# Differing Trade Elasticities for Intra- and International Distances: a Gravity Approach

Jason Query*

October 13, 2014

**Abstract**

Using the gravity model of trade, I estimate the impact of the internal distance a traded good travels within its originating country on international trade levels. Combining multiple data sets, I create a measure for the distance a U.S. produced good travels before leaving the country. Using potential, rather than actual, production of agricultural goods as an instrument, I find that internal distance is statistically significant and large in magnitude. A 10% reduction in the distance a good travels within the exporting country increases trade by roughly 16%.

*Keywords*: Trade, Distance, Gravity Model, Intranational Trade Costs
*JEL classification*: F14, L90

*Department of Economics, University of Oregon, Eugene, OR, 97403-1285, USA, jquery@uoregon.edu

# 1 Introduction

International trade economists have long been interested in trade costs. In recent years, as artificial trade barriers, such as tariffs and quotas, fall to low levels, trade economists have become more interested in transportation costs as a trade barrier. Many trade models proxy for these trade costs, which are often times difficult to measure, with distances between countries. While the distance between countries, which I call external distance, functions well as a proxy for trade costs (and has since Tinbergen 1962), it does not give an explicit explanation of trade costs. In addition, it fails to capture anything not correlated with external distance.

Because of this limitation, there is a significant literature that attempts to explicitly define trade costs (see for example Anderson and van Wincoop 2004). Transportation costs can be anything from actual shipping costs to time delays associated with shipping (Hummels and Schaur 2013) to uncertainty associated with maritime piracy (Burlando et al. 2014). The trade costs literature includes examinations of free trade agreements (Baier and Bergstrand 2007), culture (Rauch and Trindade 2002), historical and political costs (Head et al. 2010), and the border effect (McCallum 1995, Anderson and van Wincoop 2003) among many others.

One potentially important trade cost that has received little attention until recently is the costs incurred in trading before a good leaves the country of origin, which I call internal costs (see Agnosteva et al. 2014). One probable avenue in which internal costs are important is that a firm located in a country with high internal costs is at a competitive disadvantage compared to those firms that can move a good through its location cheaper. While we have many estimates for the effect of external distance on trade, we know very little about the magnitude of internal distance at the importer-commodity level.

Little research has been done in this area because internal trade costs and internal distance can be both extremely difficult to measure and difficult to estimate. No comprehensive data set exists that allows a researcher to perfectly track the movements of a traded good within the country of origin. To overcome this obstacle, I combined two data sets, a data set including by-commodity exports at the U.S. port level and a data set with by-state agricultural production, to create a unique weighted-average measure of internal distance.

The impact of internal distance on trade is difficult to estimate because it is plausible that internal distance is endogenous to trade levels. *Ceteris paribus*, firms which export heavily will tend to locate production closer to the port of export as a means of reducing internal trade costs, biasing the estimate of the impact of internal distance. To alleviate this concern, I limit my sample to agricultural goods, which are constrained in production location by climate and soil factors. To further alleviate endogeneity concerns, I develop an instrumental variable strategy, instrumenting actual agricultural production with the Food and Agricultural Organization's (FAO) Global Agro-Ecological Zone project's suitability index. This index essentially ranks the ability of each state to grow a given agricultural good. The measure is created by the FAO using historical measures of climate and soil, measures which are independent of U.S. trade patterns.

I find that the internal distance elasticity of trade is statistically significant and large in magnitude, having a larger impact on trade flows than external distance. I find that, using conservative estimates, a 10% decrease in the distance a good must travel before leaving the United States would increase the exports of that good by 16%. These findings have potentially significant policy implications, particularly with policy makers' decisions to fund internal infrastructure.

This paper proceeds as follows. Section two of this paper gives a brief review of the relevant literature. Section three outlines the empirical specifications and section four details the data used in this paper. Section five details the results. Section six concludes.

## 2 Literature Review

This paper contributes to a key literature in international trade which focuses on gravity models of trade and trade costs. Trade economists have long attempted to explain what exactly the trade cost component in the gravity model should properly consist of. Authors have examined the border effect, or the notion that regions are more likely to trade with other regions within their country as compared to international regions. Examples of these papers include McCallum (1995), which found the border effect between Canada and the U.S., Anderson and van Wincoop (2003), which outlined the theoretical justification and an empirical methodology for properly estimating the border effect, and Query (2014) which shows that the border effect is smaller for importers with greater GDP. Many papers have examined the potential trade-encouraging effects of currency and trade unions. Glick and Rose (2002) finds that joining a currency union nearly doubles trade between the sharing countries. Head et al. (2010) finds that strong colonial ties can boost trade. Blonigen and Wilson (2008) and Clark et al. (2004) show that the efficiency of a country's ports significantly impact trade. Most gravity-based papers examine between-country trade costs. In this paper, I further attempt to understand how trade costs drive trade flows, but I am examining the costs a trading firm incurs before a good leaves the country.

Two papers that are directly related to this paper are Agnosteva et al. (2014) and Cosar and Fajgelbam (2014). A recent paper, Agnosteva et al. (2014) outlines a methodology for measuring intra-national border barriers and intra-regional trade costs. The paper finds that intra-regional trade costs are an important consideration in conducting comparative statics. The paper notes the importance of future research "exploring the connection between intra-regional and inter-regional trade costs." Cosar and Fajgelbaum (2014) develops a model which demonstrates that costly trade leads production to move to areas with easy access to foreign ports. Reduction in trade costs results in migration to coastal areas. This finding is supported with data showing that U.S. export-oriented industries are more likely to be located near international ports. This finding is important to the endogeneity discussion in this paper.

Many previous papers examined internal distance, but these papers typically refer to how far a good bought and sold in the same country travels; they do not consider the effect on related international trade flows. For a discussion of the how various papers measure this type of internal distance, see Head and Mayer (2002). A few papers do examine internal distance in a similar fashion to this paper. Blonigen and Wilson (2006) estimates a gravity model which includes inland transport prices, as well as inland transport distances, though the latter is included in a market potential variable so no direct internal distance elasticity is estimated. Malchow and Kanafani (2004) uses the distance a good would have to travel before leaving a port to estimate the internal distances effect on port choice, though the paper does not look at trade level effects. While containing no direct measure of distance, Volpe Martincus and Blyde (2013) finds that firms in Chile which experienced a shock to their transportation network saw a decrease in the total value of exports. However, because this paper does not contain a measure of internal distance it does not estimate an internal distance elasticity. Cosar and Demir (2014) uses internal distance measures to calculate the remoteness of Turkish provinces. The authors demonstrate that improved road infrastructure results in increased trade levels. Atkin and Donaldson (2014) uses price gaps to show that intranational trade costs are significantly larger in Ethiopia and Nigeria than in the United States. For a review of literature related to transportation infrastructure, see Redding and Turner (2014).

# 3 Empirical Methodology

## 3.1 Gravity Model

The model used to econometrically estimate the elasticities associated with internal and external distance is the gravity model of trade pioneered by Tinbergen (1962) and given theoretical justification by Anderson (1979). The typical log-linearized specification for trade is given by:

$$\ln X_{ijk} = \beta_1 + \beta_2 \ln Y_i + \beta_3 \ln Y_j + \beta_4 Z_{ijk} + \alpha_i + \alpha_j + \alpha_k + \epsilon_{ijk} \tag{1}$$

where $X_{ijk}$ is the trade in product $k$ exported by country $i$ to country $j$, $Y_i$ is the GDP of country $i$, $Z_{ijk}$ is a vector of explanatory variables, $\alpha_i$, $\alpha_j$, and $\alpha_k$ are importer, exporter, and product fixed effects, and $\epsilon_{ijk}$ is an i.i.d. error term with mean zero and variance one. As outlined in in Anderson and van Wincoop (2003), this specification should include multilateral resistance terms for both the importer and exporter, which Feenstra (2002) shows can be accounted for using importer and exporter fixed effects. Theoretically, $\beta_2$ and $\beta_3$ should be equal to one.[1] As such, the gravity equation can be rewritten as:

$$\ln \frac{X_{ijk}}{Y_i Y_j} = \beta_1 + \beta_4 Z_{ijk} + \alpha_i + \alpha_j + \alpha_k + \epsilon_{ijk} \tag{2}$$

Variables often included in $Z_{ijk}$ are the distance between $i$ and $j$, whether $i$ and $j$ share a common language, whether $i$ and $j$ border each other, and whether $i$ and $j$ share a common colonial tie. For this paper, I include the distance between country $i$ and $j$ as well as adding a new variable, the distance good $k$ travels within country $i$ before being exported to country $j$.

There are a few complications that lead to modifications of equation 2 for my data set. First, because the only exporter in my data set is the United States, $Y_i$ and $\alpha_i$ cannot be seperarely identified from the regression's constant.[2] In addition, because the external distance between the United States and a given country is fixed, I cannot include both a measure of external distance and the importer fixed effect $\alpha_j$. To allow for a comparison of the magnitude between internal and external distance elasticities, I will exclude $\alpha_j$ from my specifications. This could lead to a potential bias in my specification, as the importer multilateral resistance term is not fully accounted for and is potentially correlated with included variables. In addition, I allow the importer income elasticity of trade to differ from

[1] However, as noted in my result section, when I allow the income elasticity to differ from one, I find an income elasticity significantly lower than one.

[2] The main importance of the exporter fixed effect is to account for country specific effects including that country's multilateral resistance term, but because there is only one exporter, this fixed effect simply gets subsumed into the constant without generating any bias.

one.

One potential issue with omitting importer fixed effects is that importer-specific factors that impact trade which are typically subsumed into the importer fixed effect are not accounted for.[3] These omitted variables are unlikely to be correlated with internal distance but may bias coefficient estimates, especially with regards to external distance. As a result, regressions I estimate in this paper will be a variation of the following equation:

$$\ln X_{jk} = \beta_1 + \beta_2 \ln Y_j + \beta_3 \ln EXTDIST_j + \beta_4 \ln INTDIST_{jk} + \beta_5 COMLANG_j \qquad (3)$$
$$+\beta_6 AREA_j + \beta 7 COMCURRj + \beta_8 RTA_j + \beta_9 CONTIG_j + \alpha_k + \epsilon_{jk}$$

where $X_{jk}$ is the value of commodity $k$ exported from the United States to country $j$, $Y_j$ is the GDP of country $j$, $EXTDIST_j$ is the external distance between the United States and country $j$, $INTDIST_{jk}$ is the internal distance of commodity $k$ before being exported to country $j$, $COMLANG_j$ is a dummy variable indicating if the United States and country $j$ share an official language, $AREA_j$ is the geographic area of country $j$, $COMCURRj$ is a dummy variable taking the value of one if the United States and country $j$ share a common currency, $RTA_j$ is a dummy variable indicating if the United States and country $j$ are in a regional trade agreement, $CONTIG_j$ is a dummy variable taking the value of one if the United States and country $j$ have a contiguous border, and $\alpha_k$ is a commodity fixed effect.

[3] Any U.S.-specific trade determinants will be in the constant term $\beta_1$ as the U.S. is the only exporter.

## 3.2 Accounting for Production Location Endogeneity

Within-country distance is likely endogenous. It is possible that producers, at least to some extent, locate their production as close to their customers as is feasible. Thus, producers of exported commodities likely move closer to the coasts, limiting the within-country distance traveled (see, for example, Cosar and Fajgelbaum 2012). To accommodate for this potential endogeneity, I limit my data to examining agricultural goods. Because of the nature of agricultural goods, production is limited to a specific area where the climate and soil are suited for growing a crop. As such, agricultural goods are less likely to be subject to this endogeneity concern.

Measuring the distance a good travels within the exporting country before leaving a port is difficult. One potential problem is that a given good-importer pair will almost certainly not originate from the same location. Countries which import corn may get corn from both Indiana and Iowa, for example. In addition, many goods travel through multiple ports for the same importer. Of the 1,175 country-commodity pairs available in my port trade data set, a mere 421 go through only one port. The mean number of ports a country-commodity pair goes through is 3.7, and the most ports a commodity-pair goes through is 48, which is Canada's corn. As a non-Canadian example, Japan's imports of soybeans went through 28 different U.S. ports in 2007.

In this paper, I generate a measure of the internal distance of agricultural goods for a given importer. Due to data limitations, this measure is limited to exports from the United States. I use state shares of agricultural production, as well as port shares in agricultural exports to a given importer, to generate a weighted average of all distances a good can travel before leaving the United States for its final destination. The measure of internal distance for a given commodity and importer pair, $INTDIST_{jk}$, is given by:

$$INTDIST_{jk} = \sum_{s} \sum_{p} \frac{X_{jkp}}{\sum_{p'} X_{jkp'}} \frac{P_{ks}}{\sum_{s'} P_{ks'}} d_{sp} \tag{4}$$

where $j$ is the importer, $k$ is the commodity, $s$ is a given state, $p$ is a given port, $X_{jkp}$ is country $j$'s imports of commodity $k$ through port $p$, $P_{ks}$ is the total production of commodity $k$ in state $s$, and $d_{sp}$ is the distance from state $s$ to port $p$. By constructing a weighted average measure of internal distance, I place a higher importance on ports frequently used for a given importer-commodity pair and place little weight on infrequently used ports. In addition, I place a higher weight on states with larger volumes of production of a given commodity and little weight on state's that grow very little.

Table 1 includes summary statistics for $INTDIST_{jk}$ by commodity. Because the United States is one of the largest countries in the world, by area, it is not surprising to see such large internal distances for the U.S. Only one good (Tobacco) travels less than 1000 miles, on average, to a port before being shipped to a country. Tomatoes and Potatoes tend to travel the furthest within the United States before being exported.

This measure is still subject to endogeneity concerns. Specifically, it could be the case that crop production locations are determined by the trade demand for these goods. Producers potentially choose to grow crops as close as possible to the ports their crops will ship through, even if that location is not the most ideal location for production. To correct for this possible endogeneity, I create an instrument for $INTDIST_{jk}$, $IV_{jk}$, which uses the FAO Global Agro-Ecological Zone project's crop suitability index to instrument for actual state agricultural production. The suitability index uses comprehensive soil and climate data to determine how suitable the land of a given U.S. state is for growing a given crop. Further details on construction of this index is in the data section of this paper. The crop suitability index is based only a state's climate and soil conditions and thus is independent of the trade process. Thus, I construct $IV_{jk}$ as follows:

$$IV_{jk} = \sum_s \sum_p \frac{X_{jkp}}{\sum_{p'} X_{jkp'}} \frac{PP_{ks}}{\sum_{s'} PP_{ks'}} d_{sp} \quad (5)$$

where all variables previously described are the same and $PP_{ks}$ is the crop suitability from the Global Agro-Ecological Zone project. This measure is used to instrument for $INTDIST_{jk}$ in a two-stage least squares process.

# 4 Data

The data used in this paper come from a variety of sources. The trade data come from the UN Comtrade database. Comtrade provides trade data for a variety of years and levels of aggregation. In this paper, I use 4-digit harmonized system 2007 data for the year 2007 and 4-digit harmonized system 1992 data for years 1992, 1997, 2002, and 2007. For most specifications, I use the HS2007 data as it is the most detailed data available in 2007. Unfortunately, HS2007 data does not exist prior to 2007, so for the year-by-year comparison I use HS1992 even for the 2007 estimation. The Comtrade data does not report which port a good went through, information that is necessary in constructing my internal distance measure. To handle this, I use the U.S. port to foreign country trade data for the year 2007 available from the U.S. Census Bureau's USA Trade Online, which is provided at the 4-digit harmonized system level.

Agricultural data comes from two main sources. Actual agricultural production for the U.S. comes from the 2007 Agricultural Census. The Agricultural Census has total production by state for the selection of agricultural products listed in Table 2. A measure of how good a state is at producing a crop, the crop suitability index, comes from the Food and Agricultural Organization Global Agro-Ecological Zone (GAEZ). The crop suitability index is a value between zero and one hundred and is derived from the GAEZ model which estimates the potential crop yield of each state's land using average climate data from 1961 to 1990 as well as data on soil resources and terrain-slope conditions. The index is estimated separately for differing input levels as well as source of water. In this paper, I use the intermediate input level, rain fed suitability index. The FAO GAEZ data have previously been used by Costinot and Donaldson (2011), using both actual and potential production of agricultural goods to measure gains from economic integration. In addition, it was used in Nunn and Qian (2011) to estimate the potential productivity of potatoes in Europe, allowing them to estimate the impact of potatoes on population growth.

State-to-port distances are calculated "as the crow flies," using longtitude and latitude data for each state and port. The location data for each state comes from Google Maps,[4] and is measured at the center of the city with the largest population according to the 2000 U.S. Census. Port location data comes from the U.S. Army Corps of Engineers' Navigation Data Center[5]. Gravity variables come from the CEPII Gravity dataset (see Head et al. 2010).

[4]Found using http://www.mapcoordinates.net/en.

[5]The data can be found at http://www.navigationdatacenter.us/ports/ports.asp.

# 5 Results

## 5.1 Internal distance elasticity

To generate estimates of the various distance elasticities, I estimate variations of equation 3 using basic OLS and the IV strategy outlined above. Table 3 reports these estimations using the HS2007 data for the year 2007. Column one is equation 3 estimated using the variable $INTDIST_{jk}$ without any instrumental or control variables, while column two excludes control variables but uses 2SLS to instrument for $INTDIST_{jk}$ with $IV_{jk}$. Column three adds the importer-specific control variables using OLS. Column four uses both importer-specific controls and the IV strategy.

Both columns one and two have an external distance elasticity near unity in magnitude. The distance elasticity of trade being close to one makes intuitive sense, but this result is not often found in empirical estimations. Columns three and four have an external distance elasticity between -0.445 and -0.465, smaller in magnitude than the results in column one and two. This indicates that while transportation costs have a significant impact on trade, non-transportation costs such as ease of communication (proxied for by the common language dummy) also have a significant impact on trade and are highly correlated with between-country distance. As expected, the contiguity dummy is large in magnitude, though it is only statistically significant in the IV regression. The regional trade agreement dummy is statistically significant and large in magnitude in both regressions. The other control variables are not statistically significant in either specification.

In all specifications, the coefficient on the log of internal distance is negative, ranging from -1.669 to -2.002, and of a larger magnitude than external distance, which ranges from

-0.445 to -1.068. The last row of table 3 reports the p-value of the test that the coefficient on internal distance is equal to the coefficient on external distance. Note that in the non-control regressions I cannot reject the hypothesis that the coefficient on the log of internal distance is equal to the coefficient on the log of external distance at the standard significance levels. However, in my most-preferred specification which is reported in column four, I can reject this hypothesis at the 5% significance level. A 10% decrease in internal distance would cause a 16-20% increase in trade. This effect is similar to, and potentially larger than, the positive impact on trade than a 10% decrease in the distance between countries, which would only increase trade by 4%-11%.

Because $EXTDIST_j$ is simply a measure of distance between two points while $INTDIST_{jk}$ is a weighted average measure of distances, I report the standardized coefficients for the regressions in Table 4. Using column four as an example, these results say that a one standard deviation decrease in internal distance will increase trade in a given product by .234 standard deviations; a one standard deviation decrease in external distance will only increase trade by .075 standard deviations. The results here are similar to those of Table 3. The standardized coefficients on internal distance are always larger in magnitude than the standardized coefficients on external distance.

This result has potentially important policy implications. Artificial trade barriers, such as tariffs and quotas, are becoming less relevant as policies such as free trade agreements push these barriers toward zero. As such, if policy makers look to reducing non-artificial barriers as a means to promote trade, these results indicate that internal trade costs are a potentially fruitful avenue for police makers to pursue. Not only will improving domestic trade infrastructure be an effective tool in trade promotion, it also creates important externalities for the country. A more efficient transportation network in the United States will not only have a strong impact on the U.S.'s international trade but will also encourage more domestic trade. In addition, domestic investment is more politically palatable than "international" investment.

Finally, Table 3 shows an income elasticity of trade that is different from the one dictated by theory. In fact, the income elasticity of trade is around .55 in table 3. This pattern also holds true in table 5. While the income elasticity of trade varies year-by-year, going as low as .437 in 1992, it hovers around the .55 mark for most years and is never close to one.

I now explore using the potential production index to generate IV estimates. The first-stage F-stat for the instrumental variable is reported in columns 2 and 4. The F-stat is greater than 400, a strong indication that my instrument is correlated with internal distance. The difference between columns one and two and the difference between columns three and four are both negligible. The IV-estimated coefficient is always within one standard deviation of the OLS estimation. The similarity between the OLS and IV estimates is evidence that farmers are not locating their farms based on distance to the port of export. Instead, they are potentially choosing production location based on some other factors, such as soil and climate quality or distance to domestic customers.

## 5.2 Internal distance over time

One natural question that comes from this result is: does this result hold for years besides 2007? To test this, I estimate equation 3 using HS1992 data for the years 1992, 1997, 2002, and 2007.[6] These results can be found in Table 5. For the years 1992 and 2007, the trade elasticity of internal distance is about 1.5-1.6%, similar to those found in Table 3. In magnitude, the internal distance elasticity is statistically significant and greater than the external distance elasticity for all years, though the degrees to which they are different vary greatly. For 1992, the coefficients on the different distance measures are different at almost the 90% level while for the years 1997, 2002, and 2007 the coefficients on internal and external distance do not appear to be significantly different. This indicates that international trade is indeed as responsive, if not more responsive, to internal distance than external distance.

Table 5 has another interesting result to note. The external distance elasticity was significantly smaller in 1992 than it is in 2007. In fact, external distance in 1992 has less than half the impact it does in 2007. In addition, the impact of external distance is monotonically increasing in time. A plausible explanation for this result could be the United States moving towards exporting those agricultural goods that cost more to ship via boat.

# 6 Conclusion

In this paper, I estimate the elasticity of trade associated with transportation costs that occur within the exporting county. I accomplish this by generating a measure of the distance various goods must travel to a port before being exported. I find that trade barriers occuring within the exporting country have a significant impact on trade flows; the impact of these barriers are potentially more important the external trade barriers.

In order to properly estimate the internal distance elasticity of trade, I combine multiple

[6] Note that this means the 2007 results in Table 3, which uses HS2007 instead of HS1992 data, and Table 5 will differ.

data sets to arrive at a weighted average measure of the distance a good must travel before leaving the United States. I use state production values and port export data to get an estimate of internal distance. I then include this estimate in a gravity model of trade as a proxy for intranational trade costs.

The location of production for a good may be endogenous to trade levels. I accommodate this potential endogeneity in two key ways. First, I limit my sample to only include agricultural products. Because agricultural products need the correct soil and climate conditions to grow, producers have less options when it comes to production location. To further alleviate this, I use the potential of state land to grow a good as an instrument for actual state production.

I find that internal distance is statistically significant and large in magnitude. In my most conservative estimate, a 10% reduction in internal distance would result in a 16.49% increase in international trade. The IV regression coefficients are similar in magnitude to their OLS counterparts. As such, their is no evidence that agricultural producers are selecting their farm locations based on their international trading partners.

# References

[1] Delina E. Agnosteva, James E. Anderson, and Yoto V. Yotov. Intra-national trade costs: Measurement and aggregation. Working Paper 19872, National Bureau of Economic Research, January 2014.

[2] James E Anderson. A theoretical foundation for the gravity equation. *The American Economic Review*, 69(1):106–116, 1979.

[3] James E Anderson and Eric Van Wincoop. Gravity with gravitas: A solution to the border puzzle. *American Economic Review*, 93(1):170, 2003.

[4] James E Anderson and Eric Van Wincoop. Trade costs. Technical report, National Bureau of Economic Research, 2004.

[5] David Atkin and Dave Donaldson. Who is getting globalized? the size and implications of intranational trade costs. Technical report, Mimeo, MIT, 2014.

[6] Scott L Baier and Jeffrey H Bergstrand. Do free trade agreements actually increase members' international trade? *Journal of international Economics*, 71(1):72–95, 2007.

[7] Scott L Baier and Jeffrey H Bergstrand. Bonus vetus ols: A simple method for approximating international trade-cost effects using the gravity equation. *Journal of International Economics*, 77(1):77–85, 2009.

[8] Bruce A Blonigen and Wesley W Wilson. International trade, transportation networks, and port choice. *http://www. nets. iwr. usace. army. mil/docs/PortDevInternalTransport/PortChoice114. pdf*, 2006.

[9] Bruce A Blonigen and Wesley W Wilson. Port efficiency and trade flows*. *Review of International Economics*, 16(1):21–36, 2008.

[10] Alfredo Burlando, Anca Cristea, and Logan M Lee. The trade consequences of maritime insecurity: Evidence from somali piracy. Mimeo, 2014.

[11] Ximena Clark, David Dollar, and Alejandro Micco. Port efficiency, maritime transport costs, and bilateral trade. *Journal of development economics*, 75(2):417–450, 2004.

[12] A Kerem Cosar and Banu Demir. Domestic road infrastructure and international trade: Evidence from turkey. 2014.

[13] A Kerem Cosar and Pablo D Fajgelbaum. Internal geography, international trade, and regional outcomes. *University of Chicago and UCLA, unpublished mimeo*, Mimeo, 2014.

[14] Arnaud Costinot and Dave Donaldson. How large are the gains from economic integration? theory and evidence from us agriculture, 1880-2002. Technical report, 2011.

[15] Jonathan Eaton and Samuel Kortum. Technology, geography, and trade. *Econometrica*, 70(5):1741–1779, 2002.

[16] FAO and IIASA. Global Agro-ecological Zones (GAEZ v3.0). FAO, Rome, Italy and IIASA, Laxenburg, Austria. 2010.

[17] Robert C Feenstra. Border effects and the gravity equation: consistent methods for estimation. *Scottish Journal of Political Economy*, 49(5):491–506, 2002.

[18] Reuven Glick and Andrew K Rose. Does a currency union affect trade? the time-series evidence. *European Economic Review*, 46(6):1125–1151, 2002.

[19] Keith Head, Thierry Mayer, and John Ries. The erosion of colonial trade linkages after independence. *Journal of International Economics*, 81(1):1–14, 2010.

[20] David Hummels and Georg Schaur. Time as a trade barrier. *The American Economic Review*, 103(7):2935–59, 2013.

[21] Matthew B Malchow and Adib Kanafani. A disaggregate analysis of port selection. *Transportation Research Part E: Logistics and Transportation Review*, 40(4):317–337, 2004.

[22] John McCallum. National borders matter: Canada-us regional trade patterns. *The American Economic Review*, pages 615–623, 1995.

[23] Nathan Nunn and Nancy Qian. The potatos contribution to population and urbanization: Evidence from a historical experiment. *The Quarterly Journal of Economics*, 126:593–650, 2011.

[24] Jason A. Query. The impact of market size on international trade and the border effect. Mimeo, 2014.

[25] James E Rauch and Vitor Trindade. Ethnic chinese networks in international trade. *Review of Economics and Statistics*, 84(1):116–130, 2002.

[26] Stephen J Redding and Matthew A Turner. Transportation costs and the spatial organization of economic activity. *Chapter in progress for the Handbook of Regional and Urban Economics*, 2014.

[27] J. Tinbergen. *Shaping the world economy: Suggestions for an international economic policy*. Twentieth Century Fund New York, 1962.

[28] Christian Volpe Martincus and Juan Blyde. Shaky roads and trembling exports: Assessing the trade effects of domestic infrastructure using a natural experiment. *Journal of International Economics*, 90(1):148–161, 2013.

Table 1: Summary Statistics for Internal Distance by Product

| | mean | max | min | sd | count |
|---|---|---|---|---|---|
| Barley | 1706 | 2217 | 1526 | 307 | 7 |
| Grain | 1006 | 2644 | 606 | 570 | 50 |
| Maize | 1448 | 3486 | 840 | 452 | 96 |
| Oats | 1797 | 2567 | 1094 | 463 | 19 |
| Potatoes | 2148 | 3121 | 1561 | 564 | 29 |
| Rice | 1592 | 3536 | 859 | 642 | 99 |
| Soy | 1459 | 2655 | 932 | 408 | 44 |
| Tobacco | 898 | 3526 | 510 | 555 | 41 |
| Tomatoes | 2648 | 3252 | 821 | 857 | 16 |
| Wheat | 1628 | 2162 | 1107 | 273 | 78 |
| Total | 1515 | 3536 | 510 | 630 | 479 |

All numbers are reported in miles.

Table 2: Products Included in Sample

| HS Code | Description | Non-IV Sample | IV Sample |
|---|---|---|---|
| 0701 | Potatoes (except sweet potatoes), fresh or chilled | Yes | Yes |
| 0702 | Tomatoes, fresh or chilled | Yes | Yes |
| 0708 | Leguminous vegetables, shelled or not, fresh or chilled | Yes | No |
| 0805 | Citrus fruit, fresh or dried | Yes | No |
| 0806 | Grapes, fresh or dried | Yes | No |
| 0808 | Apples, pears, and quinces, fresh | Yes | No |
| 0809 | Apricots, cherries, peaches, plums & sloes, fresh | Yes | No |
| 1001 | Wheat and meslin | Yes | Yes |
| 1003 | Barley | Yes | Yes |
| 1004 | Oats | Yes | Yes |
| 1005 | Corn (maize) | Yes | Yes |
| 1006 | Rice | Yes | Yes |
| 1007 | Grain sorghum | Yes | Yes |
| 1201 | Soybeans, whether or not broken | Yes | Yes |
| 1202 | Peanuts (ground-nuts), raw | Yes | No |
| 1206 | Sunflower seeds, whether or not broken | Yes | No |
| 1214 | Rutabagas, hay, clover & other forage products | Yes | No |
| 2401 | Tobacco, unmanufactured | Yes | Yes |

Table 3: Internal Distance Gravity Regressions

| | (1)<br>Basic Reg, No IV | (2)<br>Basic Reg, 2SLS | (3)<br>Controls Reg, No IV | (4)<br>Controls Reg, 2SLS |
|---|---|---|---|---|
| Log of External Distance | -1.053** | -1.068*** | -0.445 | -0.465* |
| | (0.382) | (0.211) | (0.415) | (0.262) |
| Log of Internal Distance | -1.936*** | -1.669*** | -2.002*** | -1.754*** |
| | (0.479) | (0.504) | (0.518) | (0.497) |
| Log of Importer GDP | 0.533*** | 0.538*** | 0.431*** | 0.433*** |
| | (0.086) | (0.057) | (0.074) | (0.073) |
| Contiguity Dummy | | | 1.796 | 1.772** |
| | | | (1.039) | (0.759) |
| Common Language Dummy | | | -0.294 | -0.318 |
| | | | (0.274) | (0.272) |
| Log of Importer Area | | | 0.012 | 0.014 |
| | | | (0.087) | (0.064) |
| Colonial Dummy | | | 0.411 | 0.408 |
| | | | (0.564) | (0.475) |
| RTA Dummy | | | 1.010* | 1.002*** |
| | | | (0.529) | (0.337) |
| Common Currency Dummy | | | 0.476 | 0.467 |
| | | | (0.565) | (0.607) |
| Fixed Effects Used | Commodity | Commodity | Commodity | Commodity |
| Observations | 479 | 479 | 479 | 479 |
| $R^2$ | 0.241 | 0.240 | 0.274 | 0.274 |
| First-stage F-stat | | 400.859 | | 403.439 |
| P-value of test $\ln intdist \neq \ln dist$ | 0.272 | 0.292 | 0.100 | 0.031 |

* p < 0.1, ** p < 0.05, *** p < 0.01. Robust standard errors in parentheses.

Table 4: Internal Distance Gravity Regressions: Standardized Coefficients

| | (1)<br>Basic Reg, No IV | (2)<br>Basic Reg, 2SLS | (3)<br>Controls Reg, No IV | (4)<br>Controls Reg, 2SLS |
|---|---|---|---|---|
| Log of External Distance | -0.183** | -0.186*** | -0.077 | -0.081* |
| | (0.382) | (0.211) | (0.415) | (0.262) |
| Log of Internal Distance | -0.257*** | -0.222*** | -0.266*** | -0.233*** |
| | (0.479) | (0.504) | (0.518) | (0.497) |
| Log of Importer GDP | 0.349*** | 0.352*** | 0.282*** | 0.283*** |
| | (0.086) | (0.057) | (0.074) | (0.073) |
| Contiguity Dummy | | | 0.109 | 0.108** |
| | | | (1.039) | (0.759) |
| Common Language Dummy | | | -0.043 | -0.046 |
| | | | (0.274) | (0.272) |
| Log of Importer Area | | | 0.009 | 0.010 |
| | | | (0.087) | (0.064) |
| Colonial Dummy | | | 0.032 | 0.031 |
| | | | (0.564) | (0.475) |
| RTA Dummy | | | 0.119* | 0.118*** |
| | | | (0.529) | (0.337) |
| Common Currency Dummy | | | 0.030 | 0.029 |
| | | | (0.565) | (0.607) |
| Fixed Effects Used | Commodity | Commodity | Commodity | Commodity |
| Observations | 479 | 479 | 479 | 479 |
| $R^2$ | 0.241 | 0.240 | 0.274 | 0.274 |
| First-stage F-stat | | 400.859 | | 403.439 |
| P-value of test $\ln intdist \neq \ln dist$ | 0.272 | 0.292 | 0.100 | 0.031 |

* p < 0.1, ** p < 0.05, *** p < 0.01. Robust standard errors in parentheses.

Table 5: Internal Distance Gravity Regressions - Various Years

| | (1) 1992 | (2) 1997 | (3) 2002 | (4) 2007 |
|---|---|---|---|---|
| Log of External Distance | -0.424* | -0.809*** | -0.832*** | -1.035*** |
| | (0.252) | (0.224) | (0.221) | (0.211) |
| Log of Internal Distance | -1.609** | -1.006* | -1.061** | -1.554*** |
| | (0.653) | (0.603) | (0.540) | (0.505) |
| Log of Importer GDP | 0.437*** | 0.579*** | 0.538*** | 0.570*** |
| | (0.065) | (0.063) | (0.060) | (0.057) |
| Fixed Effects Used | Commodity | Commodity | Commodity | Commodity |
| Type of Intrsument | Yes, 2SLS | Yes, 2SLS | Yes, 2SLS | Yes, 2SLS |
| Observations | 326 | 369 | 354 | 471 |
| $R^2$ | 0.144 | 0.199 | 0.220 | 0.248 |
| First-stage F-stat | 358.056 | 368.457 | 393.704 | 418.454 |
| P-value of test $\ln intdist \neq \ln dist$ | 0.101 | 0.761 | 0.704 | 0.360 |

* p < 0.1, ** p < 0.05, *** p < 0.01. Robust standard errors in parentheses.