and the dependent variable is distance.

Fig. 3

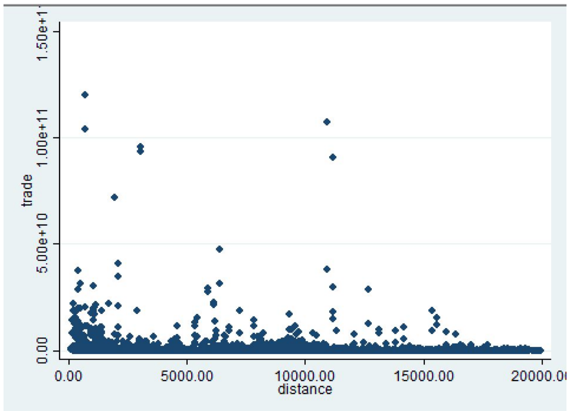
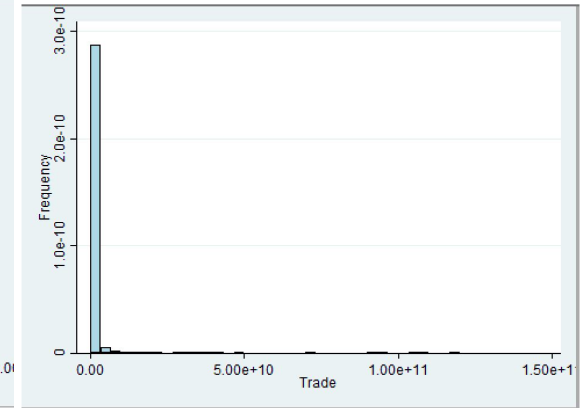


Fig. 4



We see here that there is no strong relationship between the two variables, and despite the increasing values of the variable distance, the values of the variable trade remain relatively low. We see this as well in Figure 3, the highest frequency of occurs for observations at which values of trade are relatively close to zero. These shortcomings can be corrected for by taking the natural logarithm of trade and instead using that as the dependent variable in both the simple and the multiple linear regression model. This variable will be indicated by the name ltrade.

Fig. 5

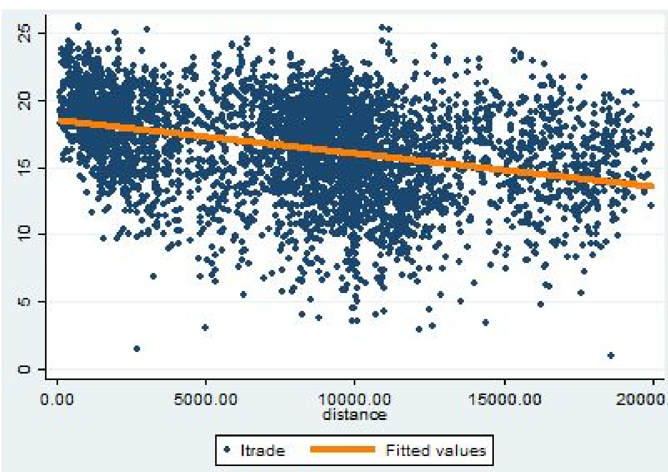
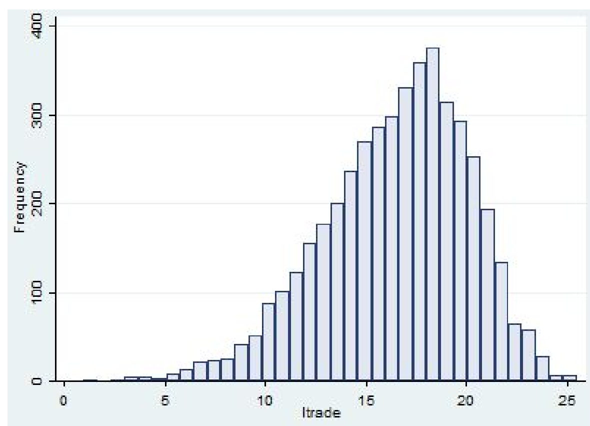


Fig. 6



Using ltrade as the dependent variable in our scatter plot, we know see a much more evenly distributed spread of observations and can much more clearly observe a negative relationship between the two variables. This observation is supported by Figure 5, the histogram of ltrade, which shows a much more normal distribution. For this reason, ltrade will be used in the regression analysis instead of trade.The natural logarithm of trade was taken so that when regression results are considered, the effects of an increase or decrease in any given independent variable can be understood in terms of the effect the percent change they will have on the volume in trade from a given country to another.

Looking at the GNP per capita of the importing country, we see that the average is $7,956.24. There is a great amount of variation in this metric, notable due to the fact that the standard deviation is a large $9,023.95 and the range of values is also quite large. The minimum GNP per capita of $252.10 is found in Nigeria, and the maximum GNP per capita of $31,206.66 is found in Switzerland. Again, it is important to note that the data set measures these values in constant 1995 United States dollars and that values would likely be much larger were a more recent year used as the base year.

The descriptive statistics for the independent variable tariff show a very large range. The lowest value is 0, indicating no tariffs imposed by the importing country on the exporting country. The large majority of values fall between 0 and 100, indicating that in most situations, tariffs imposed do not have a higher monetary value than the imports themselves. There are only eight observations in the data set for which the value of the variable tariff is greater than 100, the highest of which is 294.41 between Sri Lanka and Panama.

For the 5,550 recorded elements of regulation_ex, the variable has an average of .688 and a median of .662. The lowest value of .352 was Venezuela, showing the government often fall privy to corruption and a lack of government transparency for economic trade regulations. The highest value of 1 was Finland, as Finland has a very transparent government process that does not fall to corruption.

For the 5,550 recorded elements of services_ex, the variable has an average of .674 and a median of .645. The lowest value of .481 was Mauritius, an island country in the Indian Ocean near Madagascar. Their low value is due to the lack of internet infrastructure, as only 7.3% of people had access to internet in Mauritius in 2001. The highest value of 1 was again Finland, which has a developed internet infrastructure.

Before running a regression analysis on the data, we will check to see whether it fits the Gauss-Markov Assumptions. Before checking theses assumptions, which apply to multiple linear regression models, we will check the analogous assumptions for our simple linear regression. There are five assumptions for the simple linear regression, indicated respectively by SLR.1, SLR.2, SLR.3, SLR.4, and SLR.5. SLR.1 states that the model is linear in parameters, meaning that in the population, the relationship between x and y is linear. This can be represented by the regression equation below.

$$y = \beta_0 + \beta_1 x + u$$

Our model meets this condition because there is a linear relationship between the independent variable distance and the dependent variable trade in our estimated regression equation.

Assumption SLR.2 states that the data is a random sample drawn from the population. This assumption holds for our data set because the Commodity and Trade Database and the Trade Analysis and Information System gathered the data in such a way as to ensure that the sample size is large, gathering data from as many countries as possible and capturing relationships between various trading partners.

Assumption SLR.3 states that there is sample variation in the explanatory variable, meaning that the values of the explanatory variables are not all the same. Written using the summation operator, this assumption can be expressed as

$$\sum_{i=1}^{n} (x_i - \overline{x})^2 > 0 .$$

Our data set fits this assumption because the values of distance between countries vary depending on the countries given, meaning that the independent variable holds different values.

Finally, assumption SLR.4 is the zero conditional mean assumption, which states that the value of the explanatory variable must contain no information about the mean of the unobserved factors. Mathematically, this can be represented by

$$E(u_i \mid x_i) = 0 .$$

The model does not meet this assumption because the unobserved term contains information about the explanatory variable. For example, some information that might be contained within the unobserved term is information about the income per capita in the country, transportation used to facilitate trade of commodities, or transportation costs, the latter two of which are related to the independent variable of distance between countries. Distance is an interesting independent variable, however, because the distance between the countries is likely not strongly correlated with many of the other variables in the data set. However, because this is a simple regression and so much information is contained within the unobserved term, we will still violate assumption SLR.4.

Because our data does not meet the assumptions for a simple linear regression, we cannot say that our estimators are unbiased. This means that we will either over- or underestimate the coefficients in our model, but running a simple regression analysis will still help us understand the nature of the relationship between distance and trade, especially whether the correlation is positive or negative.

The fifth assumption is the assumption of homoskedasticity, SLR.5, which says that the value of the explanatory variable must contain no information about the variability of the unobserved factors. This can be otherwise represented by

$$Var(u_i \mid x_i) = \sigma^2 .$$

We will assume homoskedasticity for this model because it is likely that the value of variance of the residuals is constant, regardless of the value of distance as the independent variable.

The Gauss-Markov Assumptions for multiple linear regression are denoted by MLR.1-MLR.5. Assumption MLR.1 is the same as SLR.1, and for the multiple regression our model is still linear in parameters. MLR.2 is the same as SLR.2, and our data for the multiple regression is still randomly sampled. MLR.3 states that there is no perfect collinearity between any of the dependent variables in our multiple regression.

We can see that there is no perfect collinearity because the correlation coefficient between each of the independent variables is some number other than 1. Figure 7 below shows the correlation coefficients between each of the independent variables.

Fig. 7

Two Independent Variables	Correlation Coefficient
distance and gnppc_im	-0.1440
distance and tariff	0.0908
distance and regulation_ex	-0.0549
distance and services_ex	-0.0767
gnppc_im and tariff	-0.3263
gnppc_im and regulation_ex	-0.0089
gnppc_im and services_ex	-0.0102
tariff and regulation_ex	-0.0437
tariff and services_ex	-0.0615
regulation_ex and services_ex	0.6865

MLR.4 is analogous to SLR.4 in that it states that none of the values of the explanatory variables may contain information about the mean of the unobserved factors. Mathematically, this can be represented by

$$E(u_i \mid x_{i1},\ x_{i2}, ..., x_{ik}) = 0 .$$

The addition of independent variables means that there is less information contained within the unobserved term, so we will over and or underestimate our coefficients by less (our coefficients will be less biased). It is likely that given our independent variables or distance, regulation_ex, and services_ex, we have reduced the bias, but we will likely still not meet this assumption.

Finally, MLR.5 is analogous to SLR.5, as it is the assumptions of homoskedasticity. Again, we will assume homoskedasticity for the model, meaning that the value of the explanatory variables must contain no information about the variance of the unobserved factors.

IV. Results

Fig. 8

Variable	SLR	MLR 1	MLR 2
distance	-0.0002437*** (0.0000102)	-0.0001837*** (9.19e-06)	-0.0001897*** (.0000101)
gnppc_im		0.0771221*** (0.0050395)	0.0684027*** (0.0055535)
tariff		-0.0251048*** (0.0052884)	-0.0363784*** (0.0058226)
regulation_ex		-4.80806*** (0 .4448545)	4.233996*** (0.3545366)
services_ex		15.33822*** 0 .5221551	
_cons	18.47717*** (0.0972789)	10.56795*** (0.2967841)	14.95224*** (0 .2831814)
R-Squared	0.1107	0.3332	0.1872

Statistically Significant: *10% **5% ***1%

For the simple linear regression between distance and ltrade, there is a statistically significant relationship between distance and percent trade. The regression shows that as the distance between two countries' capitals grow by 1 kilometer, trade decreases by 0.02437%. Although this coefficient may seem small, if the distance between two countries' capitals grows by 1,000 kilometers, trade will decrease 24%, nearly a quarter. The R-squared value for the simple linear regression was 0.1107, showing that 11% of the data is covered by the simple linear regression.

The first multiple linear regression features many facets that are factors in world trade. With the GNP per capita of the importing country included, it shows that the wealth of the people of the importing country greatly influences trade as the more wealth a country's citizens have, the more goods and services will be imported. Tariffs also have a great influence on trade too. According to the model, a 1% increase in total tariffs results in a 2.5% decrease in total trade. The model also includes both regulation_ex and services_ex. As seen by the table, both regulation_ex and services_ex have a great impact on trade and are statistically significant, however their coefficients are quite drastic and counterintuitive for the model. The model shows that an increase in transparency/regulation by 0.01 on the index from zero to one results in a loss of -4.81% of trade, and an increase in the services infrastructure of a country by 0.01 on the index results in an increase of 15.33% of trade. When viewing the model, it is rather unintuitive that the less corrupt and more transparent a country is the less they will trade with other countries. The model seemed with become inaccurate due to the issue of possible multicollinearity as the correlation between regulation_ex and services_ex is 0.6865, which may result in inaccurate and drastic coefficients.

In the second multiple regression model, services_ex is removed to better account for issues of multicollinearity. The model better reflects that impact corruption has on trade, showing that a government that is less corrupt will have a greater amount of trade. For a 0.01 increase on the regulation_ex index, trade increases 4.23% for the exporting country. The other independent variables are not too greatly affected by the exclusion of services_ex and are all statistically significant.

As can be seen in Figure 8, each variable in the regression analysis is significant at the 1% level. This can be interpreted to mean that 99% of all possible sample can be expected to include the true population parameter. This statistical significance can be shown using various robustness tests, including t-tests, p-values, and confidence intervals. We will use the following hypotheses to test the significance of each independent variable in MLR 2. In this case, β_k is each of the coefficients, β_1 through β_4 in the model MLR 2.

$$H_0 : \beta_k = 0$$
$$H_1 : \beta_k \neq 0$$

Given the t-value for each independent variable (provided by the Stata output for the multiple linear regression, shown in the appendix), we can determine the statistical significance of the variable by comparing this t-value to the critical value of the t-distribution. In this case, we have 3,947 observations, so we have 3,942 degrees of freedom. We will use a two-tailed test because we are interested to see if the value of the coefficient is significantly different from zero and do not need to differentiate between whether it is greater or less than zero. Thus, the critical value (for large degrees of freedom, a two-tailed test, and 1% level of significance) is 2.576. If the absolute value of our t-value is greater than the critical value, it lies within the rejection region, and we can reject the null hypothesis, thus determining that the variable is statistically significant. Each t-value meets this requirement, as can be seen in Figure 9 below.

We can test the same hypotheses using the p-values of the independent variables. The p-value for the two-sided hypothesis test (again given by the Stata output) for each independent variable is 0.000. The p-value is the smallest significance level at which the null hypothesis is rejected. Thus, because our p-values are all less than 0.01 (namely, 0.000), the p-values also support the statistical significance of the independent variables in MLR 2.

The same conclusion can be drawn using the 99% confidence interval, also shown in Figure 9. At a 99% level, there is only a 1% chance that the value of β_k lies outside of the interval. Thus, we can say that each independent variable in MLR 2 is statistically significant because we can reject the null hypothesis that the associated coefficients equal zero because zero is not in any of the confidence intervals.

Fig. 9

Variable	**Coefficient (Standard Error)**	**t-value**	**p-value**	**99% Confidence Interval**
distance	-0.0001897 (.0000101)	-18.70	0.000	$-0.000216 \leq \beta_1 \leq -0.000164$
gnppc_im	0.0684027 (0.0055535)	12.32	0.000	$0.054097 \leq \beta_2 \leq 0.082709$
tariff	-0.0363784 (0.0058226)	-6.25	0.000	$-0.051377 \leq \beta_3 \leq -0.021379$
regulation_ex	4.233996 (0.3545366)	11.94	0.000	$3.32071 \leq \beta_4 \leq 5.14728$

V. Extensions

The removal of the variable services_ex as an explanatory variable in the multiple linear regression is supported by the results of an F-test to test the joint significance of the two highly correlated variables, services_ex and regulation_ex. To calculate the F-value for this test, we take into account the unrestricted model in which both of these variables are included as independent variables and the restricted model in which neither is included. The R^2 value for the unrestricted model is 0.3332, and the R^2 value for the restricted model is 0.1578. Our hypotheses are as follows.

$$H_0: \beta_4 = \beta_5 = 0$$

$$H_1: null\ hypothesis\ is\ false$$

We then use the following to calculate the F value.

$$F = \frac{(R^2_{ur} - R^2_r)/q}{(1 - R^2_{ur})/(n-k-1)}$$

$$F = \frac{(0.3332 - 0.1578)/2}{(1 - 0.3332)/(3941)} = \frac{0.0877}{0.0001692} = 518.322$$

This value is considerably larger than the critical value of 3.00, meaning that we can reject the null hypothesis that each of the coefficients in consideration is not significantly jointly different from zero. We then say that at a 5% confidence level, the variables services_ex and regulation_ex are jointly significant. Thus, we know that the variables are capturing a similar effect on the dependent variable, ltrade. Therefore, we can drop one variable, and we will drop services_ex. This is shown under the title MLR 2 in the table found in the Results section of this paper. We see here that the value of the coefficient for regulation_ex, when services_ex is dropped from the model becomes positive and is statistically significant at the 1% level.

In order to build a more comprehensive model that more fully explains the volume of trade between two countries and hopefully increases the R-squared value of the regression, we will continue to add independent variables. Namely, we will add two dummy variables that indicate whether or not the two countries in question for any given observation share a border and whether or not the two countries are part of a common trade agreement. The variables are as follows.

Fig. 10

Variable Name	Interpretation	Data Source
dadjacent	Takes on a value of 1 if the two countries for any given observation share a border and 0 if they do not.	World Bank
dtradeagr	Takes on a value of 1 if the two countries for any given observation are part of a common trade agreement and 0 if they are not.	World Bank (dtradeagr=dasean+dnafta+dlaia +daunz+dmercosur+deu)

Adding these two dummy variables to our multiple linear regression model MLR 2 from above, we obtain the following results.

Fig. 11

Variable	MLR 3
distance	-.0001353*** (.0000108)
gnppc_im	.0704556*** (.0054535)
tariff	-.038738*** (.0057099)
regulation_ex	4.281011***

	(.3477791)
dadjacent	1.971074*** (.2538717)
dtradeagr	1.465089*** (.1830582)
_cons	14.27584*** (.2825715)
R-Squared	0.2193

We can make very similar conclusions about each of the independent variables that we also included in MLR 2, and we note that each result is statistically significant at the 1% confidence level. The coefficient for the dummy variable dadjacent is 1.971 and the coefficient for the dummy variable dtradeagr is 1.465. For the variable dadjacent, this means that if two countries share a common border, the volume of trade between them will be 197.107% greater than the base case volume of trade for two countries that do not share a common border, holding all else constant. For the variable dtradeagr, this means that the volume of trade between two countries be 146.509% if they are part of a common trade agreement than it would be if they were not part of a common trade agreement, all else being equal. These are relatively large effects on the dependent variable which are statistically significant at the 1% confidence level. Including the dummy variables also affects the model by increasing the R-squared value. Model MLR 3 explains 21.93% of the variability of the dependent variable trade around its mean. This increase demonstrates the MLR 3 is a better representation of the data than the previous MLR 2 because the coefficient of determination has increased.

VI. Conclusions

The overall findings from the model support the earlier hypotheses concerning the direction and significance of the independent variables on bilateral trade flows. As the distance between the capitals of two countries increases by one kilometer, trade decreases by .0244%. This value was significantly significant and acted in the predicted direction. Distance in particular was interesting to us because despite significant advances in shipping and communication technology, physical distance is still a barrier to trade between countries. As the wealth and size of the importing country increased, the percentage of trade predictably increased as well. Therefore, the two variables included in the gravity model acted in the way we expected them to. In model MLR 2, which excludes the exporting service sector due to multicollinearity issues, GNP of the importing country, tariffs, and regulation of the exporting country were all statistically significant at the 1% level and acted in the direction predicted in our hypothesis. With the addition of the dummy variables, adjacency of the two countries and common trade agreements, both were found to also be statistically significant at the 1% level and acted in the predicted direction: two adjacent countries had higher levels of bilateral trade flow, as did two countries that participated in a common trade agreement.

With the overabundance of factors that potentially influence trade between two countries, it's difficult to narrow down the specific effects, but our model focuses on variables related not only to the gravity model but also to potential policy measures governments could undertake. While distance between capitals is not something any country necessarily has control over, the amount of tariffs imposed is a policy decision, as is negotiating future trade agreements. By gaining a better understanding of what significantly affects bilateral trade, it's possible to make more intentional, strategic choices that can improve a country's welfare and global position.

References

Chaney, T. (2013). The Gravity Equation in International Trade: An Explanation. *Journal of Political Economy, 126*, 150-177. doi:10.3386/w19285

Lampe, M., & Sharp, P. (2012). Tariffs and income: A time series analysis for 24 countries. *Cliometrica Economic Journal, 7*(3), 207-235. doi:10.1007/s11698-012-0088-5

Mack, A., Martinez-Garcia, E., & Martinez-Garcia, M. T. (2014). Speaking of Trade: Quantifying the Contribution of Multilingualism to Overcome the Language Barriers to Trade. *American Economic Association*.

Appendix

```
. regress ltrade distance

      Source |       SS           df       MS      Number of obs   =     4,557
-------------+----------------------------------   F(1, 4555)      =    567.25
       Model |  6709.75358         1  6709.75358   Prob > F        =    0.0000
    Residual |  53879.0465     4,555  11.8285503   R-squared       =    0.1107
-------------+----------------------------------   Adj R-squared   =    0.1105
       Total |     60588.8     4,556  13.2986831   Root MSE        =    3.4393

------------------------------------------------------------------------------
      ltrade |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    distance |  -.0002437   .0000102   -23.82   0.000    -.0002638   -.0002237
       _cons |   18.47717   .0972789   189.94   0.000     18.28646    18.66789
------------------------------------------------------------------------------
```

```
. regress ltrade distance gnppc_im tariff regulation_ex services_ex

      Source |       SS           df       MS      Number of obs   =     3,947
-------------+----------------------------------   F(5, 3941)      =    393.93
       Model |  15672.4038         5  3134.48077   Prob > F        =    0.0000
    Residual |   31358.423     3,941  7.95697106   R-squared       =    0.3332
-------------+----------------------------------   Adj R-squared   =    0.3324
       Total |  47030.8268     3,946  11.9186079   Root MSE        =    2.8208

------------------------------------------------------------------------------
      ltrade |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    distance |  -.0001837   9.19e-06   -19.98   0.000    -.0002017   -.0001656
    gnppc_im |   .0771221   .0050395    15.30   0.000     .0672419    .0870023
      tariff |  -.0251048   .0052884    -4.75   0.000    -.0354731   -.0147364
regulation_ex|   -4.80806   .4448545   -10.81   0.000    -5.680227   -3.935893
 services_ex |   15.33822   .5221551    29.37   0.000      14.3145    16.36194
       _cons |   10.56795   .2967841    35.61   0.000     9.986082    11.14981
------------------------------------------------------------------------------
```

```
. regress ltrade distance gnppc_im tariff regulation_ex

      Source |       SS           df       MS      Number of obs   =     3,947
-------------+----------------------------------   F(4, 3942)      =    227.05
       Model |  8806.48619         4  2201.62155   Prob > F        =    0.0000
    Residual |  38224.3406     3,942  9.69668712   R-squared       =    0.1872
-------------+----------------------------------   Adj R-squared   =    0.1864
       Total |  47030.8268     3,946  11.9186079   Root MSE        =     3.114

------------------------------------------------------------------------------
      ltrade |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    distance |  -.0001897   .0000101   -18.70   0.000    -.0002096   -.0001698
    gnppc_im |   .0684027   .0055535    12.32   0.000     .0575147    .0792907
      tariff |  -.0363784   .0058226    -6.25   0.000     -.047794   -.0249628
regulation_ex|   4.233996   .3545366    11.94   0.000     3.538904    4.929089
       _cons |   14.95224   .2831814    52.80   0.000     14.39705    15.50744
------------------------------------------------------------------------------
```

Countries Used in Research:

Argentina, Australia, Austria, Bangladesh, Belgium, Bolivia, Brazil, Bulgaria, Canada, Chile, China, Colombia, Costa Rica, Czech Republic, Denmark, Dominican Republic, Ecuador, Egypt, El Salvador, Estonia, Finland, France, Germany, Greece, Guatemala, Honduras, Hong Kong, Hungary, Iceland, India, Indonesia, Ireland, Israel, Italy, Jamaica, Japan, Jordan, South Korea, Latvia, Lithuania, Malaysia, Mauritius, Mexico, Netherlands, New Zealand, Nicaragua, Nigeria, Norway, Panama, Paraguay, Peru, Philippines, Poland, Portugal, Romania, Russia, Singapore, Slovak Republic, Slovenia, South Africa, Spain, Sri Lanka, Sweden, Switzerland, Taiwan, Thailand, Trinidad and Tobago, Turkey, Ukraine, United Kingdom, United States, Uruguay, Venezuela, Vietnam, Zimbabwe.