6-29-2016

Three Essays on International and Intranational Trade and Economic Growth

Rooholah Hadadi
Florida International University, rhadadi@gmail.com

DOI: 10.25148/etd.FIDC000738

FLORIDA INTERNATIONAL UNIVERSITY

Miami, Florida

THREE ESSAYS ON INTERNATIONAL AND INTRANATIONAL TRADE AND ECONOMIC GROWTH

A dissertation submitted in partial fulfillment of the

requirements for the degree of

DOCTOR OF PHILOSOPHY

in

ECONOMICS

by

Rooholah Hadadi

2016

To: Dean John F. Stack, Jr.
Steven J. Green School of International and Public Affairs

This dissertation, written by Rooholah Hadadi, and entitled Three Essays on International and Intranational Trade and Economic Growth, having been approved in respect to style and intellectual content, is referred to you for judgment.

We have read this dissertation and recommend that it be approved.

Cem Karayalcin

Hassan Zahedi

Tobias Pfutze

Hakan Yilmazkuday, Major Professor

Date of Defense: June 29, 2016

The dissertation of Rooholah Hadadi is approved.

Dean John F. Stack, Jr.
Steven J. Green School of International and Public Affairs

Andrés G. Gil
Vice President for Research and Economic Development and Dean of the University Graduate School

Florida International University, 2016

DEDICATION

I dedicate this dissertation to my wife, Atid, who has been a constant source of support and encouragement. This work is also dedicated to my parents, who have always loved me unconditionally.

ACKNOWLEDGMENTS

I would like to express appreciation to my adviser, Dr. Hakan Yilmazkuday, for contributing his time and effort in guiding me through all the steps of this research project. He has been an amazing teacher and mentor to me and I am extremely grateful for his feedback, support and guidance. I am also grateful to the following committee members for their contribution and support: Dr. Cem Karayalcin, Dr. Hassan Zahedi, and Dr. Tobios Pfutze. Thank you all for your feedback and help. I am grateful to the economics department staff for providing me direction through the process, paperwork, and all the deadlines. Thank you all for this help.

ABSTRACT OF THE DISSERTATION

THREE ESSAYS ON INTERNATIONAL AND INTRANATIONAL TRADE AND ECONOMIC GROWTH

by

Rooholah Hadadi

Florida International University, 2016

Miami, Florida

Professor Hakan Yilmazkuday, Major Professor

This dissertation introduced a method to construct a new measure for trade flows within a region using nighttime lights. After analyzing the relation between lights data and other proxies of economic human activity, I employed light data and econometric techniques to estimate the bilateral trade between any two regions around the world. Using these estimations, I estimated the overall internal trade volume for all countries. Moreover, I estimated the effect of internal trade within a state of the United States on the state's income. The first essay proposed nighttime lights as an alternative proxy for economic activity to be used in gravity regressions. Due to the well-known problems in the measurement of gross domestic or regional products, gravity regressions based on both international and intranational trade data suffer from potential biases. At both international and intranational levels, log nighttime lights positively and significantly enter the gravity regressions (with a coefficient of roughly one) that explain at least about half of the variance in exports. The results were shown to be robust to the inclusion of several control variables and the consideration of predicted trade flows.

Trade within a nation and internal distance are variables known to play key roles in explaining home bias and the distance puzzle in international trade literature, but data on these measures are limited to only a few countries. To address this

problem, in the second essay, I constructed micro-founded measures of internal trade and internal distances from satellite data on nighttime lights. By estimating the gravity equation coefficients using the simulated method of distance estimation, I constructed the bilateral trade flow at subnational scale and aggregated it to overall internal trade. I found my internal trade measure is highly correlated with its benchmark, the difference between GDP and total exports; however, I showed it has more information and is a more precise measure for developed countries, which have a large amount of non-tradable services included in their national income account data. The internal distance measure is generated as the lights-weighted average distances between the states within a country. While my internal distance measure is largely correlated with its alternative, which is constructed based on city-level population data, it does not suffer from the uncertainty surrounding population data.

Correlation between trade and income cannot identify the effect of trade because of the endogeneity problem. The third essay examined this relationship at subnational level and by focusing on instrumenting trade via time varying geographic factors. Proximity and economic size are determinants of trade that are uncorrelated with other income determinants. This experiment not only confirmed the effect of interregional trade, but also provided evidence that intraregional trade has a large and statistically significant impact on income. I found, however, that the effect of both trade measures is statistically similar; a one percentage point increase in the interregional and intraregional trade ratio increases income per person by 2 to 4 percent.

TABLE OF CONTENTS

CHAPTER PAGE

1. LIGHTS ON TRADE: ESTIMATION OF TRADE FLOWS USING LUMINOSITY . . . 1
1.1 Introduction . . . 1
1.2 Data Sources . . . 6
1.2.1 Nighttime Lights Data . . . 6
1.2.2 Trade Data . . . 9
1.3 Gravity and Nighttime Lights . . . 10
1.3.1 World Trade Estimation Results . . . 12
1.3.2 Subnational Trade Estimation . . . 14
1.4 Conclusion Remarks . . . 16
1.5 References . . . 27

2. MEASURING INTERNAL TRADE FLOWS AND INTERNAL DISTANCES BY LIGHTS . . . 30
2.1 Introduction . . . 30
2.2 Model . . . 35
2.2.1 Trade Model . . . 35
2.2.2 Simulated Method of Distance . . . 37
2.3 Data Sources . . . 38
2.4 Estimation and Results . . . 39
2.4.1 Internal Trade . . . 40
2.4.2 Internal Distance . . . 44
2.5 Conclusion Remarks . . . 46
2.6 References . . . 54

3. GROWTH THROUGH INTRAREGIONAL VERSUS INTERREGIONAL TRADE . . . 57
3.1 Introduction . . . 57
3.2 Constructing the Instrument . . . 63
3.2.1 The Gravity Model . . . 66
3.3 Data and Results . . . 68
3.3.1 Trade Regression Results . . . 68
3.3.2 Aggregate Trade . . . 69
3.3.3 Quality of Instrument . . . 70
3.3.4 Trade Effect on Income . . . 71
3.4 Robustness . . . 73
3.5 Conclusion Remarks . . . 77
3.6 References . . . 85

APPENDIX . . . 88

VITA 94

LIST OF TABLES

TABLE PAGE

1.1 Baseline Results for the World: 1992-2006 22

1.2 Sensitivity Analysis for Trade Flow Estimation at National Level 23

1.3 Baseline Results for the United States: 1993-2007 24

1.4 Sensitivity Analysis for Trade Flow Estimation at State Level 25

1.5 Summary Statistics 26

2.1 Simulated Method of Distance Estimations 52

2.2 Summary Statistics of Internal Trade Measures 52

2.3 Descriptive Statistics of Internal Distance 53

3.1 Gravity Model Estimation 81

3.2 Actual and Constructed Overall Trade Relation 82

3.3 Trade and Income 83

3.4 Sensitivity Analysis 84

A1 Bootstrap Results for GDP Benchmark 88

A2 Constructed Internal Trade and Distance: 2007 89

LIST OF FIGURES

FIGURE PAGE

1.1 Nighttime Lights: 2006 18

1.2 Nighttime Lights versus GDP of Countries 19

1.3 Nighttime Lights versus GSP of U.S. States 20

1.4 Log of Total Exports predicted by GDP versus Nighttime Lights 21

2.1 Internal trade measured by the SMD estimator and the benchmark . . . 49

2.2 Internal Trade vs. Benchmark: 2007 49

2.3 Fitted Line for Constructed Internal Trade and the Alternative 50

2.4 Constructed Internal Trade vs. Alternative 50

2.5 Constructed Internal Distance vs. Benchmark 51

2.6 Comparison of Alternative Internal Trade Distance 51

3.1 Overall Internal Trade Versus Two Size Measures 79

3.2 Overall Actual Trade Versus Overall Constructed Trade 80

3.3 Actual Versus Constructed Overall Trade, Controlled for Size Measures . 80

CHAPTER 1

LIGHTS ON TRADE: ESTIMATION OF TRADE FLOWS USING LUMINOSITY

1.1 Introduction

The gravity model of trade has attracted much interest, generating analytical and empirical studies in which gravity equations emerge from most mainstream trade models (Evenett and Keller, 1998; Feenstra, 2003; Arkolakis et al., 2012); these studies show a characteristic of trade patterns in which exports rise proportionally with the economic size of trade partners. Despite this tendency, what is the best measure for economic size to be used in these gravity studies? Because of its availability for virtually all countries over several years, Gross Domestic Product (GDP) has been the most commonly used measure in the literature. However, the measure of GDP has several shortcomings when it is used in gravity studies.

The first problem is the quality differences (in terms of measurement errors) in GDP data across countries. In particular, developing economies have larger non-formal sectors than developed countries, less market integration across regions, and weaker national statistical infrastructure. For example, the Penn World Table (PWT), which is a standard data set of cross-country GDP, gives countries a data quality ranking of A, B, C, and D. Most developing countries, including all sub-Saharan African countries, are graded C or D. Moreover, the PWT itself has several revisions and GDP estimates vary considerably across versions. Johnson *et al.* (2009) show that there is significant variability, both in the level and growth of GDP and Purchasing Power Parity (PPP) across alternative versions of the PWT. They also find that GDP data in levels or growth rates are not PPP-based for non-benchmark years. Given the drawbacks in the PWT data, they conclude that it is only safe to use GDP or growth data for long-term studies (of at least 10 years), and not for analyses that require annual data.

The second problem is related to the definition of GDP. Especially when the expenditure method is used to measure GDP, exports and imports are included in the calculation. Therefore, there is an accounting relationship between trade and GDP that may potentially inflate the explanatory power of gravity regressions, especially for countries with trade imbalances.

Beside the concerns on the measurement errors, the third problem is that GDP data are not available at alternative geographical aggregation levels; most countries only publish national GDP measures. The majority of trade, however, takes place within a country rather than across countries; e.g., Yilmazkuday (2012) shows that the volume of intranational trade is more than six times the volume of international trade, on average, between 1993 and 2007, according to U.S. trade data. Such intranational trade, which is important for understanding the effects of geographical factors on economic growth, simply cannot be analyzed in most countries due to the lack of GDP (or gross regional product, GRP) data at sub-national levels, where such measures are further subject to measurement errors if they are available.

In order to avoid such measurement problems of GDP/GRP, this paper proposes an alternative proxy for the economy size used in gravity regressions: the nighttime lights illuminated at disaggregated geographical locations.[1] My motivation comes from many studies that have shown high correlations between economic activity and nighttime lights (e.g., see Croft, 1978, Elvidge et al., 1997, Sutton and Costanza, 2002, Ebener et al., 2005, Doll et al., 2006, Sutton et al., 2007, and Ghosh et al, 2009, among many others). Accordingly, the main objective is to test whether

[1]The nighttime lights are measured for each 30-arc-second grids that correspond to about a space of about 25 square blocks or 0.25 square miles in the U.S.

nighttime lights are appropriate proxies for economic activity and can be further used in explaining both international and intranational trade patterns. Firstly, however, I have to discuss how nighttime lights can overcome the problems regarding the measurement of GDP/GRP.

It is intuitive that almost all investment and/or consumption activities in the evening require lights. The findings on the strong correlation between human economic activity and lights have formed a large and growing literature on using lights to estimate several socio-economic parameters. Although nighttime lights data may also be subject to measurement errors, it benefits from several unique advantages over other proxies for economic activity.[2] For instance, while measures such as GDP, means of survey, and electrical consumption are collected from multiple international data sets with unquantifiable qualities, lights data are collected from images captured for all the world's regions independently. Thus, they are homogenous/standard across countries and consistent over time; this overcomes the problem of quality differences (in terms of measurement errors) in GDP data between countries. Since neither exports nor imports are used in the calculation of nighttime lights data, the second problem (regarding the accounting relationship between trade and GDP) is also overcome.

[2]As an alternative to nighttime lights, the Geographically based Economic data (G-Econ) project at Yale University aims to offer a solution for the lack of GDP data at subnational levels by constructing economic data for every one-by-one decimal degree grid cell around the world. Nonetheless, this project faces the same problem of data quality in most of low-income countries. Similarly, to address the measurement problems in cross-country GDP values, Young (2012) creates a new proxy of growth using the changes in consumption of goods reported by Demographic and Health Surveys; it shows much higher economic growth rates for developing countries than the rates indicated by national accounts. Another proxy produced by the World Bank is the Household Consumption Survey of consumers from 123 countries; yet, the survey suffers from problems such as nonresponse bias, which is common in survey studies.

Additionally, nighttime lights are available at any temporal and geographical scales for which other proxies are of poor quality or non-existent. Illumination intensity is measured at 30-arc-second grids around the world that correspond to an area of about 0.25 square miles in the U.S. and can be aggregated to any spatial level. This helps to resolve the third problem (of using GDP) regarding the national account data limitations of studying the policy analysis at sub-national level (and the corresponding sub-national measurement errors).

In this paper, using nighttime lights as an alternative proxy for economic activity, I estimate gravity regressions at both international and intranational levels. I follow the gravity framework developed by Anderson and van Wincoop (2003), where the empirical methodology closely follows the existing literature for the determination of control variables. While international trade regressions are achieved by using export data of 192 countries/territories, as in Rose and Spiegel (2009), intranational trade regressions are achieved by using export data of 48 states of the U.S. (excluding Hawaii, the District of Columbia and Alaska, as standard in the corresponding literature) as in Wolf (2000), Hillberry and Hummels (2002) or Yilmazkuday (2012).

Although the exact estimates change across alternative regression specifications, at both international and intranational level, the log nighttime lights positively and significantly enter gravity estimations with a coefficient of about one. The results are shown to be robust to the inclusion of several control variables. The corresponding regressions explain at least half the variance in trade flows across countries/states. Moreover, nighttime lights are shown to predict changes in total export well with unitary elasticities, which are comparable to other studies that have applied the same data with other measures of economic size.

This paper is at the intersection of the two strands of literature, namely the established literature on gravity estimations and the recent literature on estimating economic activity using nighttime lights data. While gravity equations were introduced more than half a century ago (Tinbergen, 1962), the widespread use of structural gravity models is a fairly recent phenomenon. The last decade has seen an explosion of alternative micro-theoretical foundations for gravity models (e.g., Eaton and Kortum, 2002; Anderson and Wincoop, 2003; Eaton, Kortum, and Kramarz, 2008; Chaney, 2008; Arkolakis et al., 2012). These studies show that nearly all trade models require gravity in order to work, with a measure of economic activity the key ingredient in estimated regressions. In this paper, I directly follow this literature, using an alternative measure of economic activity. In order to show how my results compare with the existing literature, I also include several other regressions using GDP as the standard measure of economic activity.

Regarding the employment of nighttime lights as a proxy for economic activity, Huang *et al.* (2014) show a rapid growth in journal-published papers in a meta analysis. Many studies have confirmed that nighttime lights reflect human economic activity (e.g., Lazar 2010; Sutton and Costanza, 2002; Elvidge et al., 1997); although the potential use of nighttime lights was suggested more than 40 years ago by Croft (1973) as a means of estimating population density and energy consumption.[3] In this context, my paper is closest in spirit to the seminal paper by Henderson, Storeygard, and Weil (2012; HSW, henceforth). Using a statistical framework, HSW suggest the use of nighttime lights data to measure real income growth. Following HSW, many other efforts have been made to use nighttime lights data to improve the measurement of GDP in poor countries; e.g., Chen and Nordhaus (2011) suggest

[3]Elvidge et al. (1997) utilize lights to construct a measure for poverty and inequality. Given a log-log relation between urban areas and population, Sutton et al. (2007) estimate the urban population and GDP at sub-national level for the nations of China, India, Turkey, and the United States.

using both measured GDP and lights data to estimate the true national income.[4] In this paper, I build on previous research by using nighttime lights data as an alternative measure of economic activity in the context of explaining trade patterns.

The rest of the paper is organized as follows. The next section introduces the data on nighttime lights, international trade and intranational trade. Section 3 estimates the standard gravity equations in the literature using nighttime lights data, and compares the results with the standard approach of using GDP. Section 4 concludes.

1.2 Data Sources

In this section, I introduce the data I use in gravity estimations. My analysis is based on two data sources, namely nighttime lights and bilateral trade flows. I present a brief introduction of nighttime lights data and discuss how lights reflect economic activity. For the trade data, my analysis is based on two sources of trade flow information; one international and one intranational, and for which I provide descriptive information.

1.2.1 Nighttime Lights Data

The primary source of nighttime lights data is the images captured by satellites from the United States Air Force Defense Meteorological Satellite Program (DMSP). Since the mid-1960s, satellites have orbited the Earth every evening between 20:30 and 22:00, recording images that contain information on illumination intensity. The

[4] Instead of using lights data directly, Pinkovskiy and Sala-i-Martin (2014) utilize lights as an auxiliary tool to find the optimum weights for computing weighted average of GDP and mean surveys. They find that poverty is significantly lower and falls faster than estimations by other studies using GDP or survey-based poverty measures.

initial mission of this program was to study the extent of moonlit cloud worldwide for weather-forecasting purposes. However, as the National Oceanic and Atmospheric Administration (NOAA) processes the images and makes them available to the public, they have been used for several socio-economic studies. NOAA processes the images by screening the observations for auroral activity, summer months when sunset is late, the bright half of the lunar cycle, and for the locations experiencing ephemeral events, such as fire and background noise. After omitting problematic observations, data from all valid orbits of a satellite over the course of one year are averaged and distributed to the public.

Although the data cover the years between 1992 and 2013, I only use nighttime lights up to 2007, the last year for which I have data on bilateral trade at state and country levels. Each image in the data set contains information about every location from 65-degree north latitude to 75-degree south latitude and divides the covered area into pixels with a resolution of 30 arc-second. Each pixel is equivalent to 0.33 square miles (0.86 square kilometers) at the equator; the area of the pixel changes as one moves further from the equator. The data report the intensity of light for each pixel as a six-bit digital number (DN) between 0 and 63.

There are some problems with using the digital numbers as a measure for true light illumination. First, the pixels with a DN value of 63 may be the result of sensor saturation (top-coded). There are alternative versions of the images, which are calibrated for saturation and have the advantage of not being top coded. Using this data, which is only available for 2006, Chen and Nordhaus (2011) find small quantitative differences compared to the other version. Another problem is that the setting of sensors may vary across satellites, and with their age. Therefore, using the raw digital numbers for time series analysis could be problematic. In the next section, I control for this issue by adding the year and the satellite fix effects to my investigation. Other problems that might affect the data include lights overflowing from the source pixel to adjacent ones, and lights blooming in some specific terrain types, such as those covered by snow (Doll, 2008). Because digital numbers are, however, summed over all pixels at a large geographical level, it is unlikely these sources of errors are substantial (Pinkovskiy and Sala-i-Martin, 2014).

Most consumption and activity, especially in the evening, requires lights. This is the reason some studies focus on using light luminosity as a proxy for human economic activity. As discussed before, lights show a strong correlation with both the level and growth rates of measured income. Light consumption, however, differs in regards to the economy production composition. Agricultural economies consume less electricity for production overnight than economies that rely on industrial production. Light usage culture is another parameter that is different between countries around the world. Because there is no consistent information about consumption culture, production composition and the like, I cannot incorporate these factors into my analysis. However, I use some statistical formulations such as using time/location fixed effects or *growth* measure to account for these aspects of economies.

The world lights map on the first panel of Figure 1 suggests that developed countries in North America, eastern Asia and Europe have more lights than the rest of world. To better illustrate the link between GDP/GSP and nighttime lights, Figures 2 and 3 show the relationship between nighttime lights and GDP/GSP for 2007 at world and state levels. Countries such as the U.S., China, Russia and India, and states like California, Texas and Illinois, are at the top right-hand corner of the plot, showing the highest lights and GDP growth for 2007. The lowest GDP growth rates of Maldivies and Comoros are associated with the least lights growth. The plots depict some degree of nonlinearity in nonparametric fit graphs, especially for the case of U.S. states.[5] Higher-order terms are insignificant, however, and thus, I assume a linear relationship, meaning that lights changes have the same interpretation across countries/states with different GDP growth rates (see HSW; Chen and Nordhaus, 2011).

1.2.2 Trade Data

In order to enable a healthy comparison with the existing literature, I employ the international trade data of Rose and Spiegel (2011), in which the country export volumes come from the International Financial Statistic's Division of Trade CD-ROM, are measured in U.S. dollars, and deflated by U.S. CPI for All Urban Consumers (CPI-U); all items, 1982-84=100. The data set covers the years from 1950 to 2006, but as the lights data begin in 1992, and in order to be consistent with the intranational trade data set (to be introduced below), I use the data for the years 1993 to 2006 and update it with 2007 numbers. Real GDP data, which are employed for comparison purposes, are taken from the PWT mark 6.2.

The intranational trade data within the U.S. are from the Commodity Flow Survey (CFS) collected by the Census Bureau on behalf of the U.S. Department of Transportation; this is the same data source used in studies such as Wolf (2000), Hillberry and Hummels (2002) and Yilmazkuday (2012). The data cover the years of 1993, 1997, 2002, and 2007 and are available to the public online. The CFS provides data on interstate trade as well as intrastate shipments categorized as manufacturing, mining, wholesale or retail establishments. Although it covers most of the manufacturing sector, which is the largest goods producing sector, it excludes agriculture and some mining. For the sake of consistency with previous works, and

[5] Running the Regression Equation Specification Error Test (RESET) rejects the linearity.

because of data availability, the states of Hawaii, District of Columbia and Alaska are excluded.

Traveled distance is the best measure for the economic distance between any two trade partners, and CFS reports intranational trade within the U.S. However, as the corresponding data are not available at cross-country level, I use the great circle distance for both state and country level analyses to ensure consistency. To calculate the great circle distance across countries/states, I find the center of each country/state using the average longitude and latitude values weighted by the nighttime lights.[6] Gross State Product (GSP) data for the U.S. states, which are used for comparison purposes, are from the U.S. Bureau of Economic Analysis. Table 1.4 presents summary statistics of the key variables for the United States and the world samples.

1.3 Gravity and Nighttime Lights

This section proceeds by defining the structural gravity equation used for the analysis and providing the empirical results. Following the existing literature, I consider the gravity equation derived by Anderson and van Wincoop (2003), in which bilateral trade X_{ijt} between countries/states i and j at time t is given by:

$$X_{ijt} = \frac{M_{it} M_{jt}}{G_{wt}} \left(\frac{\tau_{ijt}}{p_{it} p_{jt}} \right)^{1-\sigma} \tag{1.1}$$

where M_{it} is an exporter-specific factor that represents the exporter's capacity to supply, M_{jt} is an importer-specific factor that represents the importer's demand, G_{wt} represents the total size of the countries/states considered, τ_{ijt} represents trade

[6]For each country/state, I calculate the longitude and latitude of the geological center points using Arcmap 10.2 application and then compute the distances between the center of trade partners.

costs, σ is the elasticity of substitution, and finally p_i and p_j indicate the ease of market access for the trade partners - the so-called *multilateral resistance* terms (MRT).[7]

The existing literature uses data on GDP or GNI to capture the effects of M_{it} and M_{jt}. Due to the measurement problems discussed previously, I instead use nighttime lights data to represent these variables. In particular, given its multiplicative nature, I take the log of both sides of the gravity equation for estimation purposes. Accordingly, I estimate the following expression:

$$\ln(X_{ijt}) = \beta_0 + \beta_1 \ln(M_{it}) + \beta_2 \ln(M_{jt}) + \beta_3 \ln(d_{ij}) + \beta_4 \ln(p_i) + \beta_5 \ln(p_j) + \mathbf{B}\ln(\mathbf{X}_{ij}) + \epsilon_{ijt} \quad (1.2)$$

where the sum of nighttime lights are used for each trade partner at time t as proxies for M_{it} and M_{jt}, d_{ij} represents distance, MRT of p_i and p_j are captured by either remoteness variables (as in Head, 2003)[8] or country/state fixed effects (as in Feenstra, 2003; Rose and Spiegel, 2009), and X_{ij} is a set of control variables as standard in the gravity literature, namely currency union, common language, regional trade agreements (RTA), common border, number of islands involved in trade, log of product area, common colonizer, currently colony, ever colony, and common/same country. I estimate this equation using both international and intranational trade data sets. To enable a comparison with the existing literature, I also consider GDP/GSP data as alternative measures of M_{it} and M_{jt}. In order to account for time-specific effects, such as satellite sensor setting variations over the years, I also include year-specific fixed effects in the gravity equation. I achieve the estimation by OLS, with a robust covariance estimator to address the plausible heteroskedasticity problem.

[7] In particular, p_i and p_j are lower when the trade partners are remote from the world markets.

[8] Remoteness is measured as the GDP or light weighted average distance between exporter (importer) and all other countries except importer (exporter); it is calculated as $R_{it} = \sum_j \frac{D_{ij}}{M_{jt}/G_{wt}}$ where the notation is the same as above.

1.3.1 World Trade Estimation Results

This section reports the gravity estimation results obtained using the international trade data introduced above. The baseline results are given in Table 1.1, where the first and third columns replicate the literature by using GDP as the measure of economic activity, while the second and fourth columns use nighttime lights as an alternative.

As expected, the estimated coefficients of GDP and nighttime lights are both positive and significant. The magnitudes of the estimates are also comparable, although the estimates are lower in the regressions using nighttime lights. The estimates for coefficients in front of nighttime lights take values of 1.097 and 0.870 when all control variables are used and MRT are captured by remoteness variables. Therefore, the elasticity of exports with respect to nighttime lights is close to unity regarding the economic size of both exporters and importers.

The control variables mostly take similar estimated signs and coefficients when the results of GDP and nighttime lights data are compared. One exception is for the effects of RTA, which are negative and insignificant in the case of GDP, but negative and significant when lights are used. Another exception is for the number of islands involved in trade, where the estimated coefficient is positive and significant when GDP data are used, but negative and significant when lights data are used. Similarly, the effects of being the same country in a particular year are estimated as negative and insignificant when GDP data are used, while such effects are estimated as positive and significant when lights data are used. Hence, changing the measure of economic size used in gravity regressions affects the coefficient estimates of control variables, which may have significant policy implications. The high explanatory power in all regressions supports the results, although the R-squared values when

GDP data are used may be inflated, especially for countries with trade imbalances (as discussed above).

To demonstrate how using nighttime lights (as opposed to GDP) can predict total trade flows, in addition to the results presented in Table 1.1, I also compute total exports by estimating the bilateral trade, using column 3 and 4, and then total all exports for each exporting country. The Figure 4 graphs estimated exports for several developed, emerging market, and developing countries. As is evident, nighttime lights predict changes in total exports in a very similar way to GDP. Therefore, in addition to the benefits of using nighttime lights (compared to using GDP, which is subject to the many previously discussed measurement problems), they also provide a means of accurately predicting trade flows.

For robustness, I also consider alternative regressions in Table 1.2, where each column shows the results of two separate estimations, one with GDP and the other with nighttime lights, as indicators of economic activity, where the control variables as in Columns 3 and 4 in Table 1.1 are still used (when possible). The only exception in Table 1.2 is Column 1, where both measures of economic activity are incorporated together to give the reader a better idea of their interaction. In Column 2 of Table 1.2, both country fixed effects and time fixed effects are considered, which is an alternative way of controlling for MRT as advocated by Feenstra (2003) and Rose and Spiegel (2009). Column 3 of Table 1.2 considers dyadic-specific dummies to capture time-invariant characteristics, which are common to a pair of countries. Alternatively, Column 4 of Table 1.2 adds country-pair time trend effects to the base model. As is evident, in Columns 1-4 in Table 1.2, the coefficients of both GDP and nighttime lights remain significant and positive, independent of the regression specification, which is an indicator of robustness for the benchmark results given in Table 1.1.

Within this picture, I also consider another problem in the literature, which is the missing (or zero) trade observations between countries. The standard gravity estimation takes the logarithm of trade and achieves the estimation in the log-linear form; however, since the log of zero is not identified, observations with a trade volume of zero are removed from any estimation. Zero trade itself, however, may reflect some valuable economic information (rather than some systematic rounding errors); therefore, dropping out the zeroes may yield inconsistent and biased results. One solution is to use the Pseudo Poisson maximum likelihood estimator (see, for example, Westerlund and Wilhelmsson, 2011). I follow the literature by using this estimation methodology, results for which are given in Column 5 of Table 1.2. Once again, the coefficients for both economic indicators of GDP and nighttime lights are positive and significant, with magnitudes around unity. Therefore, I can safely claim that the usage of nighttime lights is in fact an alternative to the usage of GDP as an economic indicator.

1.3.2 Subnational Trade Estimation

After showing that nighttime lights are good indicators of economic activity in gravity studies based on international trade data, in this section, I turn to the sample of trade flows between the states of the U.S. to examine the effectiveness of using nighttime lights in the estimation of intranational (i.e., interstate) trade. The results are provided in Table 1.3, with a balanced panel over 4 years used. Because of the change in GDP calculation method in 1997 and other measurement issues discussed above, time fixed effects are used in all specifications. Control variables differ from the international analysis above, as all states share a common currency, language, and history.

Since observations may be heterogeneous in a variety of ways, robust standard errors are used in all estimations. In order to compare the results with other studies which used CFS, Column 1 of Table 1.3 runs the regression only for the year of 1993. The coefficients of nighttime lights enter significantly and positively with an estimate close to one (i.e., 95 percent confidence interval includes one for both trade partners.) The results are consistent with those of previous works using GDP. For example, Wolf (2000) finds the trade elasticities of 1.02 and 0.98 with regards to GDP for exporting and importing states. While the R-squared of 0.53 shows a high explanatory power, it is lower than the specifications using GDP as an alternative economic activity indicator.

I use a panel data of four years in other columns of Table 1.3 in order to compare the coefficients of GDP and nighttime lights. The estimated coefficients of nighttime lights are positive and significant with an estimate of about one in all specifications, which is consistent with the estimates using GDP as the measure of economic activity.

For further robustness, Columns 4 and 5 use additional control variables as advocated by Hillberry and Hummels (2003) and Wolf (2000); in both columns, the coefficient estimates of GDP and nighttime lights are significant; almost the same and close to one. High explanatory powers in all specifications support the results.

Further sensitivity analysis is achieved in Table 1.4 (similar to what appears in Table 1.2 for the international trade data), where Column 1 incorporates both GDP and nighttime lights in the same regression. Table 1.4 shows that both coefficient estimates remain significant and positive, except when country-pair time trend effects are considered in Column 4. Therefore, nighttime lights can be used as a measure of economic activity for intranational trade regressions as well as international trade regressions. High explanatory powers in Table 1.4 further support the results.

1.4 Conclusion Remarks

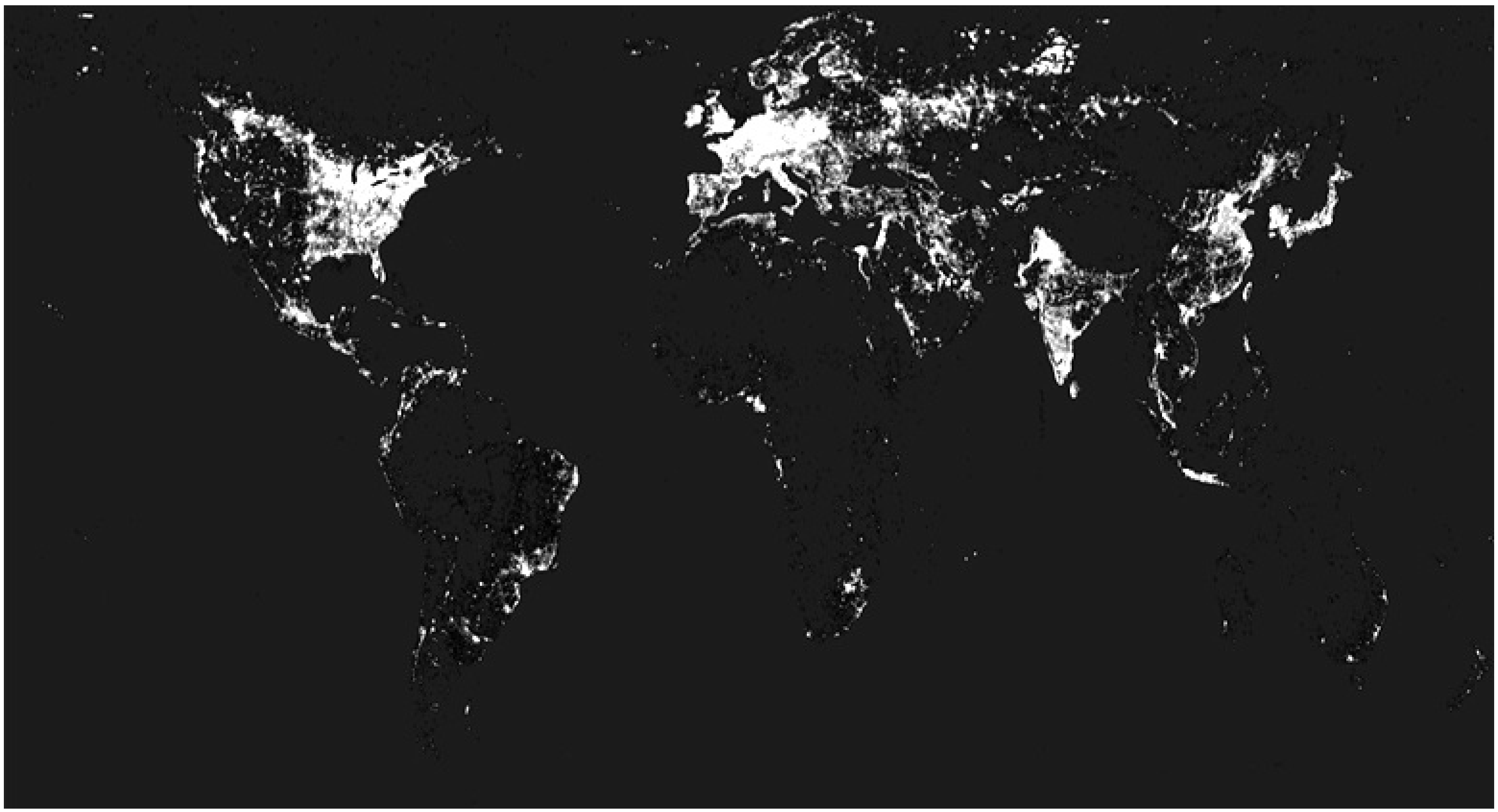
Gravity equations have proven to fit well with analytical frameworks and empirical estimations. Accordingly, many researchers have used gravity models in order to either explain trade patterns or analyze the impact of trade policies. Data-related problems, however, have prevented researchers applying this approach to the parts of the world where data are either missing/limited (e.g., intranational trade within a country) or low quality (e.g., data coming from developing/emerging countries).

This paper has attempted to use nighttime lights data as an alternative measure for economic size in gravity regressions. Since nighttime lights are required for almost all human activities in the evening, including consumption and investments, changes in light illuminations can be attributed to the change in total human activity due to income or population growth. Using statistical frameworks, several studies confirm that lights data contain valuable economic activity information at any spatial level (HSW, 2012).

I have found that nighttime lights have a high explanatory power in explaining trade patterns at both international and intranational level. Estimations using GDP and lights data are highly similar, explaining well over 50 percent of variances of exports for countries around the world; in particular, while the elasticity estimates of total exports with respect to nighttime lights are between 0.68 and 1.10, they are between 0.90 and 1.36 when GDP is considered. Therefore, nighttime lights can be used as alternative measures of economic activity in gravity studies.

In addition to being an alternative measure of economic activity, the main benefits of using nighttime lights are that they exist for poor or developing countries and are available at any geographical scale. Accordingly, one application of nighttime lights would be to simulate trade at intranational levels; e.g., one can predict the trade between any two regions by using the corresponding nighttime lights of such regions. These regions can, for example, be zip codes, counties, cities or states within the U.S. One can predict the (log) trade between such regions by using (log) nighttime lights and the coefficients estimated in this paper, which have been shown to be about one by using both international trade (around the world) and intranational trade (within the U.S.) data. I leave such an analysis for future research.

Figure 1.1: Nighttime Lights: 2006



Map source: NOAA's national Geophysical Data Center.

Figure 1.2: Nighttime Lights versus GDP of Countries

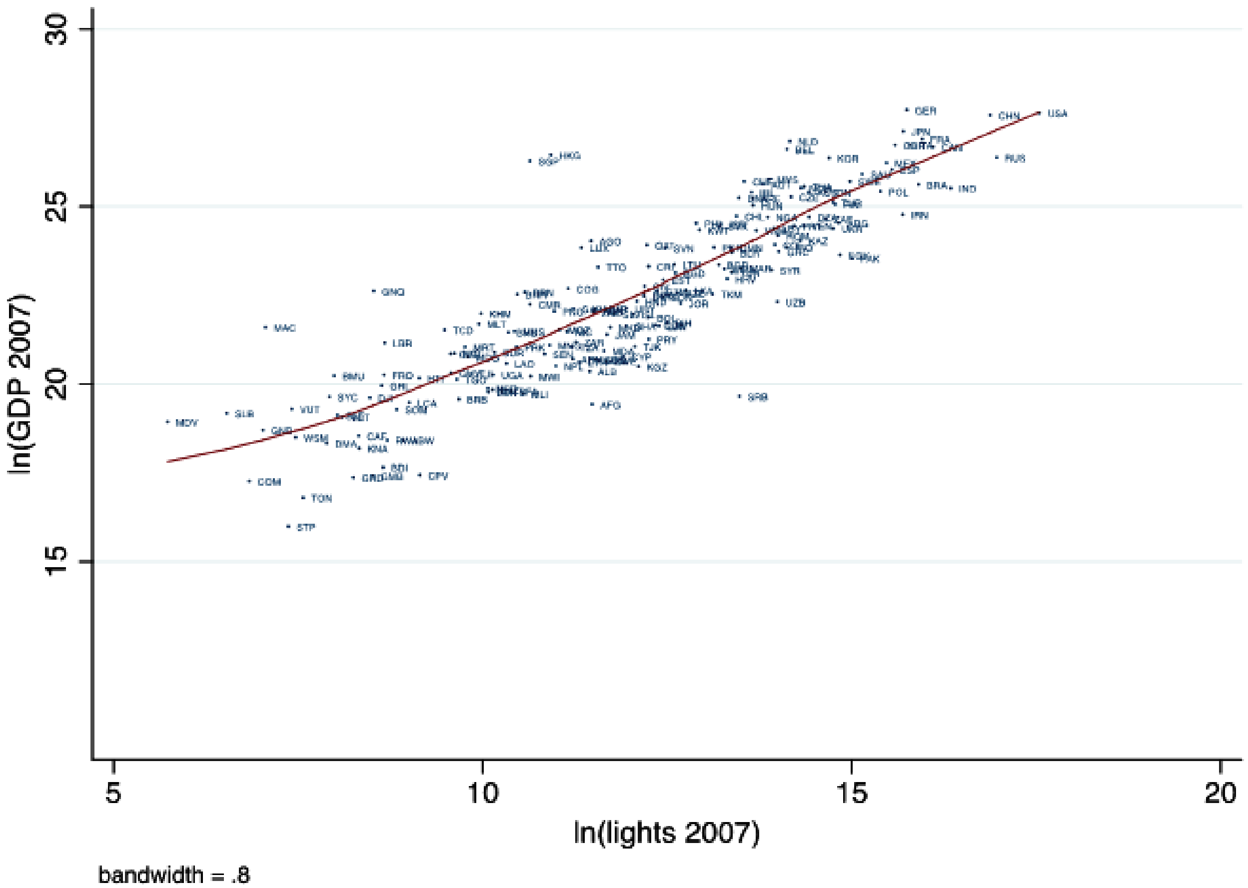


Figure 1.3: Nighttime Lights versus GSP of U.S. States

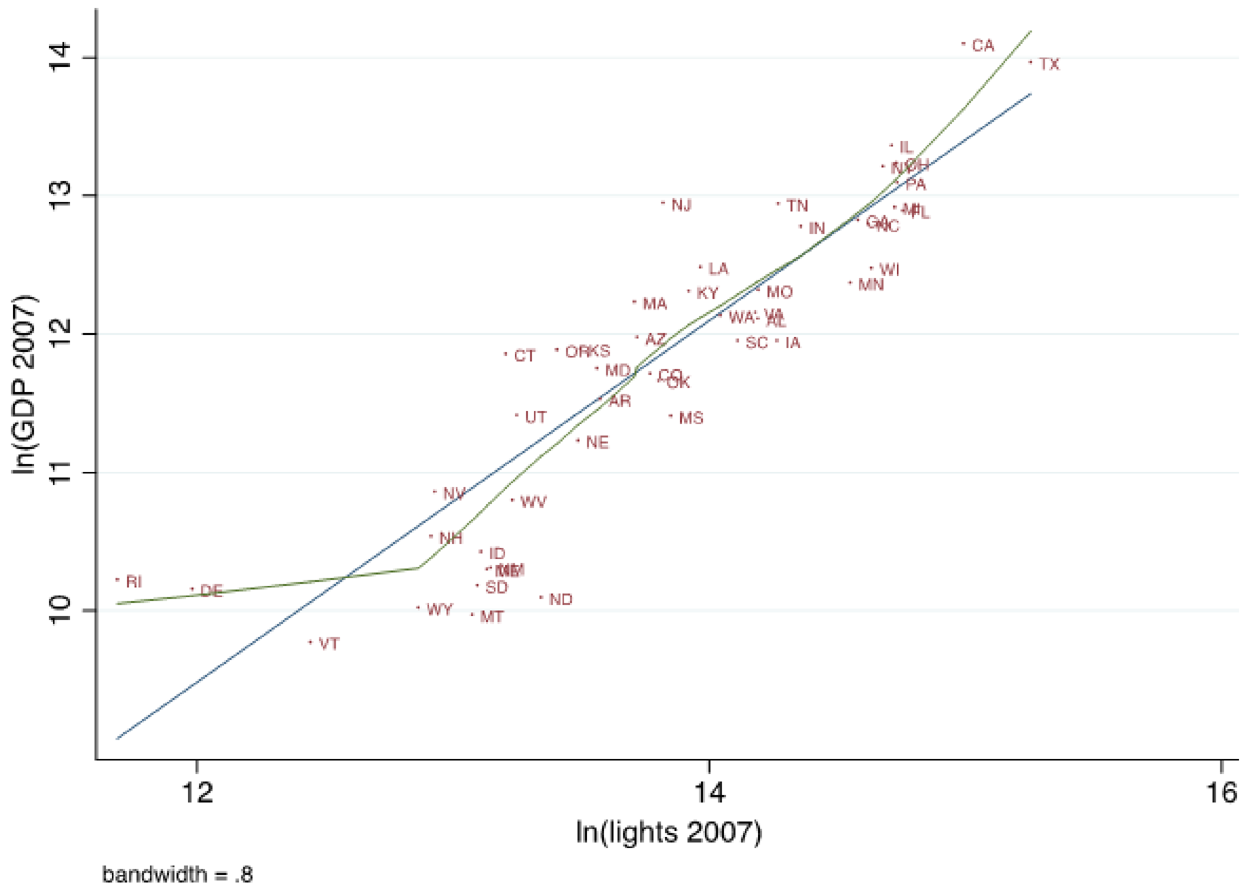


Figure 1.4: Log of Total Exports predicted by GDP versus Nighttime Lights, 1992-2006

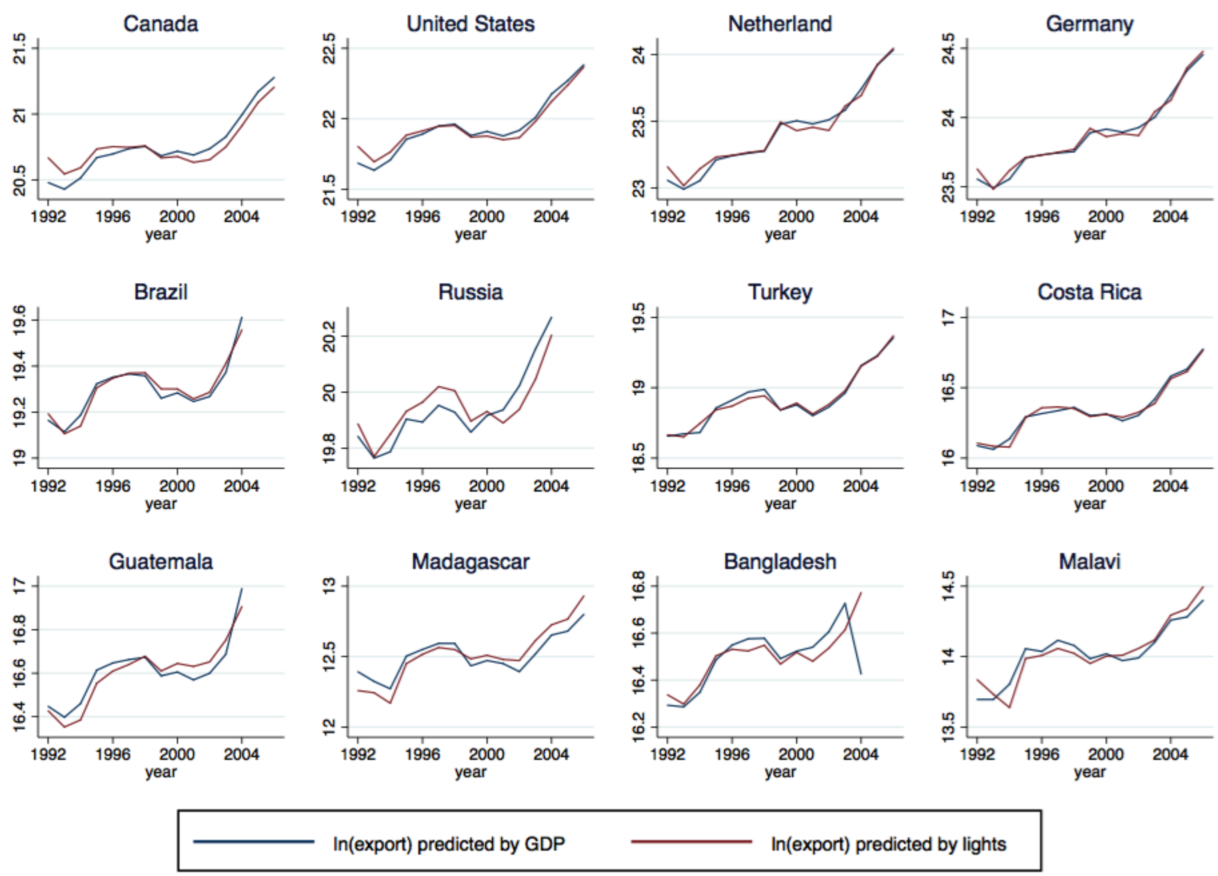


Table 1.1: Baseline Results for the World: 1992-2006

	Dependent Variable: ln(exports)			
	(1)	(2)	(3)	(4)
$\ln(\text{rgdp}_i)$	1.211***		1.360***	
	(0.0076)		(0.0084)	
$\ln(\text{rgdp}_j)$	0.903***		1.066***	
	(0.0075)		(0.0082)	
$\ln(\text{lights}_i)$		0.929***		1.097***
		(0.0078)		(0.0084)
$\ln(\text{lights}_j)$		0.678***		0.870***
		(0.0073)		(0.0084)
ln(distance)	-1.418***	-1.075***	-1.517***	-1.362***
	(0.0174)	(0.0202)	(0.0242)	(0.0260)
currency union			0.883***	1.390***
			(0.1120)	(0.1075)
common language			0.475***	0.468***
			(0.0417)	(0.0462)
RTA			-0.0499	-0.0975**
			(0.0311)	(0.0345)
common border			0.590***	0.931***
			(0.0926)	(0.0927)
# island			0.0738*	-0.231***
			(0.0356)	(0.0407)
log product area			-0.198***	-0.239***
			(0.0060)	(0.0072)
common colonizer			0.525***	0.532***
			(0.0584)	(0.0633)
currently colony			0.207	0.178
			(0.3288)	(0.8075)
ever colony			1.708***	2.260***
			(0.1047)	(0.1110)
common country			-0.715	1.636*
			(0.4949)	(0.7462)
$\ln(\text{remoteness}_i)$			1.230***	2.982***
			(0.0650)	(0.0719)
$\ln(\text{remoteness}_j)$			0.892***	2.083***
			(0.0643)	(0.0705)
constant	-16.97***	-2.101***	-34.98***	-41.02***
	(0.2636)	(0.2231)	(0.7159)	(0.7977)
R^2	0.575	0.456	0.621	0.552
Observations	215783	253519	215783	253519

Notes: All specifications include year fixed effects. Robust standard errors, clustered by dyads, are in parentheses.
*** Significant at the 1 percent level
** Significant at the 5 percent level
* Significant at the 10 percent level

Table 1.2: Sensitivity Analysis for Trade Flow Estimation at National Level: 1992-2006

	GDP & lights	GDP or lights Country FE	GDP or lights Dyads	GDP or lights Fluctuations	GDP or lights PPML
	(1)	(2)	(3)	(4)	(5)
ln(rgdp$_i$)	1.119***	0.492***	0.560***	0.564***	1.013***
	(0.0203)	(0.0482)	(0.0481)	(0.0479)	(0.0274)
ln(rgdp$_j$)	0.909***	0.373***	0.441***	0.447***	1.041***
	(0.0181)	(0.0388)	(0.0353)	(0.0350)	(0.0292)
ln(lights$_i$)	0.240***	0.203***	0.272***	0.266***	1.227***
	(0.0180)	(0.0324)	(0.0304)	(0.0300)	(0.0093)
ln(lights$_j$)	0.165***	0.237***	0.244***	0.244***	1.265***
	(0.0159)	(0.0279)	(0.0260)	(0.0259)	(0.0111)
R^2_{GDP}	0.62	0.72	0.88	0.88	
R^2_{lights}	0.62	0.72	0.88	0.88	
Observations	215783	215783	215783	214028	215783

Notes: t-statistics in parenthesis. First column uses both GDP and lights in one regression. The other columns show results of two different regressions, one based on GDP, the other based on lights as the measure of economic activity.
*** Significant at the 1 percent level
** Significant at the 5 percent level
* Significant at the 10 percent level

Table 1.3: Baseline Results for the United States: 1993-2007

	Dependent Variable: ln(exports)				
	(1)	(2)	(3)	(4)	(5)
ln(rgdp_i)		0.983***		0.986***	
		(0.0165)		(0.0141)	
ln(rgdp_j)		0.973***		0.988***	
		(0.0171)		(0.0149)	
ln(lights_i)	0.942***		1.025***		1.011***
	(0.0377)		(0.0312)		(0.0316)
ln(lights_j)	1.036***		1.088***		1.076***
	(0.0386)		(0.0321)		(0.0319)
ln(distance)	-1.409***	-1.181***	-1.350***	-0.903***	-1.262***
	(0.0434)	(0.0313)	(0.0388)	(0.0348)	(0.0563)
adjacency				0.788***	0.101
				(0.0819)	(0.1221)
intra				1.999***	1.006***
				(0.1784)	(0.1795)
ln(remoteness_i)				-0.103**	-0.233***
				(0.0376)	(0.0645)
ln(remoteness_j)				0.461***	0.312**
				(0.0699)	(0.1125)
constant	-11.96***	-8.032***	-13.94***	-15.65***	-15.41***
	(1.0199)	(0.4691)	(0.8957)	(1.3120)	(2.4919)
R^2	0.526	0.828	0.581	0.856	0.589
Observation	2137	8392	8392	8392	8392

Notes: All specifications include year fixed effects. Robust standard errors, clustered by dyads, are in parentheses.
*** Significant at the 1 percent level
** Significant at the 5 percent level
* Significant at the 10 percent level

Table 1.4: Sensitivity Analysis for Trade Flow Estimation at State Level: 1993-2006

	GSP & lights	GDP or lights State and Dyads FE	GDP or lights Fluctuations	GDP or lights PPML
$\ln(\text{rgdp}_i)$	0.672***	0.429***	0.597***	0.676***
	(0.030)	(0.071)	(0.098)	(0.018)
$\ln(\text{rgdp}_j)$	0.772***	0.437***	0.616***	0.716***
	(0.030)	(0.071)	(0.117)	(0.020)
$\ln(\text{lights}_i)$	0.494***	0.0585	-0.153*	1.007***
	(0.035)	(0.080)	(0.075)	(0.023)
$\ln(\text{lights}_j)$	0.324***	0.352***	0.186*	1.016***
	(0.036)	(0.080)	(0.075)	(0.023)
R^2_{GDP}	0.84	0.96	0.96	0.91
R^2_{lights}	0.84	0.94	0.96	0.92
Observations	8392	8392	8353	9037

Notes: Standard errors in parenthesis. First column uses both GDP and lights in one regression. The other columns, each shows two different regression results reported in one column for the sake of saving some space.
*** Significant at the 1 percent level
** Significant at the 5 percent level
* Significant at the 10 percent level

Table 1.5: Summary Statistics

Variable	Mean	SD	Min	Max	Count	Sample
ln(export)	9.5009	3.8067	-22.7028	21.1751	253519	world
ln(GDP per capita)	8.7704	1.1714	5.1391	11.1484	271036	world
ln(real GDP)	17.9056	2.0716	12.0327	23.1445	271036	world
ln(lights)	12.4525	2.3954	3.8918	17.8101	299730	world
ln(population)	9.0597	1.9403	3.6568	14.0869	298275	world
ln(distance)	8.1298	0.818	3.6841	9.4215	299730	world
ln(export)	6.5164	1.8755	0	13.6617	8392	USA
ln(GDP)	11.6768	1.075	9.4748	14.4845	9216	USA
ln(lights)	13.7967	0.7703	11.6883	15.301	9216	USA
ln(distance)	7.1577	0.798	3.4989	8.3645	9216	USA

1.5 References

Anderson, James E., and Eric van Wincoop. 2003. "Gravity with Gravitas: A Solution to the Border Puzzle - 000282803321455214." http://pubs.aeaweb.org/doi/pdfplus/10.1257/000282803321455214.

Arkolakis, C., Costinot, A., and Rodriguez-Clare, A. (2012) "New Trade Models, Same Old Gains?,"American Economic Review, 102(1): 94–130.

Aubrecht, C., C. D. Elvidge, T. Longcore, C. Rich, J. Safran, A. E. Strong, C. M. Eakin, et al. 2008. "A Global Inventory of Coral Reef Stressors Based on Satellite Observed Nighttime Lights." Geocarto International 23 (6): 467–79. doi:10.1080/10106040802185940.

Bernard, Andrew B., Jonathan Eaton, J. Bradford Jenson, and Samuel Kortum. 2000. "Plants and Productivity in International Trade." Working Paper 7688. National Bureau of Economic Research. http://www.nber.org/papers/w7688.

Bharti, N., A. J. Tatem, M. J. Ferrari, R. F. Grais, A. Djibo, and B. T. Grenfell. 2011. "Explaining Seasonal Fluctuations of Measles in Niger Using Nighttime Lights Imagery." Science (New York, N.Y.) 334 (6061): 1424–27. doi:10.1126/science.1210554.

Chaney, Thomas. 2008. "Distorted Gravity: The Intensive and Extensive Margins of International Trade." American Economic Review 98 (4): 1707–21. doi:10.1257/aer.98.4.1707.

Chen, Xi, and William D. Nordhaus. 2011. "Using Luminosity Data as a Proxy for Economic Statistics." Proceedings of the National Academy of Sciences 108 (21): 8589–94. doi:10.1073/pnas.1017031108.

Croft, Thomas A. 1973. "Burning Waste Gas in Oil Fields." Nature 245 (5425): 375–76. doi:10.1038/245375a0.

Eaton, Jonathan, and Samuel Kortum. 2002. "Technology, Geography, and Trade." Econometrica, 1741–79.

Eaton, Jonathan, Samuel Kortum, and Francis Kramarz. 2008. "An Anatomy of International Trade: Evidence from French Firms." Working Paper 14610. National Bureau of Economic Research. http://www.nber.org/papers/w14610.

Elvidge, Christopher D., Kimberly E. Baugh, Eric A. Kihn, Herbert W. Kroehl, and Ethan R. Davis. 1997. "Mapping City Lights with Nighttime Data from the DMSP Operational Linescan System." Photogrammetric Engineering and Remote Sensing 63 (6): 727–34.

Evenett, Simon J., and Wolfgang Keller. 1998. "On Theories Explaining the Success of the Gravity Equation." Working Paper 6529. National Bureau of Economic Research. `http://www.nber.org/papers/w6529`.

Feenstra, Robert C. 2003. Advanced International Trade: Theory and Evidence. Princeton University Press.

Feenstra, Robert C., Professor James, R. Markusen, and Andrew K. Rose. 2001. "Using the Gravity Equation to Differentiate among Alternative Theories of Trade." Canadian Journal of Economics, 430–47.

Head, K. (2003), "Gravity for beginners", mimeo, University of British Columbia

Head, Keith, and Thierry Mayer. 2013. "Gravity Equations: Workhorse, Toolkit, and Cookbook." SSRN Scholarly Paper ID 2210293. Rochester, NY: Social Science Research Network. `http://papers.ssrn.com/abstract=2210293`.

Henderson, J. Vernon, Adam Storeygard, and David N Weil. 2012. "Measuring Economic Growth from Outer Space." American Economic Review 102 (2): 994–1028. doi:10.1257/aer.102.2.994.

Hillberry, Russell, and David Hummels. 2002. "Explaining Home Bias in Consumption: The Role of Intermediate Input Trade." Working Paper 9020. National Bureau of Economic Research. `http://www.nber.org/papers/w9020`.

———. 2003. "Intranational Home Bias: Some Explanations." Review of Economics and Statistics 85 (4): 1089–92.

Huang, Qingxu, Xi Yang, Bin Gao, Yang Yang, and Yuanyuan Zhao. 2014. "Application of DMSP/OLS Nighttime Light Images: A Meta-Analysis and a Systematic Literature Review." Remote Sensing 6 (8): 6844–66. doi:10.3390/rs6086844.

Johnson, Simon, William Larson, Chris Papageorgiou, and Arvind Subramanian. 2009. "Is Newer Better? Penn World Table Revisions and Their Impact on Growth Estimates." Working Paper 15455. National Bureau of Economic Research. `http://www.nber.org/papers/w15455`.

Kloog, Itai, Abraham Haim, Richard G. Stevens, and Prof Boris A. Portnov. 2009. "Global Co-Distribution of Light at Night (LAN) and Cancers of Prostate, Colon, and Lung in Men." Chronobiology International 26 (1): 108–25. doi:10.1080/07420520802694020.

Kloog, Itai, Richard G. Stevens, Abraham Haim, and Boris A. Portnov. 2010. "Nighttime Light Level Co-Distributes with Breast Cancer Incidence Worldwide." Cancer Causes & Control 21 (12): 2059–68. doi:10.1007/s10552-010-9624-4.

Lazar, M. 2010. "Shedding Light on the Global Distribution of Economic Activity." The Open Geography Journal 3 (1): 147–60. doi:10.2174/1874923201003010147.

Pinkovskiy, Maxim L., and Xavier Sala-i-Martin. 2014. "Lights, Camera,... Income!: Estimating Poverty Using National Accounts, Survey Means, and Lights." SSRN Scholarly Paper ID 2384290. Rochester, NY: Social Science Research Network. http://papers.ssrn.com/abstract=2384290.

Rose, Andrew K., and Mark M. Spiegel. 2009. "The Olympic Effect." Working Paper 14854. National Bureau of Economic Research. http://www.nber.org/papers/w14854.

Sutton, Paul C., and Robert Costanza. 2002. "Global Estimates of Market and Non-Market Values Derived from Nighttime Satellite Imagery, Land Cover, and Ecosystem Service Valuation." Ecological Economics 41 (3): 509–27. doi:10.1016/S0921-8009(02)00097-6.

Tinbergen, Jan. 1962. "Shaping the World Economy; Suggestions for an International Economic Policy," January. http://repub.eur.nl/pub/16826.

Westerlund, Joakim, and Fredrik Wilhelmsson. 2011. "Estimating the Gravity Model without Gravity Using Panel Data." Applied Economics 43 (6): 641–49. doi:10.1080/00036840802599784.

Wolf, Holger C. 2000. "Intranational Home Bias in Trade." Review of Economics and Statistics 82 (4): 555–63. doi:10.1162/003465300559046.

Yilmazkuday, Hakan. 2012. "Understanding Interstate Trade Patterns." Journal of International Economics, 86: 158–166.

Young, Alwyn. 2012. "The African Growth Miracle." Journal of Political Economy 120 (4): 696–739. doi:10.1086/668501.

CHAPTER 2

MEASURING INTERNAL TRADE FLOWS AND INTERNAL DISTANCES BY LIGHTS

2.1 Introduction

Internal trade, defined as the trade within a nation, is an integral part of various international trade and economic development analyses. For example, internal trade data have been applied in the gravity equation framework to measure the bilateral trade costs and to decompose them to tariff and non-tariff components (Head and Ries 2001; Jacks, Meissner, and Novy 2008). Furthermore, internal trade becomes widely utilized in studies concerned with the distance puzzle. In fact, the gravity equation fails to support for globalization and the diminishing effects of distance on international trade without incorporating the information on internal trade (Yotov 2012). In economic growth literature, the empirical estimations find that internal trade raises income through several channels, and trade-led income models have to include both international and intranational trade variables. According to Frankel and Romer (1999), "The results also suggest that within-country trade raises income. Controlling for international trade, countries that are larger - and that therefore have more opportunities for trade within their borders - have higher incomes."

Nonetheless, data on internal trade are limited to a few countries and have not been produced consistently for countries over time. Existing literature generally defines internal trade based on market clearing conditions as the difference between income and total exports, in which GDP is usually the proxy for income (Helliwell and Verdier 2001; Yotov 2012). GDP is not, however, a proper measure of income in general and particularly in calculating internal trade. The main reason is that GDP is based on value added, whereas trade data are constructed from gross shipment figures. In addition, a growing part of GDP is services, which are largely non-tradable and are not covered by trade data. Accordingly, the use of GDP data tends to overstate internal trade, and therefore estimated trade costs. Anderson (1979) attempts to exclude non-tradable components of GDP by estimating tradable sectors as a share of GDP; yet, his adjusted GDP still suffers from the value added concept.

A strand of studies replace GDP with data on goods production captured from broad data sets across countries (Novy 2013). Even though data from these kinds of data sets, such as OECD's Structural Analysis (STAN), are comparable with shipment data in definition, they are plagued by issues with data reliability and consistency; the same problem contaminates cross-country GDP data. In addition, using these data requires conversion to one currency by purchasing power parity exchange rates based on prices for a comparable set of goods across countries. In fact, this definition of internal trade requires combining various data sources, each with unquantifiable measurement errors (Henderson, Storeygard, and Weil 2009).

In this paper, I attempt to address the limitations of internal trade data by presenting a new method of estimation, which relies on bilateral trade estimation across regions within a country. This method employs the gravity equation, which relates the bilateral trade flows between two trade partners with their distance and economic scales. Typical empirical estimations use GDP as a proxy for economic size. Because of the problems associated with GDP figures, especially for less developed countries, in addition to the unavailability of these data at subnational level, this study uses a different measure for economic activity: nighttime lights, which is the amount of light that can be observed from outer space. Several recent papers point to the advantages of using lights data as a proxy for economic activity (X. Chen and Nordhaus 2011;Henderson, Storeygard, and Weil 2012; Bleakley and Lin

2012; Storeygard 2013; Pinkovskiy and Sala-i-Martin 2014). Lights data not only are highly correlated with other income measures such as GDP and survey income, but are also available at a far greater degree of geographic fineness than the other economic activity measures.

Estimating bilateral trade flows involves using trade elasticites with respect to economic sizes, distance, and trade costs. I follow the gravity literature by using the simulated method of distance (SMD) estimator to find the optimum elasticites that match the simulated internal trade flows with the benchmarks. This estimator estimates a set of coefficients, by which the constructed internal trade from aggregating the regional bilateral trade flows has the minimum distance from the country-specific economic activity measures. Comparing my internal trade estimations with export-adjusted GDP, the measure usually considered as the internal trade in empirical studies, reveals a high positive correlation. While both measures reflect statistically similar internal trade flows for less developed countries, I find that my measure differs from its alternatives for developed countries and the difference gets larger as GDP rises. The average for GDP-adjusted internal trade for developed countries is larger than my constructed measure by 16 percent.

In section 2, I develop a measure of the internal trade distance. It is an essential factor in studies related to two puzzles; the "distance puzzle" and the "border puzzle". In empirical gravity estimations, the distance puzzle is related to the persistency of the high impact of distance on trade over time with no sign of globalization effects. The border puzzle is interpreted as the excessive trade volumes observed within a nation relative to what would be expected from a gravity equation. Yotov (2012) argues that incorporating the internal trade with its corresponding internal distance into the gravity equation solves the distance puzzle. He finds that after controlling for internal trade, the distance coefficient decreases over time. Head and Mayer (2002) confirm that the border effect is remarkably dependent on how internal distance is measured. Any overestimation or underestimation of internal distance will distort the border effect estimation. Despite the typical measures used in the prior studies being based only on geographical characteristics, they offer a method of internal distance estimation that is based on theoretical trade models. Their equation is virtually a weighted average of distances between regions within a country based on their economic activities. Because GDP is not generally available at state level, they use city-level human population as a proxy for economic activity. Their data set on internal distance, which is part of the CEPII Trade, Production and Bilateral Protection Database, are used in multiple trade literature papers. However, population numbers are collected and combined from multiple national data sets. Although they are sometimes available at a subnational level, the people or agencies that produce these databases have long warned about the uncertainty associated with the estimations in these data sets. Because of wide variations in national census timetables, methods and accuracies, it would be problematic to pool census data from individual countries to provide a data set detailing human population (C. D. Elvidge et al. 1997). Census data accuracy is limited by census takers' access to homes, frequency of repetition, resources, and, sometimes, politically motivated manipulation.

In response to the problems related to using human population as a proxy for economic activity, I consider satellite data on nighttime lights to construct an internal trade distance. Lights data are mostly free from uncertainty surrounding human population data and are available at much higher time frequency. Some studies and international databases even use lights data to estimate or ascertain human population distribution (Christopher D. Elvidge et al. 1997; Dobson et al. 2000; Lo 2001; Xi Chen and Nordhaus 2015).

I construct the internal distance measure following Head and Mayer's approach and use the trade elasticites that are estimated in section 1 of this study. I find that my constructed measure is highly correlated with the population-based measure. The mean of my internal distance data, however, is statistically larger than the mean of the population-weighted internal distance. Although I use the same analytical formulation as Head and Mayer's, using lights data suggests that population data significantly deflates the distance estimations and cannot be considered a safe measure for estimation in regards to the distance puzzle.

This work is perhaps most closely related to the stream of studies focused on measuring bilateral trade costs (Head and Ries 2001; Jacks, Meissner, and Novy 2008). Trade costs are generally estimated using price differences across borders, or are measured directly using the costs of certain items. Recent studies attempt to use various gravity equations to find trade costs based only on observable trade data. In this methodology, estimation of overall trade costs requires information on internal trade on top of international bilateral trade data. In this paper, I offer an internal trade data set as an alternative to what other studies suggest as the difference between GDP and total export (Novy 2013).

This paper is also related to papers that apply the simulated method of distance (SMD) estimator, motivated by McFadden (1989), to the gravity framework. The idea behind SMD is to generate simulated series from the economic model and match them with those computed from the actual data or the benchmark. Simonovska and Waugh (2014) use a version of the SMD estimator, simulated method of moment, to estimate the elasticity of trade with respect to trade frictions. Johnson and Moxnes (2013) estimate technology and trade costs by the same estimator, matching bilateral shipments of final and intermediate goods for sixteen countries. I use SMD to estimate the gravity equation parameters in order to match the constructed bilateral/overall internal trade with the benchmarks, which are GDP and manufacturing production.

2.2 Model

Following this section, I introduce the analytical framework, through which the internal trade and distance measures are constructed. The gravity equation and simulated method of distance are two main elements of my analysis.

2.2.1 Trade Model

In order to construct bilateral trade flows, I follow Anderson and Wincoop's (2003) gravity equation:

$$X_{ij} = \frac{M_i M_j}{y} \left(\frac{\tau_{ij}}{p_i p_j} \right)^{1-\sigma} \tag{2.1}$$

Here, X_{ij} is the bilateral trade between trade partners i and j; M_i denotes the economic size of the exporter and M_j indicates the importer-specific factors that determine the importer's demand; y is the world economic size in the cross-country estimation; τ_{ij} represents the variable trade cost factor on shipment of goods from i to j; P_i is the origin multilateral trade resistance (MTR), which measures the trade costs faced by the origin partner when exporting to a uniform world market; similarly, P_j aggregates the trade costs on the consumers in the destination as if they buy from a uniform world market. Taking logs from equation 2.1, the estimation equation is

$$\ln(x_{ij}) = \alpha + \beta_1 \ln(L_i) + \beta_2 \ln(L_j) + \beta_3 \ln(D_{ij}) + \beta_4 \ln(P_i) + \beta_5 \ln(P_j) + \epsilon \tag{2.2}$$

where the bilateral resistance term, τ_{ij}, is assumed to be a function of distance, D_{ij}. Most studies use economic scale measures such as GDP, GNP, or human population in place of M_i and M_j. In this paper, I utilize the summation of nightnighttime lights illuminated from the origin (L_i) and destination (L_j). Moreover, I consider a linearized version of the MTR term of Anderson and Wincoop's model. The linear form suggested by Baier and Bergstrand (2009) is

$$\ln(P_i) = \left[\sum_{j=1}^{N} \theta_j \ln D_{ij} - \left(\frac{1}{2}\sum_{k=1}^{N}\sum_{m=1}^{N} \theta_k \theta_m \ln D_{km}\right)\right] \tag{2.3}$$

where $\theta_j = \frac{L_j}{\sum_{s=i}^{N} L_s}$ is the relative economic size of exporter.

Unlike many studies in gravity literature, my goal is not to estimate the trade elasticites with respect to covariates in equation 2.2. If the data on regional bilateral trade flows were available, I would aggregate them to compute the actual internal trade. Given the unavailability of data on regional trade flows, my aim is to estimate them using nighttime lights data. In order to achieve that, I first need to predict bilateral trade flow using gravity equation for which I need lights and distances data in addition to the trade elasticites with respect to these variables. There are no actual data on elasticities, and these have to be estimated. One approach is to employ the findings from theoretical and empirical models; for example, unit coefficients for scale and distance variables as suggested by Anderson and Wincoop. The other approach is to use the elasticities estimated by the empirical studies on international or subnational trade flows. In this study, I estimate the elasticities by using an SMD estimator. The remainder of this section describes the specifics of how I do this.

2.2.2 Simulated Method of Distance

In order to find the trade elasticities corresponding to equation 2.2, I use a version of the simulated method of distance (SMD) estimator that minimizes the distance between the predicted internal trade constructed from the simulation process and the benchmark. The estimator maximizes the r-squared (or minimizes the sum of squared residuals relative to of total variation) of regression of the benchmark on the simulated internal trade. The general form of the SMD estimator is

$$\begin{aligned} \hat{\beta} &= \arg\min_{\beta} \frac{\mathbf{e}'\mathbf{W}\mathbf{e}}{\mathbf{y}'\mathbf{M}\mathbf{y}} \\ \mathbf{M} &= \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \\ \mathbf{e} &= \mathbf{y} - \mathbf{y}^* \end{aligned} \tag{2.4}$$

where $\mathbf{W}$ is a continuously updating weighting matrix, $\mathbf{M}$ is the $n \times n$ idempotent matrix that transforms observations into deviations from sample means; $\mathbf{y}$ is the benchmark vector and $\mathbf{e}$ is the vector of residual of regressing the benchmark on the simulated data. In equation 2.4, $\mathbf{y}^*$ denotes the overall internal trade constructed from the gravity equation.

$$y_i^* = \sum_i \sum_{j \neq i} e^{\alpha + \beta_1 \ln(\mathrm{light_i}) + \beta_2 \ln(\mathrm{light_j}) + \beta_3 \ln(\mathrm{D_{ij}}) + \beta_4 \ln(\mathrm{P_i}) + \beta_5 \ln(\mathrm{P_j})} \tag{2.5}$$

y_i^* is the aggregated bilateral trade flows between districts within the country, k. The SMD estimator is applied to estimate the trade elasticities in equation 2.2, through which the overall internal trade for a country has the minimum distance from the benchmark measure. In fact, this optimization process has three stages; at the first stage, the bilateral trades between states within a country are simulated using all possible βs. At the second stage, for each country, the bilateral trade flows are aggregated to construct the overall internal trade. The third stage is to compare the constructed internal trade with the benchmark measures. I consider two benchmarks, GDP and manufacturing productions. A set of coefficients in the simulation process is the optimum that generates the best match between the internal trade and the benchmark.

2.3 Data Sources

The primary source of data I use in this study comes from the nighttime light images captured by the United States Department of Defense satellites and made available to public as the Defense Meteorological Satellite Program Operational Linescan System (DMSP-OLS). The raw data have a high nominal resolution of 0.5 kilometer. I use data that are processed and have been filtered of noise from glare, moonlit, sunlit, aurora, and the clouds. Processed data are obtained from the National Oceanic and Atmospheric Administration, National Centers for Environmental Information, and are at a resolution of 30 arc-seconds, spanning 180 to 180 degrees longitude and -65 to 75 degrees latitude. Although the images are available from 1993, because of computational limitations on processing images and bootstrapping the MSD process, I only use the image from 2007.

The light image in graphic file format is imported into ArcGIS application along with other layers of information on gas flares and administrative divisions. The spatial database on world's administrative divisions are from the GADM database. GADM version 2.8 contains geographic information in shapefile format for 294,430 administrative areas. I process the shapefile and merge all city-level areas to 4,251 divisions at state/province district level for 190 countries around the world. I use other layers of information to clean the gas flares from the images. Geographical information on the location and the shape of gas flares is imported into the ArcMap application and applied to the lights images to remove any noise from gas flares (Christopher D. Elvidge et al. 2009). I then use the Zonal Statistics library in ArcGIS to calculate the light intensity on approximately every one-kilometer square, and aggregate the results to find the light summation for each state in my data set.

Distance calculation involves generating a geographical center for each state and then calculating point-to-point distances between these states. The center of each state is computed based on the weighted average of its geographic coordinates and also the lights distribution. Because Earth has a roughly spherical shape, and the lights images are two-dimensional maps, I have to first project the Earth's surface onto a plane. This process helps me calculate the distance between any two centroids in miles/kilometers instead of arc-minute degrees. I apply the Sphere Equidistant Conic method of projection.

Some other data sets are also used in this analysis. Data on manufacturing production and services are from the United Nations Industrial Development Organization (UNIDO) for 2007. GDP data are obtained from the Penn World Table (PWT) version 6.2. I also use international bilateral trade data to calculate total exports. Bilateral trade data are taken from the IMF Direction of Trade Statistics (DOTS) and denominated in U.S. dollars. The benchmarks on internal distance are from hyperref[2Mayer2005]Mayer and Zignago (2005) and are available on CEPII's webpage.

2.4 Estimation and Results

This section presents the process and the results of constructing the internal trade and internal distance measures.

2.4.1 Internal Trade

I construct the internal trade measure using a four-step SMD estimation:

Step 1, Simulation: Similar to other empirical studies that use SMD, I assume a range of possible values for decision variable. I set a range for positive βs between 0 and 2, and for negative ones between -2 and 0. Each coefficient assumes all possible numbers in this range, with a step size of 0.01. This translates to 320 billion sets (=200^5) of possible coefficients. Because each set requires intensive computations, I first use larger step sizes to limit the range for each variable in equation 2.2. Based on equation 2.2, and for every coefficient set, the bilateral trade between any two states within a country is computed.

Step 2, Aggregation: The bilateral trade is aggregated for all the origin states located within a country. This gives the summation of all inter-state trades for each country and for every set of coefficient βs.

Step 3, Regression: I run the regression of the benchmark on constructed internal trade in Step 2. If I simply run a benchmark such as GDP against overall internal trade, I will obviously get a large coefficient based on variation in country size alone. Therefore, I first normalize both the trade and benchmark by dividing both by the largest number. The coefficient set which has the maximum R-squared is the one I use to construct the internal trade measure.

Step 4, Bootstrap: This step computes standard errors of the coefficients using a bootstrap technique. I compute residuals from Step 3 using the estimated coefficient set. Then, I resample from the residuals with replacement, and generate a new set of data for the left-hand side variable in Step 3. Using a new series of benchmark data constructed from the resampling process, I rerun the process from

Step 1 to 3 and find a new set of estimations. I repeat this procedure 40 times and compute the standard errors as:

$$\text{S.E.}(\hat{\beta}) = \left[\frac{1}{40}\sum_{b=1}^{40}\left(\hat{\beta}^b - \hat{\beta}\right)\left(\hat{\beta}^b - \hat{\beta}\right)'\right]^{\frac{1}{2}}$$

where $\hat{\beta}^b$ denotes the coefficient from bootstrap procedures. This method of constructing standard errors is similar to the approach taken by Eaton, Kortum, and Kramarz (2008). They use a simulated method of moment estimator to estimate the parameters of a trade model of French exporters.

Table 2.1 presents the results from the steps outlined above. The estimations of trade elasticities with respect to the lights, distance and MTR have the expected sign and magnitude. They are close to the empirical results from others who have used international or subnational bilateral trade data. In a meta-analysis review, Head and Mayer (2013) analyzed 2508 estimations from 150 papers and reported the mean for the coefficients of origin's GDP, destination's GDP, and distance as 0.74, 0.58, and -1.1 with standard deviations of 0.45, 0.41, and 0.41. The standard errors from the bootstrap procedure are in Table 2.1. They confirm that the coefficients are estimated precisely; the lowest t-value is for distance coefficient and is about 4.0.

Several proxies for internal trade are considered in trade literature. One natural way of measuring internal trade is to subtract total exports from GDP. Figure 2.1 depicts how my measure compares with the alternative. The graph shows that the correlation between two measures is positive and large ($\rho = 0.80$).

The constructed internal trade data for 2007 is reported in Table A2 in the appendix. The five countries with the largest constructed internal trade are the Unites States, Russia, China, United Kingdom, and France. The GDP-adjusted measure of internal trade shows the United States, China, India, Japan, and Russia have the largest internal trade. The list for the countries with the smallest internal trade is similar from both data sets: Dominica, Tonga, Liberia, Sao Tome and Principe, and Vanuatu. The t-test of difference in the means of these measures rejects the null hypothesis (p-value= 0.03). The test, however, assumes the two measures are normally distributed variables; this might not be the case. I run a ranking test using the Wilcoxon (1945) signed rank sum test, which is essentially a non-parametric t-test. This test assumes that the difference between the two internal trade measures is ordinal. The null hypothesis is that the median of the differences is zero or, equivalently, the proportion of positive (negative) signs is one-half. The test result rejects the null hypothesis and concludes that two measures have statistically different internal trade medians. Accordingly, the conclusion is that while the internal trade measures are highly correlated, they represent different aspects of interaction within a country.

Outlier tests on both measures identify the United States as the country with the largest amount of internal trade. I normalize the data by scaling down both measures using the internal trade data for the Unites States. Accordingly, every country receives a number between zero and one. Figure 2.1 shows the relative location of each country in respect to the United States in terms of internal trade volume. For example, it shows that China, as the second country in terms of internal trade, is at least half way to the United States. In addition, it is above the 45-degree line and receives a larger internal trade figure using the GDP-adjusted measure. Figure 2.2 depicts the value of internal trade reported by both measures, while excluding the United States as an outlier. The graph shows that India, Japan, Brazil, and Canada are located above the 45-degree line, and Russia, the United Kingdom, and Italy are below that line.

Figure 2.3 includes a line from fitted values of a locally weighted regression of GDP-adjusted internal trade measure on my constructed measure. The nonparametric specification suggests that there might be a non-linear relation between the two internal trade measures. I extend this experiment by splitting the countries into two categories "high" and "low", based on the median of constructed internal trade data. Figure 2.4 shows that the fitted lines for the two categories have different slopes. Regression of log of GDP-adjusted measure on the log of my internal trade measure for low trade category reports a coefficient of 0.41 and R-squared of 0.49. The coefficient for the high trade category is 0.81 with the R-squared of 0.71. A quick look at the statistics in Table 2.2 reveals that the high trade category has a similar mean (590 vs. 690) from both measures, considering their standard deviations (100 and 160). The summary statistics for the low trade category show that the internal trade measures suggest considerably different numbers for the countries in this group. The means of my measure and the GDP-adjusted measure are 4 and 25 billion dollar respectively. The t-test of mean comparison on the high trade category cannot reject the null hypothesis of having the same mean (t=1.22). This test, however, rejects the mean equality of two internal trade measures for the low trade category (t=5.37).

One explanation comes from the data quality concerns regarding GDP and total export data. Countries with higher levels of internal trade are generally developed countries, from which the GDP and export numbers are of better quality due to a greater capacity for generating national income data. Thus, I expect to have relatively similar internal trade measures for these countries. Nonetheless, the GDP-adjusted measure would differ from my measure for the low trade category simply because of significant measurement errors in GDP or export data for countries belonging to that category. For example, the Penn World Table, which is the source of the GDP data I use in this study, gives a data quality ranking of C or D to most countries in the low trade category. Chen and Nordhaus (2011) argue that there is great uncertainty surrounding the PWT estimates and find that the margin of error (root mean squared error) for grades C and D is 20 and 30 percent respectively. Accordingly, these inconsistencies and measurement errors are present in the GDP-adjusted measure. Another explanation is related to the large and growing service sector in developed countries. While most of the services are not tradable, they inflate the difference between GDP and total export. The lower slope for low trade categories in Figure 2.4 confirms the fact that the GDP-adjusted measure reports lower numbers for the low trade category relative to the high trade one.

2.4.2 Internal Distance

I base my internal distance estimation on the general formula developed by Mayer and Zignago's (2006), in which internal distance for the country i with n states is

$$\mathrm{D}_{ii} = \left[\sum_{k \in i} \left(\frac{L_k}{L_i} \right) \sum_{l \in i} \left(\frac{L_l}{L_i} \right) d_{kl}^{\beta_3} \right]^{\frac{1}{\beta_3}} \tag{2.6}$$

β_3 is the trade elasticity with regard to distance in gravity equation 2.2 and L is the nighttime light. To ensure consistency with the constructed internal trade measure, I use SMD estimations on distance coefficients (β_3= -0.51 and -1.0). The unity coefficient for distance is what the Anderson and Wincoop model suggests and is also the mean of estimations from other empirical studies (e.g., Yotov 2012, Millimet 2007, and Head et al 2002). Equation 2.6 is similar to the arithmetic average distance formula used by Head and Mayer (2000), Helliwell and Verdier (2001), and Anderson and van Wincoop (2001) if I set β_3=1. However, gravity equation estimations suggest negative elasticities. Using negative β_3 in equation 2.6

delivers a harmonic average, which is less (equal in the case of all numbers being the same) than the arithmetic mean.

Table A2 in the appendix contains the data on internal distance for the year 2007. I use nighttime lights instead of human population or GDP as a proxy for economic activity. As previously discussed, the lights data are more consistent than other sources of economic activity data and also have the advantage of being available at a very small spatial scale. Table 3 gives a sense of several internal distance measures by providing some descriptive statistics. The first three columns use a negative distance coefficient and the forth column is essentially arithmetic weighted distances. Columns 2 and 3 are comparable, in the sense that both use the same calculation formula with the same distance elasticity (β_3=-1); however, column 2 is related to internal distance estimated based on lights data while column 3 shows the statistics from the population-weighted measure.

The description of the internal distance measure computed using the distance between two main states of each country is reported in column 5 in Table 3. Column 3 has the lowest mean and standard deviation among all the columns. To illustrate the degree of difference between using lights and population in equation 2.6, I added the internal distance estimations for three large countries at the bottom of Table 3. The internal distances for China and the United States when using lights are 1141 and 1312 kilometers respectively, while the population-weighted measure suggests 305 and 261 kilometers respectively.

Figure 2.5 depicts the internal distance computed using lights data versus using population data. While the two measures are highly correlated (ρ=0.53), the slope of the graph confirms that the population-based measure gives smaller internal distances than the lights-based measure. Figure 2.6 provides a graphical version of the comparison between measures in Table 2.3. The horizontal axis shows my distance measure while the y axis records alternative measures. The nonparametric fitted lines are based on the last three columns in Table A2. The further below the 45-degree line the fitted line is located, the lower the estimation of internal distance suggested by that alternative measure relative to my measure. The figure shows that the estimation based on the main states, arithmetic mean, and lights are reporting comparable figures for internal distance than the population-based measure (CEPII data set) suggests. While all measures are estimating similar numbers for small countries, the difference gets larger for the countries with larger inter-state distances.

Columns 4 and 5 essentially show estimation methods with no basis in theory, while the equation used in the first three columns is derived from a theoretical model. Column 3 uses population measures, which as previously discussed may suffer from several drawbacks; population data are collected from multiple census sources with various unquantifiable data qualities and collection methodologies. On the other hand, nighttime lights data are produced by a single entity, and have better information on human economic activity at subnational scale than population data. Thus, my constructed data set not only benefits from using an analytical formulation, but also utilizes a source of information on economic activity that is more reliable and is available at higher time frequencies.

2.5 Conclusion Remarks

Internal trade is an important piece of information for explaining income variation and for estimating the effect of trade barriers such as borders and distance on trade flows. However, there are no reliable and constant data on actual trade within countries across the world. In this paper, I develop a statistical framework to construct a new regional bilateral trade data set and overall internal trade data set. Estimates of bilateral trade flows are driven by employing nighttime lights data as a proxy for human economic activity in a gravity equation framework. While there is the potential for error when using lights data, these data have reliability and consistency advantages over other subnational economic activity measures such as population data.

My methodology involves using the simulated method of distance estimator to estimate the gravity equation coefficients by matching the simulated trade data to the benchmark. Given the estimation, I construct a database on bilateral trade flows between any two states around the world, and the overall trade flows within all countries in my data set. I find that my measure highly correlates to the alternative measure but has a lower average. This discrepancy is larger for the group of countries with higher GDP and trade levels; this could be attributed to the large volume of non-tradable services included in GDP data.

Internal distance data can play a key role in those international trade studies concerned with explaining the border effect or distance effect. Head and Mayer (2000) offer an equation from the gravity model to find the best measure of internal distance. Constructing internal trade through their equations requires data on economic activity at subnational level. When there are no reliable and consistent measures of economic activity at subnational and national scale over time, I suggest using lights data. My database is constructed for the year 2007 and can be extended to any year for which there exists lights data. When compared with various internal distance measures, my distance estimations provide a larger and statistically different measure than the CEPII's database, which utilizes regional human populations as the proxy for human economic activity. Questions remaining for future research include how the new database on internal trade and distances will work with estimations from models concerned with the distance puzzle and border puzzle.

Figure 2.1: Internal trade measured by the SMD estimator and the benchmark (GDP – export), data normalized based on the United State's internal trade estimate

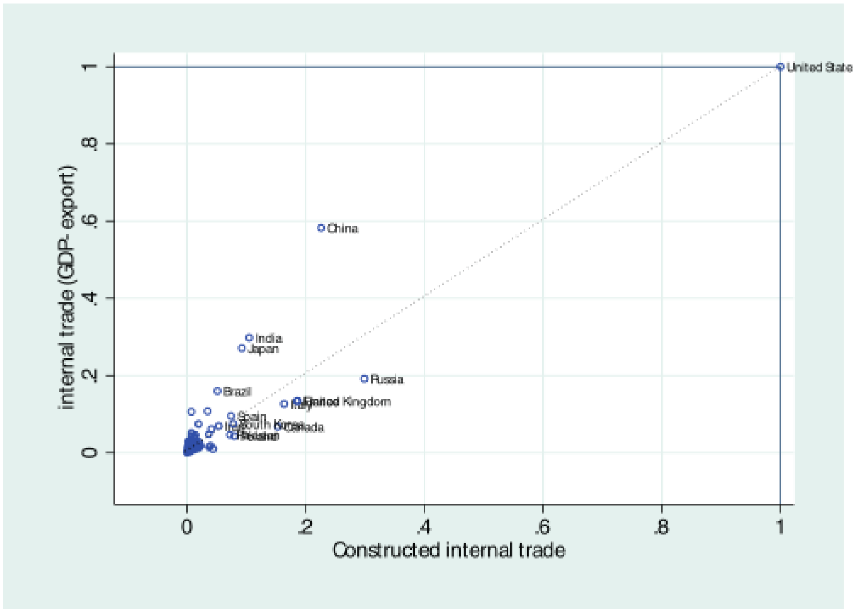


Figure 2.2: Internal Trade vs. Benchmark: 2007

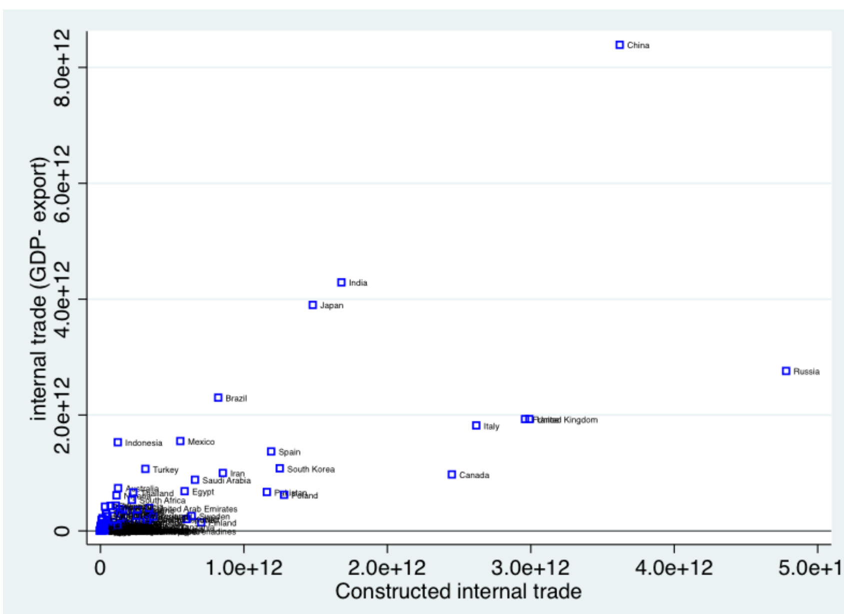


Figure 2.3: Nonparametric Fitted Line for Constructed Internal Trade and the Alternative (GDP-Export)

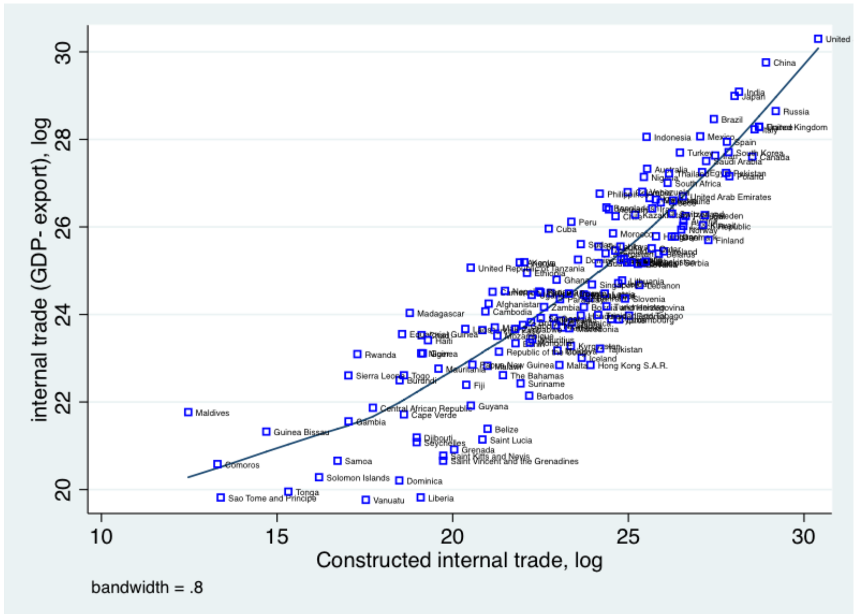


Figure 2.4: Constructed Internal Trade vs. Alternative (GDP-Export), Low and High Internal Trade Countries

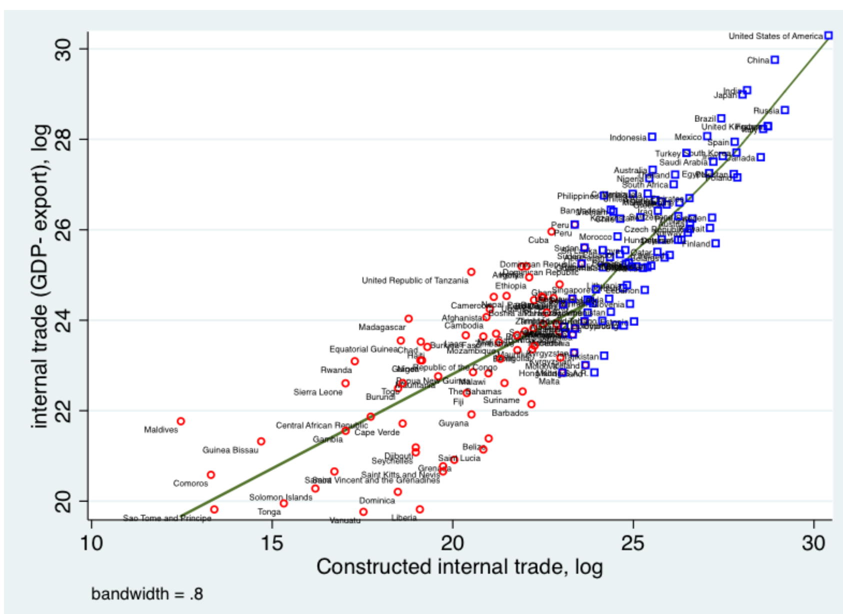
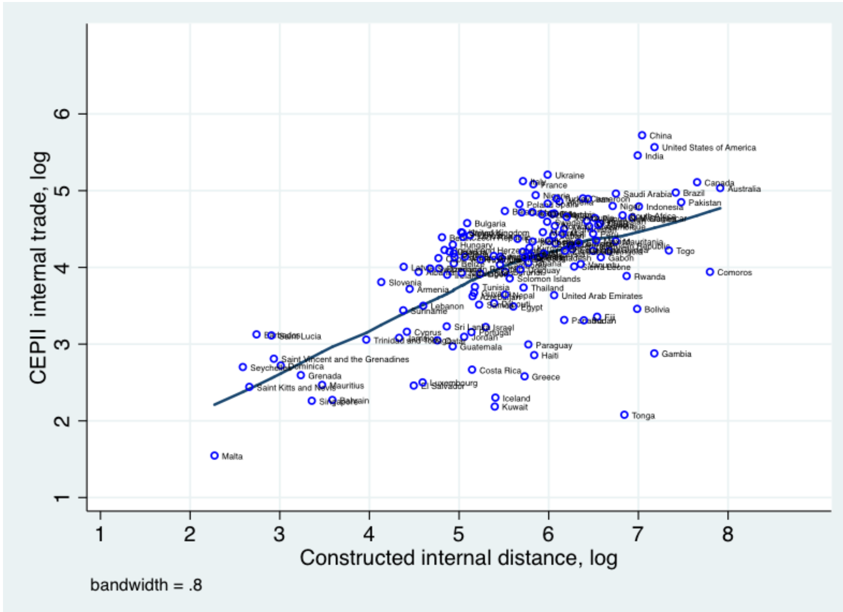
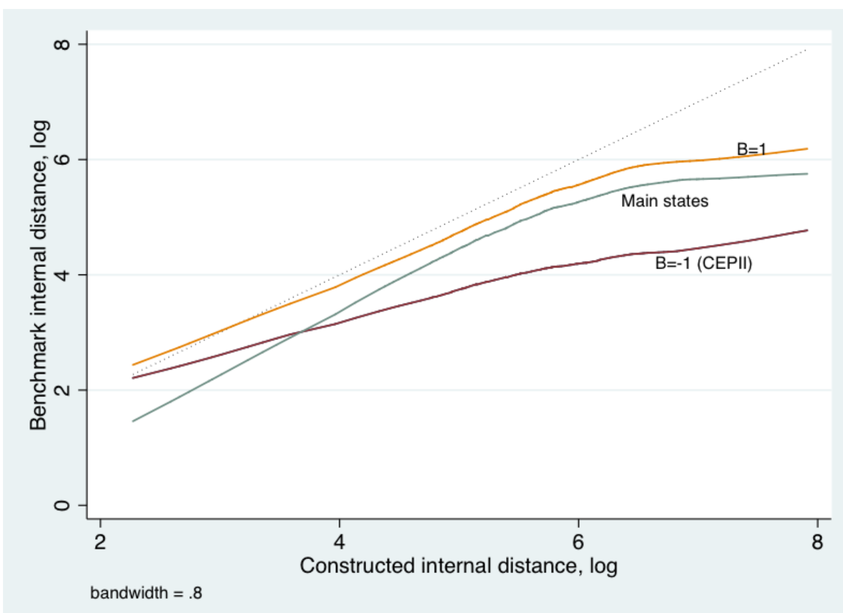


Figure 2.5: Constructed Internal Distance vs. Benchmark



Source: CEPII data set on internal trade

Figure 2.6: Comparison of Alternative Internal Trade Distance



Source: CEPII data set on internal trade

Table 2.1: Simulated Method of Distance Estimations

	light_i	light_j	distance	MTR_i	MTR_j
$\hat{\beta}$(GDP benchmark)	**0.86**	**0.85**	**-0.51**	**-0.75**	**-0.75**
Standard Error	(0.065)	(0.069)	(0.122)	(0.072)	(0.083)
$\hat{\beta}$ (Manufacturing production benchmark)	**0.7**	**0.72**	**-1.08**	**-0.52**	**-0.52**
Standard Error	(0.01)	(0.008)	(0.024)	(0.01)	(0.01)

Note: SMD estimation of gravity equation elasticities for 2007. The benchmark is GDP in the first row and manufacturing production in the third row. Standard errors are from 45 bootstraps.

Table 2.2: Summary Statistics of Internal Trade Measures

	Complete set		*High*		*Low*	
	Constructed	*GDP-export*	*Constructed*	*GDP-export*	*Constructed*	*GDP-export*
Mean	3.50E+011	4.10E+011	5.90E+011	7.00E+011	4.11E+009	2.55E+010
Median	2.00E+010	4.40E+010	1.00E+011	1.60E+011	1.51E+009	1.39E+010
Standard Deviation	1.40E+012	1.40E+012	1.80E+012	1.80E+012	5.24E+009	3.78E+010
Kurtosis	96.03	62.45	57.4	38.2	1	12.2
Skewness	9.02	7.28	7	5.7	1.4	3.2
Range	1.60E+013	1.40E+013	1.60E+013	1.40E+013	1.91E+010	2.20E+011
Minimum	2.60E+005	3.80E+008	1.00E+010	8.30E+009	2.61E+005	3.83E+008
Maximum	1.60E+013	1.40E+013	1.60E+013	1.40E+013	1.91E+010	2.20E+011
# of Countries	175	175	94	94	81	81

Note: The countries are split by the internal trade median. The constructed internal trade data are aggregated bilateral trade data computed by SMD estimations. and for year 2007

Table 2.3: Descriptive Statistics of Internal Distance

		$\beta_3 = -0.51$ (1)	$\beta_3 = -1$ (Lights) (2)	$\beta_3 = -1$ (CEPII) (3)	$\beta_3 = 1$ (4)	Main States (5)
Mean		405.7	277	68.8	290.9	207.8
Median		308.1	231.1	62.5	185	148.3
Standard Deviation		360.4	240.5	49.3	301.5	204
Kurtosis		2.4	4.6	5	7.7	8
Skewness		1.6	1.9	1.7	2.4	2.4
Range		1757.6	1369.3	305.5	1726.3	1157.5
Minimum		9.7	8.3	0.2	11.6	6.2
Maximum		1767.3	1377.6	305.7	1737.9	1163.7
Internal Distance (Km)	China	1142.9	969.9	305.7	1302	1163.7
	Brazil	1663.1	1377.6	144.8	1574.1	1097.4
	United States	1312	1130.1	261.7	1642.4	1161.1

Note: Column 1 is the constructed internal trade using the trade elasticity with regard to distance from SMD estimator. Column 3 is statistics on CEPII data set. The internal distance measure is constructed using city-level great circle distances weighted by population. Column 5 relates to distance between the centroid of the two largest states in a country. All distances are given in kilometers and for year 2007. The distances at the bottom are from constructed internal distance using lights weights

2.6 References

Anderson, James E. 1979. "A Theoretical Foundation for the Gravity Equation." The American Economic Review 69 (1): 106-16. Anderson, James E., and Eric van Wincoop. 2003. "Gravity with Gravitas: A Solution to the Border Puzzle." American Economic Review 93 (1): 170-92. doi:10.1257/000282803321455214.

Baier, Scott L., and Jeffrey H. Bergstrand. 2009. "Bonus Vetus OLS: A Simple Method for Approximating International Trade-Cost Effects Using the Gravity Equation." Journal of International Economics 77 (1): 77-85. doi:10.1016/j.jinteco.2008.10.004.

Bleakley, Hoyt, and Jeffrey Lin. 2012. "Portage and Path Dependence." The Quarterly Journal of Economics, April, qjs011. doi:10.1093/qje/qjs011.

Chen, Xi, and William Nordhaus. 2015. "A Test of the New VIIRS Lights Data Set: Population and Economic Output in Africa." Remote Sensing 7 (4): 4937-47. doi:10.3390/rs70404937.

Chen, X., and W. D. Nordhaus. 2011. "Using Luminosity Data as a Proxy for Economic Statistics." Proceedings of the National Academy of Sciences 108 (21): 8589-94. doi:10.1073/pnas.1017031108.

Dobson, Jerome E., Edward A. Bright, Phillip R. Coleman, Richard C. Durfee, and Brian A. Worley. 2000. "LandScan: A Global Population Database for Estimating Populations at Risk." Photogrammetric Engineering and Remote Sensing 66 (7): 849-57.

Eaton, Jonathan, Samuel Kortum, and Francis Kramarz. 2008. "An Anatomy of International Trade: Evidence from French Firms." Working Paper 14610. National Bureau of Economic Research. http://www.nber.org/papers/w14610.

Elvidge, C. D., K. E. Baugh, E. A. Kihn, H. W. Kroehl, E. R. Davis, and C. W. Davis. 1997. "Relation between Satellite Observed Visible-near Infrared Emissions, Population, Economic Activity and Electric Power Consumption." International Journal of Remote Sensing 18 (6): 1373-79. doi:10.1080/014311697218485.

Elvidge, Christopher D., Kimberly E. Baugh, Eric A. Kihn, Herbert W. Kroehl, and Ethan R. Davis. 1997. "Mapping City Lights with Nighttime Data1 from the DMSP Operational Linescan System." Photogrammetric Engineering and Remote Sensing 63 (6): 727-34.

Elvidge, Christopher D., Daniel Ziskin, Kimberly E. Baugh, Benjamin T. Tuttle, Tilottama Ghosh, Dee W. Pack, Edward H. Erwin, and Mikhail Zhizhin. 2009. "A Fifteen Year Record of Global Natural Gas Flaring Derived from Satellite Data." Energies 2 (3): 595-622. doi:10.3390/en20300595.

Frankel, Jeffrey A., and David Romer. 1999. "Does Trade Cause Growth" American Economic Review, 379-99.

Head, Keith, and Thierry Mayer. 2002. Illusory Border Effects: Distance Mismeasurement Inflates Estimates of Home Bias in Trade. Vol. 1. CEPII Paris. http://michau.nazwa.pl/aska/uploads/Studenci/illusionarybordereffect.pdf.

Head, Keith, and John Ries. 2001. "Increasing Returns versus National Product Differentiation as an Explanation for the Pattern of U.S.-Canada Trade." American Economic Review 91 (4): 858-76. doi:10.1257/aer.91.4.858.

Helliwell, John F., and Geneviève Verdier. 2001. "Measuring Internal Trade Distances: A New Method Applied to Estimate Provincial Border Effects in Canada." Canadian Journal of Economics/Revue Canadienne D' èconomique 34 (4): 1024-41. doi:10.1111/0008-4085.00111.

Henderson, J. Vernon, Adam Storeygard, and David N. Weil. 2009. "Measuring Economic Growth from Outer Space." Working Paper 15199. National Bureau of Economic Research. http://www.nber.org/papers/w15199.

Henderson, J. Vernon, Adam Storeygard, and David N Weil. 2012. "Measuring Economic Growth from Outer Space." American Economic Review 102 (2): 994-1028. doi:10.1257/aer.102.2.994.

Jacks, David S, Christopher M Meissner, and Dennis Novy. 2008. "Trade Costs, 1870-2000." American Economic Review 98 (2): 529-34. doi:10.1257/aer.98.2.529.

Johnson, Robert C., and Andreas Moxnes. 2013. "Technology, Trade Costs, and the Pattern of Trade with Multi-Stage Production." Unpublished Paper, Dartmouth College. http://www.freit.org/RMET/2013/SubmittedPapers/Andreas_Moxnes14.pdf.

Lo, C. P. 2001. "Modeling the Population of China Using DMSP Operational Linescan System Nighttime Data." Photogrammetric Engineering and Remote Sensing 67 (9): 1037-47.

McFadden, Daniel. 1989. "A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration." Econometrica 57 (5): 995-1026. doi:10.2307/1913621.

Novy, Dennis. 2013. "Gravity Redux: Measuring International Trade Costs with Panel Data." Economic Inquiry 51 (1): 101-21. doi:10.1111/j.1465-7295.2011.00439.x.

Pinkovskiy, Maxim L., and Xavier Sala-i-Martin. 2014. "Lights, Camera,... Income!: Estimating Poverty Using National Accounts, Survey Means, and Lights." SSRN Scholarly Paper ID 2384290. Rochester, NY: Social Science Research Network. `http://papers.ssrn.com/abstract=2384290`.

Simonovska, Ina, and Michael E. Waugh. 2014. "The Elasticity of Trade: Estimates and Evidence." Journal of International Economics 92 (1): 34-50.

Storeygard, Adam. 2013. "Farther on Down the Road: Transport Costs, Trade and Urban Growth in Sub-Saharan Africa." SSRN Scholarly Paper ID 2265755. Rochester, NY: Social Science Research Network. `http://papers.ssrn.com/abstract=2265755`.

Yotov, Yoto V. 2012. "A Simple Solution to the Distance Puzzle in International Trade." Economics Letters 117 (3): 794-98. doi:10.1016/j.econlet.2012.08.032.

CHAPTER 3

GROWTH THROUGH INTRAREGIONAL VERSUS INTERREGIONAL TRADE

3.1 Introduction

How does the trade of one region *with* others affect the standard of living of people in that area? The other question I attempt to answer is whether, and how much, the trade *within* a region impacts the income. Most economic studies answer the first question by suggesting a positive linkage at national scale. Yet, despite receiving a great deal of attention in the literature, these studies are not without their critics.

Endogeneity is the main problem in estimating the relation between trade and income. Higher income in a region due to reasons other than trade may be associated with higher trade. Estimations are then biased because of the omission of variables, reverse causality or the simultaneity issues.

In response to the endogeneity problem, a seminal study by Frankel and Romer (1999) suggests using the geographic instrument. They construct an instrument for aggregate trade by using geographic characteristics such as proximity and population. The instrument is free of reverse causality; however, given the use of a single cross section, they are not able to control for omitted variable biases. In this study, I use panel estimation to control for any time invariant correlated with income. Thus, identification in this approach comes not only through the cross sectional, but also through time series variation.

In addition to controlling for region-specific effects, I use data at the regional level. This study uses data on bilateral trade that takes place between states of the United States. There are two benefits of examining income and trade relation at subnational level; the first advantage is related to data quality. The trade or income data I use are produced by a single entity over time. The data utilized in international studies are those produced by countries around the world. They have different statistical capacities to collect and process the national income accounts. An illustration of the degree of measurement error in standard international data on income comes from a study by Dawson et al. (2001) of the Penn World Tables (PWT) Data. They claim that the suggested empirical relation between GDP volatility and income growth in PWT data is just a product of measurement error in annual income. Beside the difficulties in measuring the nominal GDP, data quality on domestic price indices and purchasing power parity (PP) exchange rate, required to measure the real GDP growth, are the other sources of uncertainty. In this study, I use data produced on a consistent basis for a single country with well-established statistical entities and hence, there may be fewer concerns about measurement errors. Second, there is less heterogeneity among states within a country relative to the countries around the world. In fact, I can examine the effect of trade on income at the subnational level in the absence of some other factors such as institution, borders, colonial background, language, and currency.

The distinctive aspect of this paper is to examine the effect of trade within a region on income. For the majority of analyses, the empirical trade research is synonymous with the use of international trade and national income data. Nonetheless, many of the economic interactions and growth variations occur within, rather than between, regions. Just as the trade between a region's residents with those of another region impacts on their income, it may also be influenced by the amount the residents trade with one another. Because the information on trade within a country are unavailable on a consistent basis, the studies at national level are unable to analyze the direct impact of *internal* trade on income. Although Frankel and Romer emphasize the role of internal trade in their model for explaining income, they are unable to identify the internal trade effects because the data on internal trade do not exist across countries. I use Commodity Flow Survey (CFS) data on subnational bilateral trade, which includes the amount of trade that takes place within each state of the United States. Thus, I are able to experiment whether, and how much, the trade within a state improves its residents' income.

This paper contains two sections. I examine whether the geographic characteristics can describe the trade pattern at regional scale. The gravity equation is the workhorse used often in international trade studies to estimate the bilateral trade flows between trade partners. The fitted values from the gravity regressions represent the component of trade related to geography. In order to construct valid instruments to be used at the second stage of the growth regression, GDP as a typical measure of economic size in gravity equations is replaced with geographic variables. In particular, economic size and proximity are the main covariates in the gravity equation.

Given the bilateral trade predictions from the first section, I aggregate the fitted values to construct the instruments. Because the gravity equation includes only the geographic variables, the instruments reflect the geographic component of overall trade. The results are consistent with cross country analysis and suggest that the geographic characteristics are important determinants of overall trade at subnational level. Not only does the geography explain the interregional trade flow, so does the intraregional trade. Actually, a larger region trades more internally simply because there are more people with whom to trade.

The second section employs several instruments including the ones constructed in the first section to examine the trade effect on income. At first, the constructed trade instruments are used to estimate the regressions of income per capita on inter- and intraregional trade by applying a two stage least square method. Comparing the results with ordinary least-square estimates, I find that in all specifications, instrumental variable (IV) estimation of trade impact on trade is larger than the OLS results.

At the second stage, the trade within a state has a large and positive effect on the income per person. Increasing the ratio of internal trade to total export by one percentage point raises the income per capita by three to five percent. Relative to within-state trade, trade with other states has larger point estimates, but is statistically similar and is within the internal trade error bands.

Identification through the heteroskedasticity method introduced by Rigobon (2003) is another method I use to examine whether the relationship between income and trade is robust. This approach exploits the differences in regions in an econometrics framework to identify the parameters of a system of simultaneous equations. This approach is applied in studies where other sources of identification, such as instrumental variables, are not available (e.g., Lee, Ricci, and Rigobon 2004; Lewbel 2012; Ehrmann, Fratzscher, and Rigobon 2011). In my studies, the estimation from this method confirms the estimation from IV; both trade measures affect income significantly and nearly the same as estimations reported by IV.

Instrumentation via lag of trade ratio is another approach I borrow from the literature (Dollar and Kraay 2001). In this approach, I address the endogeneity problem by instrumenting the internal and interstate trade changes via their lagged levels. It shows that the instruments are highly corrected with the changes of trade ratios and results from this instrument are similar to those from other methods I applied.

Furthermore, following the approach suggested by Storeygard (2013), I use the world oil price as a proxy of transport cost. I thus considered two components for the variable transport costs in the gravity equation: 1) the world price of oil, which varies across time but not across states, and 2) the area and distance between states, which varies only across states. Consistency in finding positive effects of trade for interregional and intraregional trade implies that the trade-income association is robust to changes in specification, the type of instrument applied, and the method of estimation.

Literature Review

This study contributes to a rich literature concerned with estimating the trade effect on income. Theoretical trade models discuss several channels by which trade can impact income. It improves the channels through which the technological innovations are diffused between trade partners (Barro and Sala-i-Martin 1997). Trade brings about an increase in productivity through higher competition in domestic markets (Wacziarg 2001). Alcalá and Ciccone (2004) also argue that access to larger markets, particularly for smaller local markets, is another source of trade effect on income. In addition, trade may contribute to income by potential benefits of increasing return to scale (Ades and Glaeser 1994) or by creating incentives to adopt better trade or macroeconomics policies (Bassanini, Scarpetta, and Hemmings 2001).

There is a debate over whether these theoretical implications hold empirically as well. Dollar (1992), Sachs and Warner (1995), Edwards (1998), and Ben-David (1993) are among others who advocate a positive trade-income nexus. All of these studies are subject to some empirical criticism including endogeneity of trade (Rodriguez and Rodrik 2001). In order to address the reverse causality problem as a source of endogeneity, Lee et al (2004) apply the "identification through heteroskedasticity" methodology to estimate the effect of trade openness on growth. While controlling for the reverse causality by solving a system of simultaneous equations, they find a positive but smaller effect of trade on growth than what is suggested by previous studies. In this study, I apply this methodology at the subnational level and find that a rise in trade improves the income per person significantly, which is different from the cross-country analysis.

Frankel and Romer deal with endogeneity by using geography variables such as distance, population and area to estimate a gravity equation and then demonstrate that predicted trade has enough information to explain the income per person differences. This work has initiated a stream of empirical works since two decades ago, and several attempted to use geographic characteristics in trade and growth estimations.

Yet, in a review of most influential studies, Rodriguez and Rodrik (2001) show that the results from the cross-sectional studies are not robust after controlling for omitted variables such as distance from the equator or institutions. Noguer and Siscart (2005) attempt to address the critics by using geographic controls in the second stage. That study proposes a lower effect of trade on income, but still does not eliminate it.

In response to Rodrigues and Rodric, some other studies include institutional variables in IV regression. They attempt to explicitly control for factors, such as institutions that might be associated with both the geography and income. Frankel and Rose (2002) run the income regression on trade with and without additional controls for both geography and institutions. They find that Frankel and Romer's findings on the positive effect of trade on income are robust to the inclusion of additional control variables. Feyrer (2009) addresses the omitted variable problem by generating a time-varying geographic instrument. The instrument is constructed based on the heterogeneity among countries on benefiting from improvements in aircraft technology. The instrument is generated using the changes in the effects of air and sea distance on trade over time, as a result of changes in transportation technology. This approach allows him to use country fixed effects to eliminate the bias from time invariant variables such as distance from the equator, institutions and the like. I follow this approach to construct a time varying instrument to control for all state specific differences. Unfortunately, Feyrer's estimates are not directly comparable to my estimates, as I use trade share measure while he uses the level of trade as an explanatory variable.

I also contribute to an emerging literature examining the effect of trade on income at the subnational scale. Topalova (2005) studies how trade at the subnational level can impact the income of regions with different degrees of exposure to trade liberalization. He finds that trade effects are not equal across regions in India and some areas benefited more from liberalization. Another study by Donaldson (2010) examines how construction of a new railroad network in India reduced trade costs and as a result increased each district's opportunities to trade and, accordingly, income. He finds that interregional trade increases the real income per person significantly. Storeygard (2014) explores the effects of inter-city trade on the income of Sub-Saharan African cities. His findings show that the regions with better access to port cities benefit more from terms of trade shocks. He uses world oil price fluctuations as a proxy for annual changes in transport costs. I borrow his approach in this study to investigate the impact of internal and interstate trade on growth.

3.2 Constructing the Instrument

This section explores the relation between income and trade based on a three-equation model for which Frankel and Romer provide empirical support; although my approach introduces a number of variants. The basic idea is that the average income in each state depends on the economic interactions with other states ("inter-state" trade) and economic interactions within each state ("intra-state trade") as well as other parameters.

$$\ln(Y_{i,t}) = \alpha + \beta \ln W_{i,t} + \gamma \ln T_{i,t} + \epsilon_{i,t} \tag{3.1}$$

where $Y_{i,t}$ is real income per capita, $W_{i,t}$ is intra-state trade, $T_{i,t}$ is inter-state trade, and $\epsilon_{i,t}$ is the residual term. Intra-state trade is itself a function of state's size $S_{i,t}$ and other factors.

$$\ln W_{i,t} = \alpha_0 + \lambda \ln S_{i,t} + v_{i,t} \tag{3.2}$$

I know from gravity equations that trade between two trade partners can be estimated using the economic size and the trade barriers such as proximity $P_{i,t}$.. Distance and remoteness are two main determinants of proximity in gravity equations.

$$\ln T_{i,t} = \alpha_1 + \varphi \ln P_{i,t} + \sigma_{i,t} \tag{3.3}$$

As I do not consider all determinants of income in equation 3.1, the error terms in these three equations are likely not independent from each other and might be correlated. For example, better infrastructure in transportation has an impact on income and on both the internal and interstate trade. Yet, I assume that the geographic terms in equations 3.2 and 3.3 are independent from the error terms. Size and proximity impact income through trade and are not changed by income or other income determinants. I assume that there is no correlation between $\epsilon_{i,t}$ and geographic characteristics, $P_{i,t}$ and $S_{i,t}$. Because these two measures are correlated with intrastate and interstate trade, I use them as instrument variables to reach an unbiased estimation of equation 3.1. In contrast to Frankel and Romer, CFS includes data on trade within each state and it lets us identify λ and φ, the effect of intra- and inter-state trade on income.

Proximity is the key variable I use to construct the instrument for interstate trade. Distance is one natural measure of proximity. Each state has different distance to all other states and I need to generate an average distance to be considered as an instrument for overall interstate trade ratio. The distance of one state from its trade partners receives weights based on the economic size of its partners; distance from a state with large markets and higher GDP affects trade more than the same distance from a state with limited markets. In the next section, I explain how I drive the distance weights using a gravity equation.

I consider two measures of trade, which are standard in literature; trade ratio of total economic size and log of trade level. Trade ratio is usually measured by summation of import and export as a ratio of GDP. The second trade ratio for interstate trade is the summation of export to and import from other states divided by total traded production, the summation of internal trade and export. The identification strategy of equation 3.2 comes from employing two methodologies. As I show in the next section, the land size of each state explains a large part of internal trade differences among states. Yet, because I consider panel estimation, it requires some time varying instruments. I use area and population as measures of size; however, it might be a problematic instrument for internal trade at the subnational scale, given higher migration mobility among the states within a country. In order to control for inter-state migration, I abstract from any changes in population density and focus on the sum of the coefficients of population and area, both in log form (Frankel and Rose 2002). The second approach is to use transport cost as an instrument for internal trade. I follow studies on the impact of oil price on trade costs and income (e.g., Mirza et al. (2009), Storeygard (2013), Below et all, 2015). As oil prices rise, so do the transport costs. These studies find that the changes in transport costs due to changes in oil price have a critical implication for regional growth. For example, Storeygard (2013) considers instrumenting the exogenous annual changes in transport costs by oil price fluctuations and finds a positive trade-led growth at subnational level in Africa. Employing this methodology, the interaction of log world oil price and log distance are used in the gravity equation to construct the interstate trade instrument. Interaction of log oil price and log area are an instrument for intrastate trade. Accordingly, I consider the variable transport costs as two components: the price of oil, which varies across time but not across states, and the road distance between states, which varies across states but not time. I virtually assume that no state in the sample is capable of substantially affecting world oil prices individually.

3.2.1 The Gravity Model

Empirical and analytical models show that the distance between two trade partners can explain most of the bilateral trade flow between them. The empirical findings confirm that the trade is negatively correlated with the distance, which itself is a proxy for trade costs. Although originating from analogy with gravity equation in physics, the gravity equations have been an essential integral of empirical trade literature. Anderson and Wincoop (2001) is the first study that introduces a theoretical foundation for the gravity equation. Recent studies show that the gravity equation emerges from most of the mainstream trade models (Feenstra 2003). The gravity equation says that trade rises proportionally with economic size of trade partners and declines with trade costs such as distance and tariffs. My goal in this section is to apply the geographic factors as proxy for economic size and trade costs in the gravity equation and construct an exogenous trade instrument. A minimal version of the gravity equation I use is as

$$\ln(T_{i,j,t}) = \beta_0 + \gamma_{i,t} + \gamma_{j,t} + \gamma_t + \beta_1 \ln S_{i,t} + \beta_2 \ln S_{j,t} + \beta_3 t_{ij,t} + \mathrm{B} X_{ij} + \vartheta_{ij,t} \quad (3.4)$$

where $S_{i,t}$ and $S_{j,t}$ are measure of states' sizes, $t_{ij,t}$ is the trade costs between two states, and X_{ij} is the set of time-invariant control variables. In most studies, GDP is a proxy for the economic size. Using GDP in generating fitted trade values contaminates the instrument with income information. I replace the GDP with two geographic measures of size: population and area. My goal is to generate instruments for trade using the geographic information exclusively. Therefore, the purpose of equation 3.4 is not comparative statics analysis on the economic size or distance as it is a common goal in gravity equation studies.

Equation 3.4 considers control variables (X_{ij}), which include dummy variables for states' adjacency and internal trade. Several studies on subnational trade find that sharing a border at the subnational level has a significant impact on bilateral trade (Wolf 2000; Hillberry and Hummels 2002).

In analytical models, $\gamma_{i,t}$ and $\gamma_{j,t}$ are multilateral trade resistance terms (MTR). The rationale for including MTR terms is that two trade partners surrounded by other states with large economic sizes will trade less between themselves than if they were remote. I account for MTR terms by including a linearized version of the system of price equations introduced by Anderson and Wincoop. Baier and Bergstrand (2009) evaluate the first-order log-linear Taylor-series of theoretical motivated MTR as

$$\ln \gamma_i = \left[\sum_{j=1}^{N} \theta_j \ln t_{ij} - \left(\frac{1}{2}\right) \sum_{k=1}^{N} \sum_{m=1}^{N} \theta_k \theta_m \ln t_{km} \right] \quad (3.5)$$

Time subscripts are omitted for simplicity. t_{ij} denotes the trade costs, which are road distance or the interaction of distance and oil prices. $\theta_i = \frac{S_i}{\sum_{j=i}^{N} S_j}$ is the economic size of the state$i$ relative to the whole country.

3.3 Data and Results

I use the bilateral trade data from the Commodity Flow Survey (CFS) in a panel of 4 periods of 5 years each, spanning from 1992-3 to 2007. The trade data set consists of 8392 bilateral trade observations for 48 states within the United States (excluding Hawaii, District of Columbia and Alaska, as standard in the corresponding literature). The CFS is a shipper-based survey reporting the trade between and within U.S. states. Road distances between the geographical centers of states are collected from Google Map. Inflation and Real GDP per capita data are from the Bureau of Economic Analysis. Education and population information comes from the Bureau of Census.

3.3.1 Trade Regression Results

The first step to construct instrument variables for the actual interstate trade is to estimate the gravity model using equation 3.4. Table 3.1 shows the OLS regression results. The first column is the point estimations and the corresponding standard errors for the variants in equation 3.4. Column 1 uses MTR terms computed by equation 3.5. I consider state fixed effects in column 2 in place of MTR, which it implicitly assumes are constant over time. Both regressions have time dummies and standard errors that are clustered at the bilateral pair level.

Significance and similarity of coefficients, as well as the high goodness of fit in both columns, tell us that geographic characteristics have enough information to explain the bilateral trade variations. The coefficients of determination (R-squared) are 0.64 in column 1 and 0.92 in column 2.

The results closely match the findings from studies at national level. For example, the log of distance as a proxy for trade costs has a large negative effect on trade. Both coefficients of scale variables in column 1 are significant. Because area is constant over time, when state effects are included in column 2, it becomes insignificant. The home bias variable, which is a dummy with a value of one if trade takes place within the state and zero otherwise, has a significant coefficient in both columns. Adjacency dummy enters highly significant. States adjacent to each other trade more by a factor of 2.5; results that are comparable with Wolf's study, in which GDP is the measure of economic size. In the next section, I construct the trade instrument by aggregating the fitted values for each state and examine if it has enough information to explain the total trade.

3.3.2 Aggregate Trade

Fitted values from specification in column 1 of Table 3.1 are generated for all 48 states over the four time periods. The total constructed trade is the fitted values, which are unlogged and then summed for each state. I also summed all actual bilateral trade volume to obtain overall actual trade.

$$\hat{T}_{i,t} = \sum_{i \neq j} e^{\hat{\beta}_0 + \hat{\gamma}_{i,t} + \hat{\gamma}_{j,t} + \hat{\gamma}_t + \hat{\beta}_1 \ln S_{i,t} + \hat{\beta}_2 \ln S_{j,t} + \hat{\beta}_3 d_{i,j} + \hat{B} X_{ij}} \tag{3.6}$$

Because I assume the homoscedasticity of error terms in the gravity equation ($\vartheta_{ij,t}$), they are the same for all observations, which means $E(e^{\vartheta_{ij,t}})$ assumes a constant number. As I use the log of $\hat{T}_{i,t}$ in the second stage, any constant multiplier of $E(\hat{T}_{i,t})$ goes to the regression constant and would have no implication. Thus, given my assumption of homoscedasticity of error terms, I omit the error term. Time dummies in equation 3.6 control for growth rate shared by all states. As discussed before, the idiosyncratic time variation is from the changes in trade costs or population changes.

In my panel data, as long as there is a single observation for a pair of states in any time cycle, an estimate for the bilateral trade will be generated, as distance is always available. Hence, there will be some observations with fitted values, but not an actual trade number. It causes problems as in IV regression, the actual and instrument variables need to be matched. I use Feyrer's (2009) suggestion by imputing the missing actual trade using a full set of state and time dummies.

3.3.3 Quality of Instrument

Figure 3.1 illustrates the partial association between internal trade and the two size measures. Both graphs show positive and strong relations; population and area have high correlation with interstate trade, 0.90 and 0.51 respectively. Regressing of log intrastate trade on constant, log population and log area yields positive and significant coefficients and high R^2 of 0.76.

Figure 3.2 plots the constructed trade from column 1 of Table 3.1 and the actual trade for 2007. The scatterplot illustrates that constructed trade using geographic information explains much of the variance in actual trade. The correlation between two variables is 0.71. The first column in Table 3.2 shows that $\hat{T}$ becomes significant and with the coefficient of 0.8 as an explanatory variable for T in the OLS estimation. The r-squared of 0.9 confirms high explanatory power of the regression.

The constructed trade measure has a high correlation with size measures. Running a regression of $\ln\hat{T}$ on log of area and log of populations reports the R^2 of 0.88

and statistically significant coefficients. The coefficient in front of the area is negative, though. It implies that a bigger state in terms of land size has more distance from other states and it leads to less interstate trade, after controlling for other factors such as population and MTR. The summation of population and area has a significant coefficient of 0.5.

To examine if geographic variables have information beyond the size variables in order to explain the interstate variable, I regress the T on size measures with and without constructed trade. Comparing columns 2 and 3 tells us that population is an important factor in explaining the interstate trade variations. When constructed trade is added to column 2, $\hat{T}$ remains significant and falls by more than a half; the regression results imply that the instrument for interstate trade contains information beyond the size measure I use as an instrument for intrastate trade.

Figure 3.3 shows that there is still a strong positive partial correlation between the interstate trade and the constructed trade given the size measures. California and Florida are two outliers with high predicted trade, after controlling for the population and area. Tennessee is also an outlier with large actual trade relative to the constructed measure.

3.3.4 Trade Effect on Income

Table 3.3 reports the growth regression results. Column 1 runs the OLS regression of log income per capita on a constant, intrastate and interstate trade ratios, time dummies, and size variables. The results show a statistically and economically significant relation between intrastate trade and income. The estimated coefficient of internal trade indicates that an increase of one percentage point in intrastate trade share is associated with an increase of 2 percent in income per capita. Controlling for intrastate trade, the interstate trade coefficient is insignificant. If interstate trade share enters solely in the regression without the presence of intrastate trade share, the coefficient will turn positive but still remains statistically insignificant; it suggests, as expected, there is some correlation between the two trade measures. Besides, population in column 1 has a negative coefficient, in spite of the positive point estimates in other columns of Table 3.3.

Column 2 shows the IV estimates of both stages, in which trade measures are treated as endogenous. Following Frankel and Romer, the constructed trade share and the log of population are instruments for interstate and intrastate trade, respectively. Moving from OLS to IV increases the estimated effect of size and trade on income. The coefficient on interstate trade is significant and rises sharply while the intrastate trade remains positive but insignificant. The point estimates imply that the OLS understates the effect of trade, similar to Frankel and Romer's findings. In addition, the coefficient of population at the first stage is what I expect from other studies. It is positive, significant and also is estimated with more precision ($t = 5.7$). However, it is much lower than what it is estimated using cross-country data. Increasing the population by one percent translates to an increase in income per capita of 0.05 percent, while Frankel and Romer find a much larger effect of 0.2 percent. Column 3 reports the IV estimates of the same specification but considers the interaction of log of area and log of population. At the first stage, both coefficients are positive and statistically significant. Higher population is associated with higher trade within a state. However, one might expect a negative area impact on internal trade after controlling for population. The positive coefficient of area in column 3 might be due to sampling error. There is another possibility that area has a negative effect on internal trade, but a larger positive impact through increased natural resources. If I focus on the sum of log of area and log of population, the coefficient of 0.08 reflects the effect of size while population density is held constant. The first-stage F-tests on excluded instruments on interstate and intrastate trade are 33 and 61, confirming the instruments are sufficiently strong. The coefficients are much more precisely estimated when log of area is added to column 2.

An alternative instrument for intrastate trade can be generated through the same method I use to construct the instrument for interstate trade. I construct an internal trade instrument using the gravity equation, but it requires a measure for internal distance. I use the measure employed by Wei (1996), which is one-half of the distance from the domestic capital to the capital of the nearest trade partner. Another proxy I consider is the square root of each state's area.

Adding the constructed intrastate trade measure from the gravity equation raises the trade coefficients slightly in column 4. I also find that the results are not sensitive to the proxy of internal distance used at the first stage. The confidence intervals of the intrastate and interstate trade shares have overlap, which implies both coefficients are the same statistically. Column 4 predicts that a one percentage-point increase in intrastate trade and interstate trade share increases the income per person by 5 and 6 percent, respectively.

3.4 Robustness

How sensitive is the trade-led income estimation to some changes in basic model specifications? Table 3.4 presents some robustness along five dimensions. First, I consider the average of trade ratios and the income at five-year intervals from 1993 to 2007 (except 1993-97, which is a four-year interval). Real GDP per capita and intra- and inter-state trade shares are averaged over the time intervals. Therefore, the panel includes data on 48 states over three periods. Column 1 of Table 3.4 shows the estimation results utilizing the instruments in column 3 of Table 3.3 but with the trade ratios for the first year (for example, trade ratios of year 1993 for the 1993-97 cycle) and long-term income growth. When one compares these two columns, it is evident that point estimates for inter-state and intra-state trade ratios are as anticipated. Both trade measure estimates are positive, and, while the inter-state trade measure is significant, the intra-state trade is marginally significantly different from zero ($t = 1.87$). The coefficient of determination at the second stage shows a lower explanatory power of growth regression.

A variation to column 1 is to use GDP measure as a measure of income. Following Frankel and Rose, I consider inflation, population, high school attainment and initial income as the other determinants in the second stage. Applying the average values and also lagged variables reduces the observations to 96. The IV estimation of internal trade ratio is still positive and significant in column 2, while interstate trade is not significant anymore. As expected, education has a large positive effect on a state's GDP growth. Negative initial log of GDP is intuitive and comparable to other studies' findings on growth convergence at subnational scale. As an alternative way of addressing the endogeneity of repressors, I follow Dollar and Kraay (2003) where they suggest estimation the growth in first difference and instrumenting via lagged level of endogenous variables. Thus, instead of using constructed trade ratios and size measures, lag of internal and inter-state trade ratios are considered in column 3. Because I use the lag variables, the number of observations without missing values drops to 144. The positive and significant coefficients again confirm the findings in Table 3.3. Even though the point estimate for internal trade is similar to column 1, the interstate trade ratio and its explanatory power decrease considerably relative to basic results in the previous section.

Third, I use oil price as an instrument for trade variables. As I discussed, oil price is a major determinant of transport costs and can impact the trade volume. Because the data is at subnational level, I assume that the individual income per person does not affect the world oil price. Accordingly, I use the interaction of log of distance and log of oil price as a proxy for the trade costs ($t_{ij,t}$) in the Anderson and Wincoop model. Oil price is time varying and area is time invariant. As the oil price is the same for all states in a time period, the time varying changes in income are explained by oil price and time fixed effects. Similarly, the interaction of log of population and log of area is the instrument for the internal trade (Storeygard, 2013). The IV estimation in column 4 reflects positive and statistically significant associations between trade ratios. The hypothesis that the IV coefficient for internal trade is zero is rejected at 90 percent confidence level. The explanatory power of the regression is high (Uncentered R2 is 0.98) and Cragg-Donald (1993) weak identification test (F-statistic=25) rejects the null hypothesis of having weak instruments. This IV estimate, besides the ones from other alternatives, shows that OLS estimation is biased and understates the trade effects.

An alternative approach for addressing the endogeneity problem is to solve the simultaneous equations. In this approach, finding extra moments to solve the system of equation plays the key role. For example, in a supply-demand identification problem, I need to find some variables or shocks that shift the supply schedule, so the slope of the demand can be identified. Accordingly, the standard IV practice searches for something that moves the means. Instead of using common instruments that move the trade measures while being uncorrelated to income, I apply the identification through the heteroskedasticity (IH) methodology introduced by Rigobon et al (2004). In this method, data are split into two samples according to the heteroskedasticity of the residuals. Rigobon (2004a) shows that if the relative variance of the residuals shifts across the sub-samples, there would be then enough moments to identify the equations. IH methodology searches for something that shifts the variance instead of mean. Actually, the difference in variance of sub-samples provides enough information to identify the coefficients. Column 5 presents the estimations for the growth equation using IH methodology. The regression shows a statistically and economically significant relationship between trade and income.[1]

In this section I considered different alternatives to estimate the effect of trade on income. The estimations are not considerably different when moving from one instrument to alternative ones. This supports the argument that there is a significant positive association between trade and income and confirms the validity of the constructed instrument using geographic characteristics. The data have shortcomings in some ways. The time dimension of the data set is only four and there are not many time variations in trade in the 14 years from 1993 to 2007. When lagged or differences in variables are considered, the temporal dimension drops even more to three or two time cycles. Moreover, the method of GDP calculation for the U.S. was changed in 1997. The GDP numbers for that year are reported in both calculation methods. The average and standard deviation of figures, however, are different between the two methods. Because I utilize the old method for the year 1993 and the new system of calculation for the other years, I use time fixed effects in order to take into account these changes in data.

There might be a possibility that the population is itself endogenous in the long run. To make sure that there is no individual variable driving the results, I consider only area as a measure of economic size. I redo the construction of the instrument at the cross-section level for the year of 2007. This practice has no major impact on

[1] We use the Stata module called "ivreg2h" written by Christopher F Baum and Mark E Schaffer to compute the coefficients and standard deviations. It can be found at the address: `https://ideas.repec.org/c/boc/bocode/s457555.html`

results. In fact, similarity between the results from this experiment and the one from panel data, using oil or population measures, and the one using IH methodology, implies that my findings do not hinge on the method I use to address the endogeneity problem.

3.5 Conclusion Remarks

Whether and how much trade affects income are open but imperative questions for academics and policymakers. The main difficulty in estimating the association between trade and income comes from the endogeneity of trade measures. Because of reverse causality and omitted variables, trade is not an exogenous variable and the link between trade and income cannot identify the trade effects. Geography is a common instrument for trade which looms large in recent studies of trade-led economic growth. Feyrer (2009), Rodriguez and Rodrick (2000) and Frankel and Romer (1999) among others have attempted to use the geographic component of trade to explain the economic outcomes. However, most of these studies are concerned with an international level of analysis; cross-sectional and due to data unavailability, they do not estimate the effect of internal trade directly.

The trade-led income is an open but imperative question for academics and policymakers. The main difficulty in estimating the association between trade and income is due to endogeneity of trade measures. Because of reverse causality and omitted variables, trade is not an exogenous variable and the link between trade and income cannot identify the trade effects.

This paper is an attempt to investigate the link between trade and income at regional level. Assuming the geographic characteristics are not affected by policy or income, at least in the short term, I use these factors to construct an instrument for identifying the income effect of interregional and intraregional trade. Consistent with most studies at international level, I find that there is a significant relationship between the geographic components of trade and income. A one percentage point increase in interstate trade ratio increases the income per capita by at least two percent. In addition, my results show that after controlling for interstate trade, intrastate trade affects income at the same magnitude as the interstate trade.

My findings are consistent across multiple instruments and specifications. I consider using the lagged variable as instruments for endogenous variables. In another variation, I consider transport cost instead of distance to explain trade patterns. In fact, world oil price is a proxy for transport costs in the gravity equation in order to construct the trade instrument. Moreover, I estimate the trade and income relation in a system of simultaneous equations by identification through the heteroskedasticity method introduced by Rigobon (2003). In this methodology, identification is based on the heteroskedasticity of the structural shocks. In these alternatives, the point estimates of interstate and intrastate trade impact are positive and mostly significant. My findings are subject to some caveats. The time dimension of the data set is limited to four, and in some specifications two, periods. That may mean I cannot reach a precise estimate. Another concern is related to the population density of a region, which may change over time due to changes in income.

Figure 3.1: Overall Internal Trade Versus Two Size Measures

(A) Internal Trade Versus Area

(B) Internal Trade Versus Population

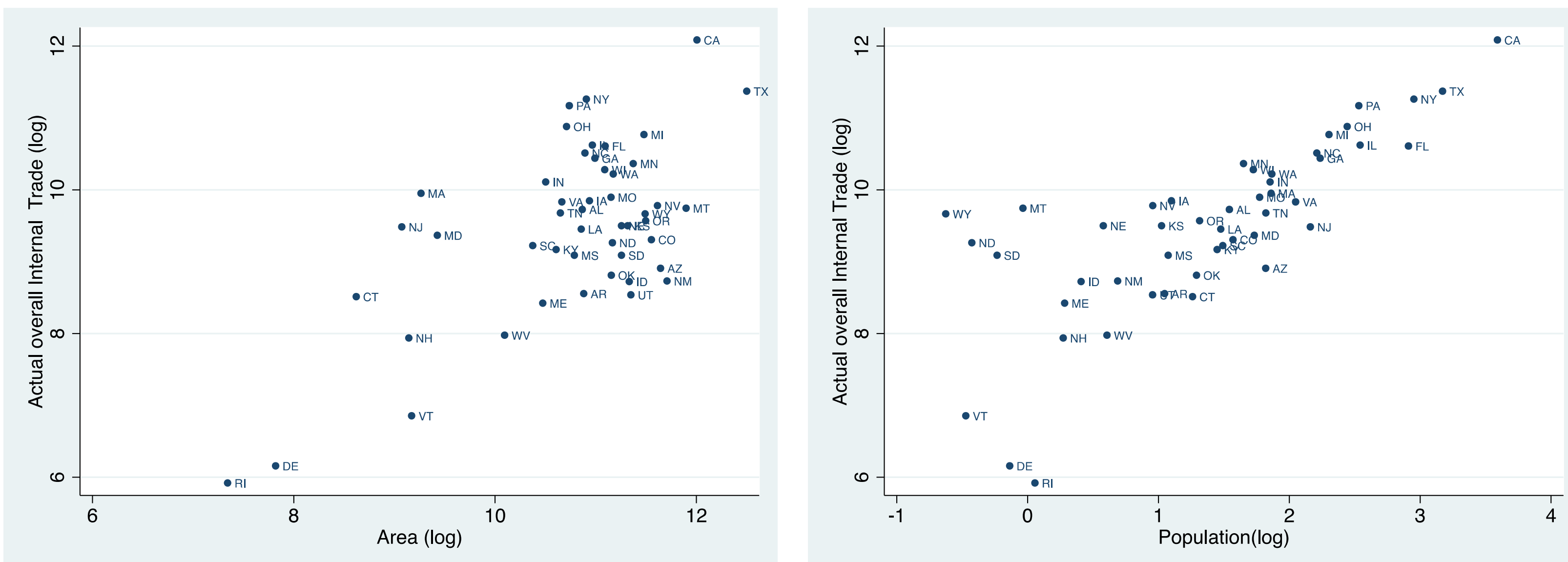


Figure 3.2: Overall Actual Trade Versus Overall Constructed Trade

Figure 3.3: Actual Versus Constructed Overall Trade, Controlled for Size Measures

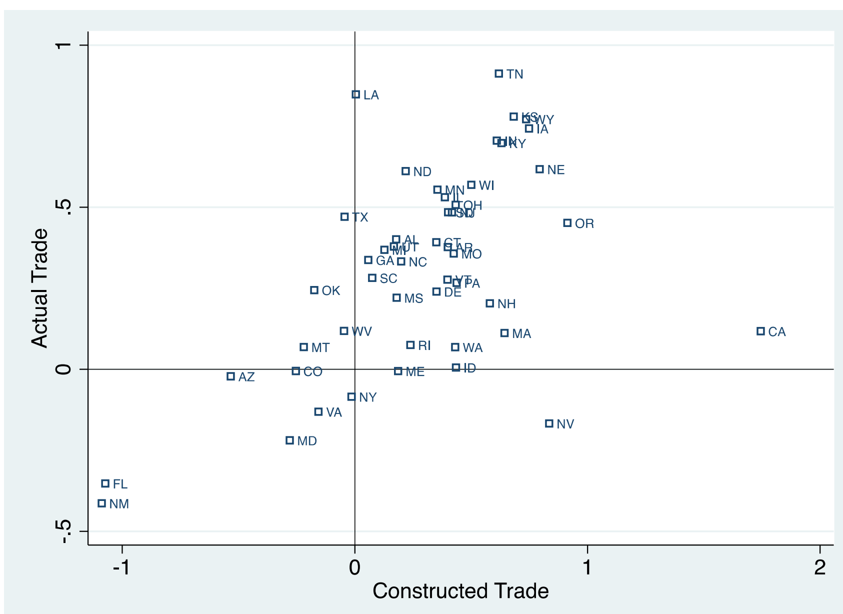


Table 3.1: Gravity Model Estimation

	(1)	(2)
Log distance	-0.842***	-0.975***
	(0.0245)	(0.0140)
Log product of area	0.0512***	-0.0825
	(0.0105)	(0.127)
Log population State i	0.643***	0.279*
	(0.0170)	(0.169)
Log population State j	1.011***	0.493***
	(0.0148)	(0.156)
Intra	2.176***	1.830***
	(0.112)	(0.0784)
Adjacent	0.927***	0.662***
	(0.0550)	(0.0299)
MRT State i		
	-0.670***	
	(0.0578)	
MRT State j	0.142***	
	(0.0442)	
Constant	-0.903***	3.605
	(0.265)	(2.524)
Adjusted R^2	0.643	0.92
Observation	8392	8392

Note: The dependent variable is log bilateral trade for 48 states within United States. The first column runs the regression using remoteness variables as a proxy for multi resistance term. Second column report the estimation using state's effect. Year effects are added in both columns. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.2: Actual and Constructed Overall Trade Relation

	Interstate trade			Internal trade
	(1)	(2)	(3)	(4)
Ln constructed trade	0.787***		0.330***	
	0.03		0.07	
Ln area		0.0285	0.0254	0.629***
		0.0209	0.0239	0.0432
Ln population		1.115***	0.675***	0.701***
		0.0242	0.0884	0.0451
Constant	2.514***	10.30***	6.547***	2.192***
	0.333	0.204	0.867	0.453
adj. R^2	0.899	0.929	0.947	0.762
observations	192	192	192	192

Note: Standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.3: Trade and Income

	(1)	(2)	(3)	(4)
Internal trade ratio	2.091 ***	2.914	4.686*	5.253**
	(.520)	(4.255)	(2.774)	(2.435)
Inters-state trade ratio	-0.161	6.868***	6.010***	6.280***
	0.238	(2.631)	(1.655)	(1.543)
ln area	.397***		.0468***	.0384***
	(.131)		0	(0.000)
ln population	-.517 **	.051***	.0369***	.0483***
	(.178)	0.009	0	(0.000)
constant	10.15***	-0.00454	0.467	-0.0972
	(0.701)	(4.829)	(3.100)	(2.818)
Adj. R2	0.99	0.9723	0.9772	0.9763
observations	182	178	178	178

Note: Standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.4: Sensitivity Analysis

	Average		Difference	Using oil price	IH
	(1)	(2)	(3)	(4)	(5)
internal trade	4.238*	2.295**	4.631**	3.910*	4.300*
	(2.266)	(1.114)	(1.248)	(2.446)	(2.577)
Interstate trade	4.771***	0.632	2.192**	3.110***	2.800***
	(1.355)	(0.695)	(0.884)	(1.120)	(1.104)
population	0.049***	1.015**			
	(0.008)	(0.165)			
log GDP (lag)		-0.409**			
		(0.148)			
inflation		0.709**			
		(0.186)			
log education		6.082*		0.89	2.000**
		(3.516)		(1.044)	(0.978)
constant	2.939	6.337**	5.635**	1.144	
	(2.534)	(2.428)	(0.212)	(4.503)	
R2	0.0951	0.573	0.238	0.98	0.98
observation	144	96	144	182	182

Note: Standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

3.6 References

Ades, Alberto F., and Edward L. Glaeser. 1994. "Evidence on Growth, Increasing Returns and the Extent of the Market." Working Paper 4714. National Bureau of Economic Research. http://www.nber.org/papers/w4714.

Alcalá, Francisco, and Antonio Ciccone. 2004. "Trade and Productivity." The Quarterly Journal of Economics 119 (2): 613–46. doi:10.1162/0033553041382139.

Anderson, James E., and Eric van Wincoop. 2001. "Gravity with Gravitas: A Solution to the Border Puzzle." Working Paper 8079. National Bureau of Economic Research. http://www.nber.org/papers/w8079.

Baier, Scott L., and Jeffrey H. Bergstrand. 2009. "Bonus Vetus OLS: A Simple Method for Approximating International Trade-Cost Effects Using the Gravity Equation." Journal of International Economics 77 (1): 77–85. doi:10.1016/j.jinteco.2008.10.004.

Barro, Robert J., and Xavier Sala-i-Martin. 1997. "Technological Diffusion, Convergence, and Growth." Journal of Economic Growth 2 (1): 1–26.

Bassanini, Andrea, Stefano Scarpetta, and Philip Hemmings. 2001. "Economic Growth: The Role of Policies and Institutions. Panel Data Evidence from OECD Countries." SSRN Scholarly Paper ID 265091. Rochester, NY: Social Science Research Network. http://papers.ssrn.com/abstract=265091.

Ben-David, Dan. 1993. "Equalizing Exchange: Trade Liberalization and Income Convergence." The Quarterly Journal of Economics 108 (3): 653–79. doi:10.2307/2118404.

Dawson, John W., Joseph P. DeJuan, John J. Seater, and E. Frank Stephenson. 2001. "Economic Information versus Quality Variation in Cross-Country Data." The Canadian Journal of Economics / Revue Canadienne d'Economique 34 (4): 988–1009.

Dollar, David. 1992. "Outward-Oriented Developing Economies Really Do Grow More Rapidly: Evidence from 95 LDCs, 1976-1985." Economic Development and Cultural Change 40 (3): 523–44.

Dollar, David, and Aart Kraay. 2001. Trade, Growth, and Poverty. World Bank, Development Research Group, Macroeconomics and Growth.

Donaldson, Dave. 2010. “Railroads of the Raj: Estimating the Impact of Transportation Infrastructure.” Working Paper 16487. National Bureau of Economic Research. http://www.nber.org/papers/w16487.

Edwards, Sebastian. 1998. “Openness, Productivity and Growth: What Do We Really Know?” The Economic Journal 108 (447): 383–98. doi:10.1111/1468-0297.00293.

Ehrmann, Michael, Marcel Fratzscher, and Roberto Rigobon. 2011. “Stocks, Bonds, Money Markets and Exchange Rates: Measuring International Financial Transmission.” Journal of Applied Econometrics 26 (6): 948–74. doi:10.1002/jae.1173.

Evenett, Simon J., and Wolfgang Keller. 1998. “On Theories Explaining the Success of the Gravity Equation.” Working Paper 6529. National Bureau of Economic Research. http://www.nber.org/papers/w6529.

Feenstra, Robert C. 2003. Advanced International Trade: Theory and Evidence. Princeton University Press.

Feyrer, James. 2009. “Trade and Income – Exploiting Time Series in Geography.” NBER Working Paper 14910. National Bureau of Economic Research, Inc. http://ideas.repec.org/p/nbr/nberwo/14910.html.

Frankel, Jeffrey A., and David Romer. 1999. “Does Trade Cause Growth?” The American Economic Review 89 (3): 379–99.

Frankel, Jeffrey, and Andrew Rose. 2002. “An Estimate of the Effect of Common Currencies on Trade and Income.” The Quarterly Journal of Economics 117 (2): 437–66.

Hillberry, Russell, and David Hummels. 2002. “Explaining Home Bias in Consumption: The Role of Intermediate Input Trade.” Working Paper 9020. National Bureau of Economic Research. http://www.nber.org/papers/w9020.

Lee, Ha Yan, Luca Antonio Ricci, and Roberto Rigobon. 2004. “Once Again, Is Openness Good for Growth?” Journal of Development Economics 75 (2): 451–72. doi:10.1016/j.jdeveco.2004.06.006.

Lewbel, Arthur. 2012. “Using Heteroscedasticity to Identify and Estimate Mismeasured and Endogenous Regressor Models.” Journal of Business & Economic Statistics 30 (1): 67–80. doi:10.1080/07350015.2012.643126.

Noguer, Marta, and Marc Siscart. 2005. "Trade Raises Income: A Precise and Robust Result." Journal of International Economics 65 (2): 447–60. doi:10.1016/j.jinteco.2003.11.006.

Rodriguez, Francisco, and Dani Rodrik. 2001. "Trade Policy and Economic Growth: A Skeptic's Guide to the Cross-National Evidence." In NBER Macroeconomics Annual 2000, Volume 15, 261–338. MIT PRess. http://www.nber.org/chapters/c11058.pdf.

Sachs, Jeffrey D., Andrew Warner, Anders \AAslund, and Stanley Fischer. 1995. "Economic Reform and the Process of Global Integration." Brookings Papers on Economic Activity 1995 (1): 1–118.

Storeygard, Adam. 2013. "Farther on Down the Road: Transport Costs, Trade and Urban Growth in Sub-Saharan Africa." SSRN Scholarly Paper ID 2265755. Rochester, NY: Social Science Research Network. http://papers.ssrn.com/abstract=2265755.

Topalova, Petia. 2005. "Trade Liberalization, Poverty, and Inequality: Evidence from Indian Districts." NBER Working Paper 11614. National Bureau of Economic Research, Inc. https://ideas.repec.org/p/nbr/nberwo/11614.html.

Wacziarg, Romain. 2001. "Measuring the Dynamic Gains from Trade." The World Bank Economic Review 15 (3): 393–429. doi:10.1093/wber/15.3.393.

Wei, Shang-Jin. 1996. "Intra-National versus International Trade: How Stubborn Are Nations in Global Integration?" Working Paper 5531. National Bureau of Economic Research. http://www.nber.org/papers/w5531.

Wolf, Holger C. 2000. "Intranational Home Bias in Trade." Review of Economics and Statistics 82 (4): 555–63. doi:10.1162/003465300559046.

Appendix

Table A1: Bootstrap Results for GDP Benchmark

light_i	light_j	distance	MTR_i	MTR_j
0.86	0.85	-0.51	-0.75	-0.75
0.8	0.99	-0.69	-0.75	-0.7
0.86	0.95	-0.63	-0.9	-0.65
0.7	0.93	-0.35	-0.8	-0.8
0.88	0.89	-0.69	-0.7	-0.65
0.7	0.99	-0.47	-0.75	-0.7
0.7	0.95	-0.35	-0.8	-0.8
0.72	0.91	-0.47	-0.65	-0.9
0.88	0.95	-0.69	-0.7	-0.7
0.82	0.87	-0.65	-0.75	-0.75
0.8	0.79	-0.37	-0.8	-0.8
0.8	0.83	-0.43	-0.65	-0.9
0.76	0.81	-0.35	-0.7	-0.65
0.88	0.99	-0.59	-0.8	-0.8
0.76	0.99	-0.51	-0.7	-0.75
0.7	0.99	-0.41	-0.65	-0.9
0.86	0.83	-0.53	-0.75	-0.8
0.76	0.99	-0.51	-0.7	-0.8
0.88	0.87	-0.57	-0.65	-0.65
0.82	0.95	-0.55	-0.9	-0.65
0.88	0.95	-0.69	-0.7	-0.7
0.84	0.85	-0.69	-0.7	-0.7
0.76	0.89	-0.47	-0.65	-0.8
0.74	0.99	-0.41	-0.8	-0.75
0.82	0.79	-0.35	-0.8	-0.9
0.8	0.91	-0.52	-0.74	-0.76

Table A2: Constructed Internal Trade and Distance: 2007

	Internal trade		Internal distance				
Country	Constructed Internal Trade	GDP-Export	$\beta_3 = -0.51$	$\beta_3 = -1$	$\beta_3 = 1$	Major states	$\beta_3 = -1$ pop
Afghanistan	1.35E+009	3.39E+010	478.6	368.7	430.3	303.8	90.3
Albania	1.06E+010	2.30E+010	94.5	78.1	78.7	63.8	51.5
Algeria	1.32E+011	3.78E+011	454.6	366.9	560.8	580.5	129
Angola	3.27E+009	8.69E+010	787.9	576.7	520.8	420	67.5
Armenia	1.09E+010	1.96E+010	85.5	70.9	70.3	64.9	41.2
Australia	1.23E+011	7.38E+011	2730.9	1988.2	1948.7	1042.8	153.6
Austria	3.46E+011	2.25E+011	234.7	186.5	167.7	108.9	62.8
Azerbaijan	3.77E+010	1.07E+011	172.5	139	166	110.7	37.5
Bahrain	2.36E+010	3.86E+010	36.1	25.1	16	9.8	9.7
Bangladesh	3.86E+010	3.05E+011	360.7	276.1	185	142.7	62.4
Barbados	4.28E+009	4.14E+009	15.5	13	11.6	7.8	22.8
Belarus	1.71E+011	1.06E+011	247.3	166.9	188.3	171.4	113.9
Belize	1.31E+009	1.94E+009	140	109.6	77.6	57	57.5
Benin	2.90E+009	1.37E+010	122.7	89.1	116	126.2	80.8
Bolivia	5.93E+009	4.38E+010	1081.8	834.1	614.7	394.2	31.8
Bosnia and Herzegovina	2.03E+010	3.14E+010	126.1	104.1	101.6	85.1	68.5
Brazil	8.21E+011	2.30E+012	1663.1	1377.6	1574.1	1097.4	144.8
Brunei	4.51E+009	2.22E+010	257	120.6	28.6		
Bulgaria	6.36E+010	9.49E+010	162.4	141.7	177.2	125.3	97.2
Burkina Faso	1.13E+009	1.85E+010	435	284.9	219.7	197	74.7
Burundi	1.07E+008	5.89E+009	248.6	84.7	22	62.8	51.5
Cambodia	1.23E+009	2.84E+010	234.2	125.8	148.8	160	56.6
Cameroon	1.51E+009	4.44E+010	590.3	431	337.7	259.3	134.2
Canada	2.45E+012	9.74E+011	2111.1	1621	1563.8	1188	165.7
Cape Verde	1.21E+008	2.70E+009	117.9	71.9	166.3	23.9	61.6
Central African Republic	5.00E+007	3.14E+009	513.6	308.6	235	296.9	72.7
Chad	2.00E+008	1.65E+010	715.9	494.6	371.5	426.2	96.7
Chile	5.00E+010	2.50E+011	661.3	498	735.3	327.2	67.7
China	3.62E+012	8.39E+012	1142.9	969.9	1302	1163.7	305.7
Colombia	7.06E+010	4.33E+011	419	362.1	433.4	401.9	109.6
Comoros	6.00E+005	8.68E+008	2436.6	640	45.5	16.2	51.5
Costa Rica	2.17E+010	4.05E+010	171.5	136.3	121.2	85	14.4
Croatia	1.06E+011	8.56E+010	139.7	100.4	168.3	89.4	65.6
Cuba	7.47E+009	1.88E+011	298.5	227.8	375.5	125.2	111.9
Cyprus	4.42E+010	2.38E+010	82.9	61.9	42.9	36.2	23.6

Continued on next page

Table A2 – Continued from previous page

Country	Internal trade		Internal distance				
	Constructed Internal Trade	GDP-Export	$\beta_3 = -0.51$	$\beta_3 = -1$	$\beta_3 = 1$	Major states	$\beta_3 = -1$ pop
Czech Republic	3.44E+011	2.00E+011	155.3	126.2	150.7	105.6	81.2
Denmark	2.72E+011	1.57E+011	238.5	181.7	128.6	78.1	61.1
Djibouti	1.74E+008	1.59E+009	219.3	109.3	31.8	57.3	34.2
Dominica	1.06E+008	5.96E+008	20.3	16.5	15.9	10.3	15.2
Dominican Republic	1.72E+010	9.27E+010	118.8	97.3	123	83	53.9
Ecuador	4.96E+010	1.14E+011	285.1	241.1	275.2	197.5	79.2
Egypt	5.87E+011	6.85E+011	271.8	222.2	342.9	376.4	32.8
El Salvador	2.03E+010	4.18E+010	89.5	73.3	85.3	54.6	11.7
Equatorial Guinea	1.15E+008	1.69E+010	523.6	312	181.1	63	69.3
Estonia	7.34E+010	2.58E+010	141.1	114.8	99.2	80	62.5
Ethiopia	4.02E+009	6.85E+010	562.7	416.9	347.8	399.3	75.3
Fiji	7.14E+008	5.30E+009	691.9	304.9	275.2	50.9	28.7
Finland	7.02E+011	1.45E+011	301.8	256.8	316.8	218.2	66.9
France	2.96E+012	1.93E+012	339.7	299	418	278.2	161.4
Gabon	3.58E+009	2.08E+010	721.3	439.7	202	194.6	62.3
Gambia	2.50E+007	2.30E+009	1308.9	420.6	65	40	17.8
Georgia	8.59E+009	2.43E+010	158.3	125	142.6	99.3	62.7
Ghana	9.34E+009	5.86E+010	321.2	252.6	204.1	183.7	58
Greece	1.81E+011	3.42E+011	308.1	260.9	265.7	136.5	13.2
Grenada	5.06E+008	1.21E+009	25.4	17.6	13.1	7	13.4
Guatemala	3.10E+010	8.54E+010	138.2	116.6	134.4	124.1	19.5
Guinea	2.07E+008	1.09E+010	322	208	301.8	186.5	63.2
Guinea Bissau	2.41E+006	1.82E+009	1023	280	37.3		
Guyana	8.14E+008	3.30E+009	175.4	123.4	121.1	174.4	39.2
Haiti	2.40E+008	1.47E+010	343	167.8	51.8	62.7	17.4
Honduras	1.86E+010	2.59E+010	189	155.9	148.5	125.9	60.5
Hong Kong S.A.R.	2.45E+010	8.35E+009	22.4	19.1	19.1		
Hungary	1.58E+011	1.58E+011	138.4	115.5	157.5	114.7	73.4
Iceland	1.91E+010	9.82E+009	222.5	168.9	171	120.6	10
India	1.68E+012	4.29E+012	1088.1	939.5	1039.5	682	235
Indonesia	1.21E+011	1.53E+012	1102.7	852.9	1177.8	523	120.8
Iran	8.53E+011	9.99E+011	624.9	542.6	666.3	482.9	133.3
Iraq	1.42E+011	2.98E+011	429.2	337.7	335.9	248.7	110.3
Ireland	1.97E+011	1.12E+011	130	114.2	140.1	99.7	49.6

Continued on next page

Table A2 – Continued from previous page

Country	Internal trade: Constructed Internal Trade	Internal trade: GDP-Export	Internal distance: $\beta_3 = -0.51$	Internal distance: $\beta_3 = -1$	Internal distance: $\beta_3 = 1$	Internal distance: Major states	Internal distance: $\beta_3 = -1$ pop
Israel	2.48E+011	1.57E+011	199.8	126.2	109.7	55.3	25.1
Italy	2.62E+012	1.82E+012	302.6	251.8	432	206.5	168.1
Ivory Coast	7.89E+009	4.32E+010	288.7	243.4	244.7		
Jamaica	1.44E+010	2.21E+010	76	62.5	71.4	39.4	21.8
Japan	1.48E+012	3.90E+012	424.6	343.9	585.4	231.2	83.1
Jordan	5.40E+010	5.41E+010	157	126.8	136.3	114	22.1
Kazakhstan	8.74E+010	2.58E+011					
Kenya	3.73E+009	8.73E+010	403.8	287.6	275.4	287.1	77.2
Kuwait	6.00E+011	2.04E+011	220.9	128.6	49.1	50.2	8.9
Kyrgyzstan	1.40E+010	1.29E+010	325.1	226.3	222.8	167.6	70.6
Laos	6.95E+008	1.90E+010	451.5	293.2	260.8		
Latvia	3.68E+010	4.26E+010	80	62.3	111.8	95.6	55
Lebanon	9.88E+010	5.18E+010	99.5	72.2	53.9	38.5	33.1
Liberia	1.95E+008	4.05E+008	187.6	119.7	96.6	125.5	50.4
Libya	5.75E+010	1.25E+011	543.2	419	587.6		
Lithuania	6.03E+010	5.76E+010	167.7	138.9	117.8	96.1	82.9
Luxembourg	5.52E+010	2.36E+010	98.8	53.5	16.1	19.1	12.2
Macedonia	1.34E+010	1.93E+010	66.5	56.3	83.3		
Madagascar	1.43E+008	2.74E+010	1023.2	570.9	258.1	288.2	104.7
Malawi	1.30E+009	8.19E+009	377.9	265.2	220.4	129.5	86.4
Malaysia	1.57E+011	3.65E+011	382.3	288.3	570.8	216	65.4
Maldives	2.61E+005	2.84E+009	-0.6	-	60.9	6.5	74.4
Mali	1.61E+009	1.97E+010	471.5	341.9	367	418.9	84.6
Malta	1.01E+010	8.34E+009	9.7	8.3	12.8	6.7	4.7
Mauritania	3.24E+008	7.67E+009	852	495.5	349.5	381.9	76.9
Mauritius	4.66E+009	1.51E+010	32.2	25.5	72.1	17	11.8
Mexico	5.57E+011	1.55E+012	684.2	566.1	876.7	527.5	77.5
Moldova	9.57E+009	1.16E+010	84.4	71.7	101.9		
Mongolia	4.37E+009	1.39E+010	336.7	193.9	303.2	470.6	76.5
Morocco	4.65E+010	1.69E+011	335.7	285.6	339.3	317.1	112.3
Mozambique	1.73E+009	1.63E+010	633.4	357.6	673	336.8	93.3
Nepal	2.15E+009	4.54E+010	247.9	194.8	221.3	144.3	38.3
Nicaragua	9.17E+009	2.07E+010	152.7	113.3	105.3	136	50.8
Niger	1.98E+008	1.09E+010	823.1	536.5	589.5	423.4	121.5
Nigeria	1.12E+011	6.13E+011	348.6	276	388.7	361.5	140
Norway	3.25E+011	1.84E+011	431.7	341.9	560.7	214	93.7
Oman	5.86E+010	9.23E+010	329.9	271.5	301.3	173.3	63.9
Pakistan	1.16E+012	6.71E+011	1767.3	977.9	338.6	353.4	127.7

Continued on next page

Table A2 – Continued from previous page

	Internal trade		Internal distance				
Country	Constructed Internal Trade	GDP-Export	$\beta_3 = -0.51$	$\beta_3 = -1$	$\beta_3 = 1$	Major states	$\beta_3 = -1$ pop
Panama	5.70E+009	4.47E+010	479.2	305.3	179.2	103.5	27.5
Papua New Guinea	8.50E+008	8.40E+009	675.1	475	482.9	255.9	103.8
Paraguay	1.03E+010	3.75E+010	321.3	237.6	301.7	239.9	20
Peru	1.42E+010	2.20E+011	661.4	511.8	760.7	426.4	84.7
Philippines	3.19E+010	4.18E+011	234.6	162.3	452.8	206	0.2
Poland	1.28E+012	6.23E+011	290.5	257.2	263.7	210.3	124.7
Portugal	3.68E+011	2.51E+011	170.3	144.8	183.4	114.3	23.5
Qatar	1.40E+011	1.20E+011	116.5	78.6	41.7	40.2	21.1
Republic of Serbia	7.85E+010	8.70E+010	129.9	111.2	113.5		
Republic of the Congo	1.81E+009	1.13E+010	1472	533.4	212.4		
Russia	4.78E+012	2.76E+012	1402	1167	1737.9		
Rwanda	3.22E+007	1.07E+010	961.9	316.4	36.2	61	48.8
Saint Kitts and Nevis	3.71E+008	1.05E+009	14.3	11.4	14.4	6.2	11.5
Saint Lucia	1.13E+009	1.52E+009	18.3	15.1	14.5	9.3	22.5
Saint Vin cent and the Grenadines	3.70E+008	9.36E+008	18.8	13.7	12.3	7.4	16.6
Samoa	1.83E+007	9.35E+008	185.6	87.4	41.1	20.1	33.6
Sao Tome and Principe	6.58E+005	4.04E+008	-	-	-	11.7	14.9
Saudi Arabia	6.59E+011	8.82E+011	853.5	703.9	634	551.9	142.9
Senegal	5.95E+009	2.44E+010	218.5	164	156.9	166.8	63.4
Seychelles	1.74E+008	1.43E+009	13.3	10	93.4	8	14.9
Sierra Leone	2.49E+007	6.57E+009	535.9	328.3	134	101.9	55.2
Singapore	2.56E+010	5.25E+010	28.7	20.4	13.6	9.6	9.6
Slovakia	9.41E+010	8.41E+010	154	123.1	130.7	83.3	85.7
Slovenia	6.56E+010	3.81E+010	62.3	53.7	81.7	53.5	45.1
Solomon Islands	1.08E+007	6.42E+008	261	111.4	127.3	64.9	47.3
South Africa	2.20E+011	5.35E+011	919.1	738.3	778.4	415.4	107.6
South Korea	1.25E+012	1.08E+012	195.1	151	163.8		
Spain	1.19E+012	1.37E+012	398.4	346	508.6	267.5	125.4
Sri Lanka	3.08E+010	1.25E+011	129.3	113	131.8	96.3	25.3
Sudan	1.86E+010	1.32E+011	597.2	452.7	450.6	595.4	27.4

Continued on next page

Table A2 – Continued from previous page

Country	Internal trade: Constructed Internal Trade	Internal trade: GDP-Export	Internal distance: $\beta_3 = -0.51$	Internal distance: $\beta_3 = -1$	Internal distance: $\beta_3 = 1$	Internal distance: Major states	Internal distance: $\beta_3 = -1$ pop
Suriname	3.35E+009	5.48E+009	79.6	54.9	77	152.2	31.2
Sweden	6.36E+011	2.56E+011	396.3	333.2	442.2	252.3	99
Switzerland	2.50E+011	2.65E+011	107.4	93	106.1	76.4	53.8
Tajikistan	3.21E+010	1.21E+010	309.3	216.3	114.8	142.3	62.4
Thailand	2.29E+011	6.63E+011	304.6	242.3	451.2	269.4	42
The Bahamas	2.03E+009	6.61E+009	317.8	184.7	175.6		
Togo	1.21E+008	6.59E+009	1540.8	487.8	72	89.6	68.1
Tonga	4.51E+006	4.62E+008	937.7	380.7	102.7	10.3	8
Trinidad and Tobago	3.05E+010	2.62E+010	52.7	43.9	48.4	26.9	21.3
Tunisia	6.96E+010	8.67E+010	177	150.5	197.7	152.4	42.4
Turkey	3.13E+011	1.07E+012	440.3	383	553.2	332.3	133.8
Turkmenistan	3.86E+010	3.19E+010	617.3	473	346.2	262.8	100
Uganda	4.60E+009	4.14E+010	133.4	82	101	184.7	67
Ukraine	2.58E+011	3.58E+011	398.4	331	442.3	292.2	182.8
United Arab Emirates	3.40E+011	3.94E+011	428.4	228.4	135.1	108.8	38
United Kingdom	2.99E+012	1.93E+012	152	132	191.9	185.8	86.2
United Republic of Tanzania	8.11E+008	7.71E+010	786	590.3	578.3		
United States of America	1.60E+013	1.44E+013	1312	1130.1	1642.4	1161.1	261.7
Uruguay	1.33E+010	4.27E+010	294.2	231.1	268.4	157.9	53.1
Uzbekistan	1.18E+011	8.86E+010	374.7	302.2	368.6	251.6	108.8
Vanuatu	4.09E+007	3.83E+008	576.6	365.1	182.1	45.7	57
Venezuela	1.07E+011	4.36E+011	492.9	407.8	542.1	359.2	105.4
Vietnam	4.14E+010	2.92E+011	354.8	260.7	630.4		
Zambia	6.54E+009	3.14E+010	700	493.7	336.7	326.3	95

VITA

ROOHOLAH HADADI

2005	B.Sc., Insustrial Engineering Isfahan University of Technology Isfahan, Iran
2008	M.S., Soci-economic System Engineering Sharif University Tehran, Iran
2012	M.A., Economics Florida International University Miami, Florida
2012–2016	Teaching Assistant and Doctoral Student Florida International University Miami, Florida