An Examination of Gravity Relationships at the Subnational Level

Paul Phillips

ECONOMICS WORKING PAPER SERIES
Working Paper 2024–05–A

U.S. INTERNATIONAL TRADE COMMISSION
500 E Street SW
Washington, DC 20436

May 2024

An Examination of Gravity Relationships at the Subnational Level
Paul Phillips
Economics Working Paper 2024–05–A
April 2024

Abstract

This paper uses data on shipments between U.S. regions to estimate a subnational version of the canonical international gravity equation. Using a CES utility framework, I estimate gravity regressions using a data set comprised of 132 subnational regions and then extract trade cost values for each origin–destination pair, breaking down results by sector and mode of transportation. All forty–two sectors show a statistically significant negative relationship between distance and trade, with the elasticity of trade with respect to distance being higher for raw materials than for manufactured goods. Among modes of transportation, distance more negatively affects pipeline shipments, and the effect of distance on air shipments is not statistically significant. With other factors being equal, shipping goods between regions costs around four times as much on median as shipping goods within a region, although this relationship varies widely depending on the mode of transportation or the sector.

Paul Phillips
Research Division, Office of Economics
U.S. International Trade Commission
Paul.Phillips@usitc.gov

1 Introduction

Since Tinbergen (1962) debuted the first gravity model, gravity estimation has been a prevalent field within international trade economics. A typical gravity estimation procedure involves the regression of bilateral trade between two countries on the distance between those two countries as well as shared characteristics including a shared border, the presence of a common language, the presence of a common currency, or colonial ties. Researchers use gravity estimations to inform their understanding of the factors influencing trade between two nations, or to calibrate the iceberg transportation costs that they can then input into a structural model.

However, economists have not yet, to this author's knowledge, applied gravity regression estimation to analyze the factors predicting trade between granular regions *within* the United States. The majority of papers analyzing subnational trade focus on the distributional effects of international trade policies throughout the country, and they do so either by assuming that import exposure rates are the same across the country or by using data at the level of states or aggregated Census areas,[1] where data is more reliable and more readily available than at the level of counties or metropolitan areas. Another strand of literature has used subnational trade statistics and GIS (Global Information System) data on transportation networks to map out trade costs within the United States. In this paper, I perform the more basic task of examining how the characteristics of a pair of U.S. regions influence the amount of trade conducted between these two regions. The results of this analysis provide insight into how gravity variables affect trade at the subnational level, as well as providing further guidance on how to calibrate trade costs within the United States in a structural model.

I obtain subnational shipment data from the Freight Analysis Framework (FAF), which reports the value and weight of shipments between 132 subnational regions. I use a simplified gravity estimation method that allows me to divide data into forty–two sectors and five modes of transportation. Explanatory variables within the gravity equation include distance and binary indicators of whether the two regions are in the same state or broader Census area, as well as other geographic and demographic factors that the regions may share in common. I then apply these gravity equation results to a standard CES model to obtain

[1] The Census Bureau uses the term 'regions' when delineating subdivisions of the United States. The Freight Analysis Framework, my main data source for this paper, also describes its U.S. subdivisions as regions. To avoid confusion, I will use the term 'areas' when referring to Census subdivisions.

estimates of trade costs[2] between different regions of the country.

Gravity equation results reveal that trade elasticities with respect to distance tend to be lower for manufactured goods than they are for raw materials, while distance elasticities are lower for goods transported across rail and air than for goods transported by truck, pipeline, or water. Gravity equation results do not differ significantly between 2018 and the following years, suggesting that the pandemic and trade wars with China did not have a strong effect on the gravity mechanisms behind subnational trade. The median trade cost between two subnational regions is around five, meaning that shipping goods between two regions of the country costs around 400% more than shipping goods within the same region. Median trade costs in the agricultural sector are higher than median trade costs in manufacturing sectors and median trade costs in air, water, and trade costs in rail transportation are lower than those in truck and pipeline transportation.

Section 2 reviews some of the literature on economic geography and subnational trade. Section 3 goes through the data, while Section 4 introduces the gravity equation methodology. Section 5 presents the gravity equation results, and Section 6 presents estimates of transportation costs. Section 7 concludes.

2 Literature Review

The typical gravity estimation procedure entails regressing trade between two countries on distance and a bevy of bilateral country characteristics, including shared border, common language, common currency, colonial ties, or the presence of a regional trade agreement between the two countries. Fixed effects control for factors specific to one country, and gravity regressions either employ a fixed effects estimation or Pseudo–Poisson Maximum Likelihood (PPML) framework. The U.S. International Trade Commission provides a database of country–level gravity variables across time (Gurevich and Herman, 2018). Recent papers in the academic literature that perform gravity estimation include Anderson, Larch, and Yotov (2022); Yotov, Larch, and Heid (2015); Agnosteva, Anderson, and Yotov (2014); Olivero and Yotov (2012); Lin and Sim (2012); Helpman, Melitz and Rubinstein (2008). To the author's knowledge, the gravity literature has not yet considered estimating the relationship between gravity factors and trade at the subnational level.

Papers that use subnational trade data will often do so in order to estimate the regional effects of a national trade policy. These papers often assume that import exposure is the same regardless of location within the country. Reduced–form papers such as Hakobyan and McLaren (2016); Acemoglu, Autor, Dorn,

[2]Throughout this paper, I use the terms 'trade costs' and 'transportation costs' synonymously.

Hanson, and Price (2016); and Autor, Dorn and Hanson (2013) determine local labor market exposure to trade competition using localized employment data and national import data. The model–based analyses in Caliendo and Parro (2021) and Caliendo, Dvorkin and Parro (2019) contain trade costs at the local level, but calculate subnational import shares based on national trade data from the World Input–Output Database and regional variation in employment. Other papers do not make such a strong assumption about import exposure, but due to data limitations they perform their analysis at the state or Census region level rather than any more disaggregated subnational areas. Riker (2022) and Feenstra and Hong (2023, working paper) use subnational trade data to determine how changes in trade policy will affect different states, while Riker (2019) performs a similar analysis for U.S. census regions. Riker (2022) uses subnational shipment data from the Commodity Flow Survey, while Feenstra and Hong use the FAF. Analysis done at the state or regional level can help increase data reliability but nonetheless obscures potentially large regional differences within states. The gravity analysis discussed here employs data for 132 sub–state regions, allowing for insights into how these more granular factors can influence trade.

Another strand of the economics literature uses granular subnational data to estimate a network of trade costs across the country. Recent papers do so in order to model the welfare effects of a specific transportation policy. These papers include Fuchs and Wong (2023) for port congestion, Allen and Arkolakis (2022) for road congestion, Jaworski, Kitchens, and Nigai (2023) for the Interstate Highway System, and Donaldson and Hornbeck (2016) for the historical development of U.S. railroads. Allen and Arkolakis (2014) use a general equilibrium model to estimate the distribution of trade costs, productivities and amenities in the United States, while also calculating the welfare effects of constructing the Interstate Highway System. Meanwhile, Agnosteva, Anderson and Yotov (2014) and Anderson and van Wincoop (2003) use subnational data to estimate different transportation costs, such as the transportation cost imposed by distance and the trade cost associated with crossing a state or international border. Agnosteva et al. focus on Canadian provinces, while Anderson and van Wincoop's model uses data from Canadian provinces and U.S. states. This literature often focuses on one mode of transportation, while not distinguishing between sectors. In comparison, my use of a more simplified gravity framework allows me to examine forty–two disaggregated sectors and five modes of transportation simultaneously, allowing for insights into how gravity results may differ at the sectoral level.

By using a CES–driven gravity framework with detailed subnational trade data, this paper provides clear insight into how geographic and demographic variables relate with trade at the subnational level. The set–up of the paper also allows for insights into how gravity results may differ between sectors or between different modes of transportation.

3 Data

I use data from the Freight Analysis Framework, which provides figures on inflows and outflows between 132 subnational regions. In this paper, I focus on how geographic or demographic characteristics affect a shipment that passes between two subnational regions, regardless of whether this shipment was necessarily produced in the first region or consumed in the second. Although the FAF breaks subnational trade flows down into shipments traded exclusively within the U.S. and domestic components of international shipments, I use information on shipments sent from one region within the U.S. to another, aggregated across all initial origins or ultimate destinations.[3]

The regions represent both metropolitan areas and non–metropolitan areas, although not all U.S. metropolitan areas are represented in the dataset. The FAF either represents each state in its entirety, as is the case with less–populated states such as Alaska and Idaho, or divides states into one or more metropolitan areas and the 'rest of the state', which includes all counties not represented by one of the aforementioned metropolitan areas. For example, the FAF divides Georgia into Atlanta, Savannah, and 'Rest of Georgia.'

The FAF data includes forty–two SCTG2 sectors and eight modes of transportation. The eight modes of transportation are truck, rail, water, air (including routes with both truck and air components), multiple modes and mail, pipeline, other and unknown, and no domestic mode. The 'multiple modes and mail' category includes all trade involving intermodal switching other than truck–air, but does not include more detailed information on the different modes involved or when the switching occurs. 'Other and unknown' and 'no domestic mode' are limited in their number of observations, and any results in these areas would be difficult to interpret. I therefore focus on the five major transport modes: truck, rail, water, air, and pipeline.

[3] The FAF does not distinguish between shipments that terminate at their recorded destination and transshipments whose recorded destination is a temporary warehousing. While this would complicate the task of estimating subnational import exposure rates, it has no bearing on the gravity analysis contained within this paper.

3.1 Descriptive Statistics

Subnational trade flows encompass 132 regions, forty–two sectors, and five modes of transportation, for a total of around six million observations. Any kind of comprehensive analysis would be impossible, but this section highlights some of the general trends observed in U.S. subnational trade in 2018.

As shown in Figure 1 on the following page, the overwhelming majority of goods within the United States are shipped by truck, with truck travel accounting for almost three–quarters of subnational trade by value. Trade over water, rail, and air routes all account for less than five percent of total trade, with goods shipped by pipeline covering about six percent of subnational trade values.

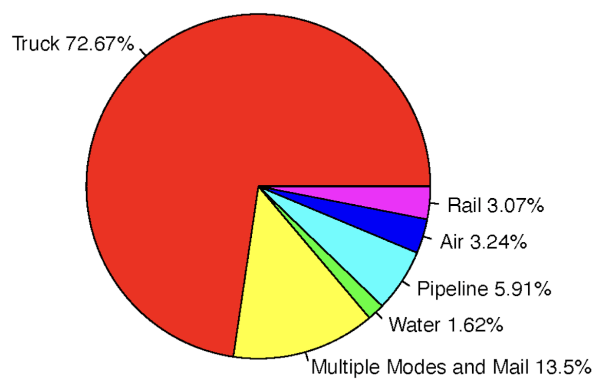


Figure 1: Breakdown of subnational trade by type of transportation, NAICS

Statistic	Truck	Rail	Water	Air	Pipeline
Proportion of agricultural trade	.898	.0389	.0160	2.36×10^{-3}	0
Proportion of trade in energy	.464	.0352	.0690	1.14×10^{-5}	.417
Proportion of trade in motor vehicles	.764	.0862	.00464	.00599	0

Table 1: Breakdown of subnational trade by mode of transportation for selected sectors

Table 1 shows that the breakdown of subnational shipments by type of transportation varies by sector. For agricultural goods, which include FAF sectors 1–9, almost ninety percent of all shipments are by truck, while rail and air shipments are nearly nonexistent. Energy goods, which include FAF sectors 15–19, are transported by pipeline almost as often as by truck, and the fraction of goods shipped by water is also higher for energy goods than it is overall. Around three–quarters of the value of motor vehicles trade is by truck, with rail shipments comprising 8.6 percent of motor vehicles trade compared with three percent of trade generally.

	Truck	Rail	Water	Air	Pipeline
Pairs with zero trade	.468%	11.9%	36.5%	9.53%	11.2%
Largest origin (volume)	Los Angeles	Detroit	New Orleans	Los Angeles	Rest of TX
Largest destination (volume)	New York, NY	Detroit	New Orleans	Los Angeles	Houston, TX
Largest origin (fraction)	Rest of MD	ND	AK	AK	ND
Largest destination (fraction)	Fresno, CA	Portland, OR	New Orleans	AK	Beaumont, TX

Table 2: Breakdown of subnational trade by mode of transportation for regions

The vast majority of U.S. regions send shipments to each other. Only 144 regional pairs out of the 17424 total pairs do not trade with each other in any capacity. Of the regional pairs that trade with each other, fewer than one percent do not conduct trade using trucks. However, the other modes of transportation are more specialized. 11.9% of regional pairs do not send goods to each other by railroad, 36.5% do not trade over water, 9.53% do not trade via airplane and 11.2% do not send goods via pipeline.

Due to its large population and proximity to China and Mexico, Los Angeles, CA is the largest origin region by volume for trade transported by truck and air. Los Angeles is also the highest destination region for air trade, but New York City, NY is the largest destination region for truck due to its large customer base. Detroit, MI conducts the most trade by rail both as an origin and as a destination, with New Orleans, LA conducting the most trade by water. As Texas is a center of oil production, the highest volume of pipeline trade both by origin and destination is recorded in Texas. When regions are compared by the *fraction* of trade conducted by each mode of transportation, statistics change somewhat. North Dakota records the highest fraction of outbound rail and pipeline trade in the country, with Fresno, CA receiving a high proportion of inbound truck trade. Alaska has a high proportion of air trade both incoming and outgoing, as its position renders rail or truck trade to be difficult.

In Figure 2, I display the amount of subnational trade that passes through each FAF subnational region, either as an origin or as a destination. Los Angeles, CA is the region through which the most subnational shipments pass, comprising 2.553 trillion dollars and 6.34% of the value of all subnational shipments. Chicago, Detroit, Dallas–Fort Worth, Houston, and Rest of Texas also contribute more than a trillion dollars of subnational trade.

Large amounts of subnational shipments pass through the South and Midwest, with regions within Wis-

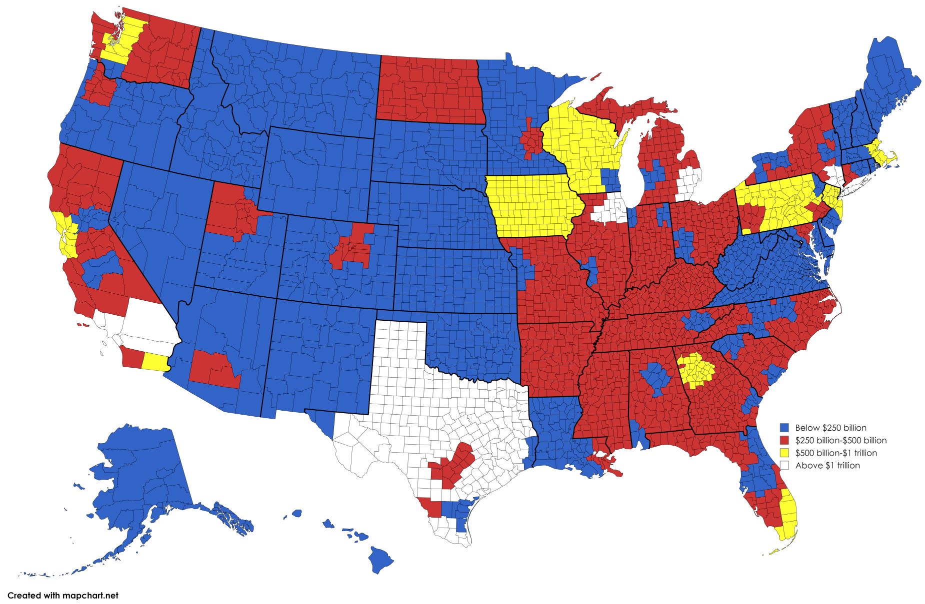
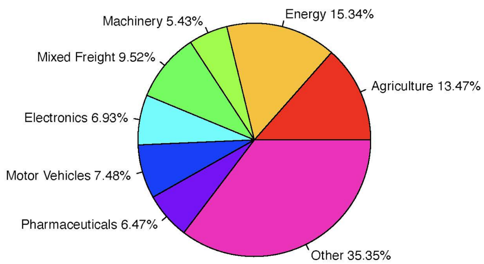


Figure 2: Geographical distribution of subnational shipments

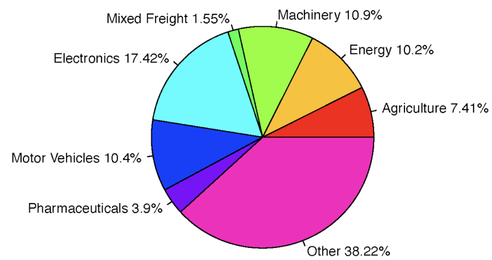
consin, Iowa, Illinois, and Texas all for the most part contributing more than 500 billion dollars of subnational trade. Fewer shipments pass through the Western and Northeastern states, with the exceptions of major metropolitan areas in the Northeast (Boston, New York City, Philadelphia), and California. Results are inconsistent as to whether more trade passes through major metropolitan areas within a state, or the 'rest of the state' as defined by the FAF which includes all counties not in that metropolitan area. For example, 'Rest of Ohio' and 'Rest of Pennsylvania' both process more than $500 billion of subnational shipments, but the metropolitan areas in those states do not. In Washington, Minnesota, and the Mountain West, higher shipment values go through the major metropolitan areas than the counties outside these areas. Florida and New York have one metropolitan area that processes more shipments than non–metropolitan areas within the state, while other metropolitan areas within the state process fewer shipments.

When comparing the geographical distribution of shipments among sectors, the identities of the regions processing the highest amount of shipments change substantially. For example, Detroit, MI processes 11.7% of shipments in motor vehicles but 2.52% of all subnational shipments, while Iowa processes 3.03% of all shipments in agricultural goods but 1.27% of all subnational shipments. However, regardless of the sector, a median .5% of all subnational trade passes through each individual U.S. regions, with the most trade–

intensive region contributing between five and ten percent.



(a) Domestic Destination



(b) International Origin/Destination

Figure 3: Subnational Trade Breakdown by Industry Category

Figure 3 displays the percentage of subnational trade belonging to major industry categories; I have further broken down the trade into shipments that a) start and finish within U.S. borders or b) either originate abroad or are eventually bound for a foreign destination. Energy goods, agricultural goods and pharmaceuticals are more commonly shipped domestically for domestic consumption, while subnational electronics shipments more often have a foreign origin or destination. Mixed freight accounts for nearly ten percent of purely domestic shipments, but a much smaller fraction of shipments that have an international component.

Individual FAF sectors[4] that account for a large fraction of subnational trade include motor vehicles, electronics, machinery, and pharmaceuticals.

I obtain an estimate of how decentralized trade is within each sector by computing a Herfindahl–Hirschman Index (HHI) using each FAF region's share of trade within that sector. HHIs below 1000 indicate a highly competitive industry, and as such trade in Figure 4 is generally quite decentralized, with most sectors having HHIs below 500. Tobacco, coal, and crude petroleum are the sole sectors with HHIs above 500, while tobacco is the sole sector to have an HHI above 1000, indicating moderate concentration. 25.7% of tobacco shipments are sent from or received in Greensboro, NC, and 18.8% of tobacco shipments are sent from or received in Richmond, VA.

[4] Agriculture and energy are not individual FAF sectors, but rather amalgamations of sectors that appear separately in the FAF's categorization. Manufacturing sectors are in general broader than non–manufacturing sectors. For example, agricultural sectors include such categories as cereal grains, meat/seafood, and tobacco, while machinery is itself a category.

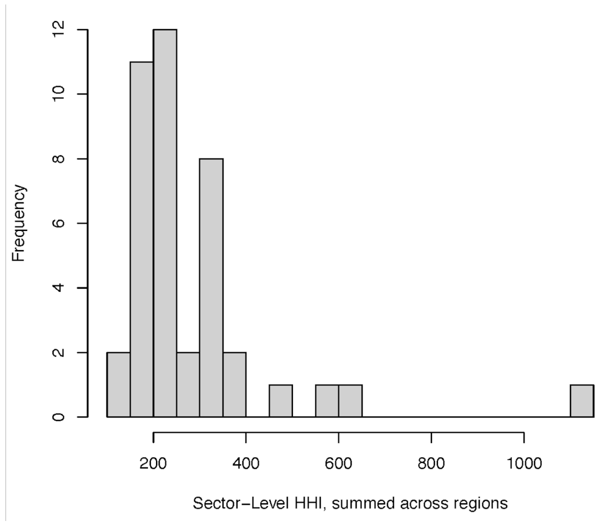


Figure 4: Histogram of the Herfindahl–Hirschman Index by Sector

4 Gravity Regression

4.1 Theoretical Foundation

I assume that there are K regions in the United States, and a representative consumer in city i has utility function

$$\left(\sum_{k=1}^{K} c_{kij}^{\frac{\theta_j-1}{\theta_j}}\right)^{\frac{\theta_j}{\theta_j-1}}$$

for each type of good j. This utility function implies that the representative consumer will be spending

$$E_{kij} = P_{kij}c_{kij} = P_{ij}^{\theta_j-1}P_{kij}^{1-\theta_j}(P_{ij}c_{ij}) \tag{1}$$

on imports in sector j from region k, where P_{ij} is the aggregate price of sector j goods in region i and P_{kij} is the specific price of goods in sector j that originate in region k.

Next, I assume that $P_{kij} = P_{kkj}\tau_{kij}$, where P_{kkj} is the price of sector j goods produced and sold in k and τ_{kij} is the additional cost of shipping the good from k to i. This τ_{kij} is the quantity of greatest interest.

Taking logs of both sides of (1) yields the equation

$$\log(P_{kij}c_{kij}) = (\theta_j - 1)\log(P_{ij}) + (1 - \theta_j)\log(P_{kkj}) + (1 - \theta_j)\log(\tau_{kij}) + \log(c_{ij})$$

The left–hand side of the equation corresponds to the trade flows observed in the data. θ_j is the trade elasticity for sector j, and I can represent P_{ij}, P_{kkj}, and c_{ij} with fixed costs since they do not vary across k and i simultaneously. The gravity equation then becomes

$$\log(\text{Imports}_{kij}) = \beta_0 + \sum_{n=1}^{N} \beta_n X_n + \alpha_{ij} + \alpha_{kj} + \varepsilon_{kij} \tag{2}$$

where α_{ij} and α_{kj} are origin–sector and destination–sector fixed costs, respectively. These fixed costs account for regional factors that do not vary across origin–destination pairs, and I cluster standard errors by origin and destination.

Unlike in Anderson and Van Wincoop (2003), the CES demand function introduced in this section does not contain expenditure weights across goods k and does not normalize prices to one. This difference becomes irrelevant by the writing of the applied gravity regression equation (2), which is the same in both setups. This paper innovates through its application of a CES gravity structure to U.S. *sub*national trade, and the different data framework required for such an estimation.

A major problem with estimating subnational trade is that a significant fraction of locations do not trade with each other at the level of individual sectors. Feenstra and Hong (2023, working paper) addresses this issue by using a translog utility function to allow for zero trade levels between regions. However, with translog preferences, the measurement and interpretation of elasticities becomes less straightforward. A translog gravity regression as in Novy (2013) also requires the researcher to obtain measurements of the variety of goods available in each location, which is all but impossible for U.S. subnational regions. I therefore use a more straightforward CES consumer demand function and run the gravity regressions using a Pseudo–Poisson Maximum Likelihood (PPML) methodology.[5]

[5] Heckman (1979) proposes another method for estimating a log regression when many observations of the left–hand side variable are zeroes. However, his methodology does not allow for fixed effects to appear in the regression.

4.2 Gravity Variables

I model trade costs τ_{kij} as a function of the (logged) distance[6] between regions and a binary variable indicating whether regions k and i are in the same state, as shipping goods across state lines may impose additional administrative or regulatory costs beyond the value of the distance. I also include gravity variables designed to be subnational counterparts to the language, currency, and other categories of binary variables commonly introduced into the standard gravity models that include multiple countries. I outline these variables below.

- Political Lean: Two regions with similar political views may be more likely to trade with one another than two regions that do not, due to shared cultural or business linkages. I obtain county–level 2020 vote totals for the two major parties from the MIT Election Data and Science Lab and aggregate these figures to get the total fraction of votes allocated to each major–party candidate at the regional level. The *Political Lean* variable is equal to one for a pair of regions if both regions awarded more than sixty percent of their votes to the same candidate in 2020, and zero otherwise.

- Elevation: Since the majority of subnational shipments in the U.S. are transported by land, a substantially higher increase in elevation could inhibit trade between areas for a given level of distance. I include in the regression the product of the logged distance between regions and the average distance in elevation between those regions. I compute the average change in elevation between two regions with the same method that I use for computing the average distance between two regions.[7]

- Coastline: Two regions that both have access to the ocean may be more likely to trade with one another than two regions that do not both have access to the ocean. Alternatively, they might be less likely to trade with one another if trade with foreign nations is a substitute for shipments sent within

[6] I construct distance between U.S. regions using the NBER's County Distance Database, a data set of great–circle distances between U.S. counties based on internal points in each geographic area. I then aggregate the counties to form FAF geographic regions. For regions that are also metropolitan areas, I cross–reference the metropolitan area with a list of its constituent counties found in the Bureau of Labor Statistics' County–MSA crosswalk, and for regions that are outside metropolitan areas, I include the list of all counties in the relevant state that were not included in any of that state's metropolitan areas.

I find the distance between regions by taking the average of the distances between the constituent counties in each regions. Mathematically, suppose that region 1 has N_1 counties and region 2 has N_2. The average distance between the two counties would then be

$$\frac{1}{N_1 N_2} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \text{Distance}_{ij}$$

[7] This measure only takes into account the change in elevation between the origin region and destination region. If two regions were at the same elevation but are separated by a mountain range, the elevation variable would be zero.

the United States. I model this presence as a binary variable equal to one if the relevant FAF regions each border at least one of the following: the Atlantic Ocean, Pacific Ocean, or Gulf of Mexico. For example, if one region bordered the Pacific Ocean and one region bordered the Gulf of Mexico, their binary value is one. If one or both of the two regions does not border any ocean, their binary value is zero.

- Canada: The presence of a border with Canada may affect trade between two regions of the U.S. I model this variable as a binary variable equal to one if the FAF regions both border Canada and zero if at least one does not. I include regions that border one of the Great Lakes as regions that border Canada, since they may access Canada by water.[8]

- Black/Hispanic: The presence of large minority communities may enhance trade between two U.S. regions due to cultural or business ties between those communities. The effect of these cultural ties is especially pronounced in the case of Hispanic communities, who may conduct business in Spanish. I model *Black* and *Hispanic* as separate variables.[9] For each variable, I include in the regression a binary variable equal to one if the relevant minority group comprises at least ten percent of the population of both regions and zero if it comprises less than ten percent of at least one region's population.

  Minority population estimates are available from the U.S. Census Bureau at the county level, and I sum up county estimates by region to find the relevant population fractions in each region. I measure *Black* as the fraction of the population that is Black or African American alone or in combination, and I measure *Hispanic* as the fraction of the population that is Hispanic alone or in combination. I obtain population estimates by county and sum them up by region to find relevant population fractions in each region. The median FAF region is 11.1% black and 10.6% Hispanic. Laredo, TX has the highest concentration of Hispanic Americans at 95.3%, and Memphis, TN has the highest concentration of Black Americans at 52.3%.

- Region: The *SameState* variable accounts for the effect of two regions being in the same state, but regional or cultural ties may also enhance trade between Census areas. I include in the regression a

[8] I was unable to model this border effect for Mexico because not enough FAF regions border Mexico. Out of the 17424 regional pairs, only forty–two included two regions that both have a Mexican border. I was likewise unable to model the Great Lakes and Canada separately.

[9] I do not include an indicator variable for Asian population due to lack of variation; the median FAF region is only three percent Asian by population.

binary variable equal to one if both regions are part of the same broader Census area[10] and zero if at least one of the regions is not.

- Urban/Rural: Since urban regions and rural regions specialize in producing different types of goods, regions with different population concentrations may be more likely to trade with one another than regions with similar concentrations. I define a pair of regions as both being 'urban' if they both have population densities above 500 people per square mile, and 'rural' if they both have population densities below 100 people per square mile. I gather data on county–level populations and land areas from the Census Bureau and add them up to get population densities for the FAF regions.

  Unlike the other subnational gravity regressors, this variable is clearly vulnerable to reverse causality as trade patterns can obviously affect the concentration of population within a given geographic location. I avoid this issue by classifying regions into broader population density–based categories rather than including the population density itself as a regressor. While subnational trade could affect population movement across regions, the general categorization of regions as 'urban' or 'rural' is more fixed. The inclusion of this variable also removes a large source of omitted variable bias from the political lean variable, as the urban/rural nature of an area is closely related to its political orientation.

5 Gravity Equation Results

I express all non–zero subnational shipment values in logs.

5.1 Aggregate Results

Before breaking down gravity results by sector or mode of transportation, I present results of gravity regressions on all subnational trade observations. I run the Column 1 regression on data aggregated across all sectors and modes, while Column 2 and Column 3 are run on pooled data across sectors and modes, respectively. I interact origin– and destination–fixed effects with sectors in Column 2 and regions in Column 3.

These preliminary results indicate that distance and state borders have strongly significant effects on

[10] States that comprise each Census area are as follows: New England (CT, MN, MA, NH, RH, VT, NJ, NY, PA); Midwest (IN, IL, MI, OH, WI, IA, KA, MN, MO, NE, ND, SD); South (DE, DC, FL, GA, MD, NC, SC, VA, WV, AL, KY, MS, TN, AR, LA, OK, TX); West (AZ, CO, ID, NM, MT, UT, NV, WY, AL, CA, HA, OR, WA).

	Aggregate	Pooled by sector	Pooled by mode
Dist	-1.29***	-.1.28***	-1.40***
SameState	.746***	.715***	.704***
Political Lean	.137***	.0885***	.0735***
Elevation $\times$ Dist	-9.61×10^{-5}***	-5.10×10^{-5}	-3.41×10^{-5} $* **$
Coastline	-.156*	-.150***	-.128**
Canada Border	-.0451	-.0923	-.0750
Black	.0400	.120***	.0641
Hispanic	.0174	2.01×10^{-3}	-.311***
Northeast	-.119***	-.140	-.0351
South	.0248	.0471	.0169
Midwest	.219**	.212***	.223**
West	-.103	-.0910	.0453
Urban	-.0862***	-.106**	-.109*
Rural	-.0528	-.0298	-.0580
Sector–region fixed effects	No	Yes	No
Mode–region fixed effects	No	No	Yes
Adjusted pseudo R–squared	.860	.779	.864

Table 3: Gravity regression results, without dividing by sector or mode
***: $p < .001$; **: $p < .05$; *: $p < .1$

subnational trade, with distance predicting lower trade and being in the same state predicting higher trade. Two regions trade more with each other if both are in the Midwest, while indicator variables for the other Census areas do not have a consistent effect on trade.

Furthermore, the combination of subnational shipments across sectors and modes appears to introduce aggregation bias. Disaggregating across sectors causes the *Black* indicator variable to become statistically significant at the one percent level, while the elevation and *Northeast* cease to be statistically significant at any level. With these findings in mind, I now proceed to a set of gravity estimations broken down by sector.

5.2 Sector–Level Results

It is not possible for me to list results from all forty–two sector–level gravity regressions in a concise form. I will instead summarize the results of these regressions, then go into more detail about the coefficients on distance and *SameState* as these results are larger and more significant.

For each variable, Table 4 shows the number of sectors, out of forty–two, that are either significantly negative, significantly positive, or not statistically significant at the ten percent level. Distance and *SameState* have consistently significant impacts, with distance elasticities being negative and SameState elasticities being

	Negative	Non–significant	Positive
Distance	41	1	0
Same State	0	6	36
Political Lean	1	31	10
Elevation × Distance	3	29	10
Coastline	8	32	2
Canada Border	6	32	4
Black	2	34	6
Hispanic	4	33	5
Northeast	7	28	7
South	1	34	7
Midwest	2	28	12
West	2	33	7
Urban	3	36	3
Rural	9	30	3

Table 4: Summary of Gravity Regression Coefficients

positive. Distance is only insignificant for transport equipment, and state borders are only insignificant for six sectors including cereal grains, coal, and crude petroleum.

The remaining variables do not display statistically significant relationships with trade for around three quarters of sectors. This result suggests that these regional characteristics do not have a strong effect on subnational trade, especially compared to distance and state borders, but are still important for certain specific commodities Two regions that both share a coastline are less likely to trade with one another than two regions who do not, suggesting that coastal regions choose to trade with foreign nations instead of shipping goods to other regions of the U.S.[11] A similar political lean is more likely to predict higher trade than lower trade between regions, while rural status has the opposite effect as rural regions are unlikely to be centers of commerce for a large variety of sectors. Urban status, however, displays a positively significant relationship with trade as often as it does a negative one.

While being in the same state strongly predicts higher shipment levels, being in the same broader Census area has much less predictive power on trade. Among the estimates that are significant, regional dummies more often have positively significant effects on trade than negatively significant effects, with Midwestern regional identity having the strongest positive relationship with subnational trade. The Northeast dummy

[11]A region bordering the Pacific Ocean might be unlikely to trade with a region bordering the Atlantic Ocean due to the expense of shipping goods across the Panama Canal or around South America. However, the inclusion of distance in the regression controls for this issue. I do not make separate variables for the Atlantic and Pacific Ocean because very few of the regional pairs border the same ocean, and so the variable would have a lack of variation.

variable, however, is just as likely to predict lower trade between regions as it is higher trade between regions.

Energy goods such as coal, gasoline, or natural gas are generally shipped between regions with differing characteristics, as regions with high energy consumption needs are often not the same regions with large endowments of oil, gas, or coal. For example, areas with a similar political lean, areas in the Midwest and areas with a high concentration of black or Hispanic Americans all trade less crude petroleum with each other than areas that do not share these characteristics. Two regions will trade less coal, gasoline, and fuel oils if both regions are rural. However, in general, the gravity regressors in Table 4 are more likely to display statistically significant results for less frequently traded sectors than for more frequently traded sectors, as idiosyncratic local characteristics may have less impact on shipments in a commodity universally traded in large values across the country than shipments in a commodity whose scope is more limited. Just as the

Variable	Min	Median	Mean	Max
Dist	-3.75	-1.38	-1.47	-.368
SameState	-.416	.816	.778	1.84
Political Lean	-.493	.167	.0732	.547
Elevation × Dist	-2.87×10^{-3}	1.89×10^{-4}	-9.58×10^{-5}	1.16×10^{-3}
Coastline	-.916	-.223	-.173	1.74
Canada Border	-.921	-.0189	.171	1.62
Black	-.548	.177	.123	1.12
Hispanic	-.827	.112	.114	1.09
Northeast	-1.43	.0102	.309	4.82
South	-.866	.328	.337	1.59
Midwest	-1.07	.267	.238	1.44
West	-2.43	.563	.498	2.03
Urban	-1.71	-.120	-.0530	1.32
Rural	-1.37	-.153	-.217	.413

Table 5: Summary statistics of gravity regression coefficients

effects of distance and state borders on trade are more statistically significant than the other variables, Table 5 shows that they are larger in magnitude as well. Distance elasticities are well above one, and the median *SameState* coefficient is close to one.

As for other regressors, the positive effect of *Black* is larger on average than that of *Hispanic*, suggesting that despite the absence of language effects shared African–American demographics are more important for trade than shared Hispanic demographics. Regional effects on trade are largest for the West and weakest for the Northeast; the median effect on trade of two regions' both belonging to the Northeast is roughly zero.

The magnitude of estimates in Table 5 are largely uncorrelated with sector–level trade frequencies. With regard to distance and state borders, the two regressors with the most explanatory power, regression estimates in a sector have nearly zero correlation with that sector's proportion of total subnational trade by value. For other regressors, the coefficients in some highly traded sectors are large in magnitude but not significant at the ten percent level, explaining why significance is somewhat related to trade frequency but magnitude is not.

Finally, the results presented in this section underscore the importance of separating data by sectors when analyzing subnational trade. Although regional characteristics such as racial demographics or political affiliation have little impact on subnational trade in the aggregate or indeed for most sectors, they do have significant relationships on trade for certain particular sectors, a nuance lost when examining data exclusively at the aggregate level. Sector–level heterogeneity is especially valuable since many trade policy changes proposed by governments will not impact all sectors equally.

5.3 Distance and State Borders

I will now go into more detail about the effect of distance and state borders on trade. Table 6 shows the coefficients on distance and a binary variable indicating whether both regions are in the same state. These results are shown both for trade by value and for trade by weight.

Sector	Logged Trade by Value		Logged Trade by Weight	
	Distance	Same State	Distance	Same State
Live animals/fish	-3.65***	-.267*	-4.14***	-.223
Cereal grains	-2.56***	.866***	-2.96***	.827***
Other agricultural products	-1.66***	1.14***	-2.15***	1.50***
Animal feed	-1.63***	.827***	-2.19***	.900***
Meat/seafood	-1.36***	.396***	-1.49***	.377***
Milled grain products	-1.41***	.0184	-1.55***	.239***
Other foodstuffs	-1.39***	.625***	-1.73***	.501***
Alcoholic beverages	-1.18***	1.87***	-1.48***	1.23***
Tobacco products	-1.46***	.738***	-1.35***	.810***
Building stone	-1.87***	1.00***	-2.49***	1.25***
Natural sands	-1.34***	.972***	-1.93***	1.58***
Gravel	-2.25***	.849***	-3.19***	1.47***
Nonmetallic minerals	-1.23***	1.03***	-1.71***	1.27***
Metallic ores	-.825***	1.35***	-1.42***	2.69***
Coal	-2.47***	.548***	-2.96***	.251***
Crude petroleum	-1.30***	.278**	-1.49***	-.150***
Gasoline	-2.22***	.760***	-2.33***	.634***
Fuel oils	-1.85***	1.05***	-2.06***	.922***
Natural gas and other fossil products	-2.54***	-.319***	-3.17***	-.480***
Basic chemicals	-1.03***	.761***	-1.54***	.412***
Pharmaceuticals	-1.47***	.858***	-.891***	1.23***
Fertilizers	-1.64***	1.30***	-1.97***	1.23***
Chemical products	-1.01***	.530***	-1.28***	.537***
Plastics/rubber	-.981***	.638***	-1.03***	.592***
Logs	-1.33***	1.20***	-1.86***	1.30***
Wood products	-1.84***	.540***	-2.09***	.639***
Newsprint/paper	-1.26***	.552***	-1.40***	.531***
Paper articles	-1.43***	.512***	-1.68***	.524***
Printed products	-.953***	.418***	-1.19***	.778***
Textiles/leather	-.808***	.823***	-1.11***	.845***
Nonmetal mineral products	-1.53***	.671***	-2.15***	.891***
Base metals	-1.40***	.583***	-1.56***	.528***
Articles made from base metal	-1.22***	.784***	-1.47***	.730***
Machinery	-.974***	.835***	-1.00***	.943***
Electronics	-.824***	1.00***	-1.09***	.807***
Motorized vehicles	-.895***	.886***	-1.02***	.757***
Transport equipment	-.384***	1.56***	-.856***	1.92***
Precision instruments	-.683***	1.03***	-1.03***	1.02***
Furniture	-1.02***	.896***	-1.21***	.857***
Miscellaneous manufacturing products	-.727***	1.20***	-1.34***	.738***
Waste/scrap	-1.68***	.670***	-2.40***	.786***
Mixed freight	-2.43***	.257***	-2.63***	.347***

Table 6: Gravity regression results, by sector (***: $p < .001$; **: $p < .05$; *: $p < .1$)

The second and third columns of Table 6 display trade elasticities with respect to distance and the presence of state borders for all forty–two sectors on which the FAF contains data. These findings suggest a consistent relationship between distance and trade at the subnational level, which contrasts with the inconsistent significance often found when estimating international gravity relationships. As the United States is a large single market with constitutional restrictions on barriers to interstate commerce, factors that complicate the gravity relationship at the international level may be absent within the United States.

Overall, agricultural goods and raw materials have higher distance elasticities than manufactured goods, a finding explained by raw materials' higher depreciation costs and likelihood to spoil. All agricultural sectors (rows two through eight) have distance elasticities above one. Live animals and fish have the highest distance elasticity of all forty–two sectors, at -3.65. Sectors with low distance elasticities, meanwhile, include transport equipment, precision instruments, and miscellaneous manufacturing products, all of which have elasticities less than one with respect to distance.

Being in the same state affects trade positively, with most SameState coefficients falling between one–half and one. Regions that display higher trade elasticities with respect to distance do not necessarily display higher elasticities with respect to state borders, as SameState elasticities are generally higher for manufactured goods than for raw materials or agriculture. Alcoholic beverages, miscellaneous manufacturing, and transport equipment show the highest elasticities with respect to state boundaries, implying the existence of stronger and more dispersed regulations that hinder trade between states in those sectors. Alcoholic beverages are more vulnerable than other sectors to state bureaucracy, as state policies differ with regard to the sale and distribution of alcohol.

5.4 Modes of Transportation

Since the number of transportation modes is lower than the number of sectors, I present mode–level gravity results in their entirety.

Table 7 shows that distance elasticities are highest for pipeline trade, followed distantly by truck trade and water trade. If the distance between locations increases by one percent, those two locations are predicted to conduct 2.39 percent less trade by pipeline. However, distance is not important at all for trade shipped via airplane, with the coefficient being small and not statistically significant. All other distance elasticities are statistically significant at the one percent level.

	Truck	Rail	Water	Air	Pipeline
Dist	-1.45***	-.604***	-1.04***	.0662	-2.34***
SameState	.760***	1.12***	4.00×10^{-4}	.656	-.472***
Political Lean	.0949**	.218	-.194	-.364**	-.366**
Elevation × Dist	7.92×10^{-5}	-4.95×10^{-4}*	-2.45×10^{-3}***	3.82×10^{-4}***	1.28×10^{-5}
Coastline	-.136*	.450*	1.06***	-.505	-.602**
Canada Border	-.114*	-.254	1.67**	.213*	.102
Black	.0791	.122	1.36***	-.754**	-.151
Hispanic	.0183	-.402*	1.26**	-.535***	.129
Northeast	-.0133	.728**	.847	.796*	2.81***
South	.113	-.190	-.477	-.314**	.187
Midwest	.146**	-.818***	-.775	.879***	.350*
West	.0367	.633	9.87***	1.21	.998***
Urban	-.0448	-.555***	-1.26***	-.206	-.296
Rural	-.0675	-.0755	-.0644	-.297	-.123
Adjusted psuedo R–squared	.875	.599	.821	.711	.803

Table 7: Gravity regression results, by mode of transportation

As with sectors, modes of transportation that show the greatest elasticity in magnitude with respect to distance do not necessarily display the greatest elasticities with respect to the SameState variable, and vice versa. Rail and truck transportation are most positively affected by two regions' being in the same state, and I speculate that truck and rail transportation, which rely on the manmade networks of highways and train tracks, are more affected by trade and travel regulations that may differ between states. Pipelines, meanwhile, show a negative relationship between trade and being in the same state. Pipelines transport oil across multiple states, and the transportation of oil via pipeline within a state might not be profitable.

PPML estimations of international gravity have typically yielded distance elasticities between -.75 and -.95 (Anderson, Larch, and Yotov (2022); Yotov et al. (2016); Agnosteva, Anderson, and Yotov (2014); Santos Silva and Tenreyro (2006)). The rule of thumb for international gravity regressions more generally is that the distance elasticity will be around -1, with estimates including -.809 (Yotov, Larch, and Heid (2015)), -1.12 (Olivero and Yotov (2012)), or -1.2 (Helpman, Melitz and Rubinstein (2008)). My distance estimates are in the general range suggested by this literature, skewing slightly larger. The trade distances pertinent to subnational trade are generally smaller than the distances observed in international trade, and an increase in decrease of ten percent could make more of a difference if the original distance is 300 miles as opposed to 3000.

Predictably, two regions that both share a coastline are more likely to trade with each over water and less likely to trade with each other by other means, with the intriguing exception of rail. A higher difference in elevation makes air travel more likely but rail and water travel less likely, implying that air transport may substitute for rail or water transport when the goods must be shipped over a large change in elevation. More intriguing are results showing that high minority representation in a pair of regions predicts higher water trade between those regions but lower trade by airplane.

Regional identity has the strongest effect on trade in the Midwest; due to vast distances and relatively little rail infrastructure, two regions in the Midwest exchange more shipments by truck, air and pipeline and fewer shipments by rail compared to two regions that are not in the Midwest. However, the *Midwest* dummy variable coefficient is not significant for water trade, despite the presence of the Great Lakes in several Midwestern states. The effect of 'West' on water travel may be large because California and Washington conduct an amount of trade disproportionate to their geographic or population size. Although two regions' being in the same *state* is negatively related to their trade via pipeline, their being in the same region has a significant effect on pipeline trade for all areas except the South.

5.5 Other Robustness Checks

As a robustness check, I also perform gravity regressions using logged trade by volume (in tons) as the dependent variable instead of logged trade in value. Results of these regressions are reported in columns three and four of Table 5. Generally speaking, the ordinality of trade elasticities with respect to distance does not change much; live animals and fish are still the most elastic sector, and agricultural trade are more elastic with respect to distance than trade in manufactured goods. However, almost all sectors report higher elasticities for trade in tons than for the current value of trade. One explanation for this pattern is that exporters can respond to longer shipping distances by charging more, which would raise the current value of the shipment without altering the amount being shipped.

For the majority of sectors, trade elasticities with respect to state borders are smaller for trade by weight than for trade by value. Pharmaceuticals and gravel have value elasticities well below one and weight elasticities above one, while electronics and miscellaneous manufacturing products have value elasticities above one and weight elasticities below one. Crude petroleum displays a negative elasticity for trade by weight and a positive elasticity for trade by value.

Gravity regression results are virtually unaltered when using data from 2019, 2020, 2021, or 2022, so President Trump's trade wars and the pandemic did not have much of an impact on how regional characteristics affect subnational shipments. I also estimate gravity regressions measuring distance as a population–weighted average of the distance between each area's constituent counties rather than a simple unweighted average. This adjustment also did not cause any noticeable changes in the results.

Lin and Sim (2012) show that the effect of logged distance on logged trade rises over time, with a coefficient of about -.75 in the 1950s and -1.25 in the 1990s. These estimates contradict my finding that gravity results vary very little over the past five years, but of course Lin and Sim are working with a much longer time period.

6 Trade Costs

In Section 4.1, I introduce a theoretical foundation for estimating subnational gravity equations, in which I replaced logged trade costs with a linear function of logged distance and other bilateral regional characteristics. I can then rearrange this formulation to get trade costs for every pair k and i.

$$\tau_{kij} = \exp\big(\frac{\sum\limits_{n} \beta_n X_n}{1-\theta_j}\big)$$

I use this formula to back out regional pair–wise trade cost estimates broken down by sector and by mode of transportation. I estimate sector–level θ_j using data from the USITC's DataWeb.

Due to the specification that $P_{kij} = P_{iij}\tau_{kij}$, I interpret a transportation cost τ_{kij} as the *ceteris paribus* additional cost to region i of consuming sector j produced in region k compared with the cost of consuming sector j produced in region i (in other words, consumption of goods produced locally). If τ_{kij} were 1.4, then a sector j good produced in k would be forty percent more expensive to a region i consumer than a good produced in region i. To render this consumer indifferent between good iij and good kij, region k would need to possess some other advantage, such as higher productivity, that compensates for the additional costs imposed by trade.

6.1 Estimating θ_j

I estimate Armington elasticities $\{\theta_j\}_{j=1}^{42}$ using a similar methodology to the gravity regression discussed in Section 4. A CES utility function gives us the regression

$$\log(\text{Imports}_{kij}) = (1 - \theta_j)\log(f_{kij}) + \alpha_{ij} + \alpha_{kj} + \varepsilon_{kij} \tag{3}$$

where α_{ij} and α_{kj} are fixed effects, f_{kij} is an international freight cost factor and ε_{kij} is the error term. The difference between estimating this equation and estimating the trade costs in (2) is that here, I run a regression on international data to calibrate values of θ_j, whereas I use (2) to extract subnational values of τ_{kij} given pre–calibrated values of θ_j.

The international data to estimate (3) comes from the USITC's DataWeb, and records sector–level import values for 225 exporting entities and forty–two ports of entry into the U.S in 2018. I calculate f_{kij} as the ratio of the landed duty–paid value of the imports to their customs value, as in Riker (2022) and Riker (2019). Using the total trade values across all sectors, I obtain a value of $\theta = 6.82$, which lies in the middle of the 5–10 range suggested in Anderson and van Wincoop (2004).

6.2 Trade Cost Results

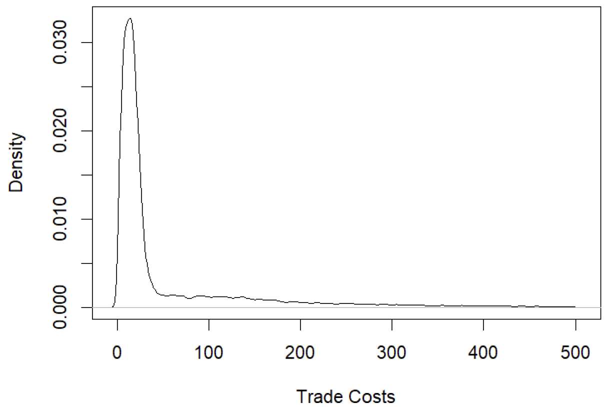
All trade cost density plots and summary statistics discussed here exclude estimated trade costs between two regions that do not trade with each other.

Trade cost estimates vary widely among sectors, with motorized vehicles having the lowest trade costs and non–metallic mineral products[12] having the highest. Geographically, a shipment from Beaumont, Texas to the 'Rest of Colorado' would incur the lowest transportation cost out of all regional pairs.

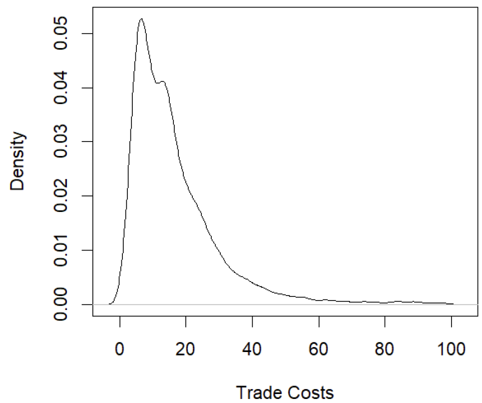
Figure 5 shows density plots for five frequently traded sectors.[13] Motorized vehicles, the most frequently traded sector, have a median trade cost of 1.40, meaning that shipping a car from one subnational region to another would, on average, add forty percent to the cost of the car. Agricultural goods generally have higher transportation costs than manufactured goods, reflecting their higher trade elasticity of trade with

[12] This sector includes articles made of ceramics, glass, or asphalt, including household appliances, as well as concrete pipes and quicklime.

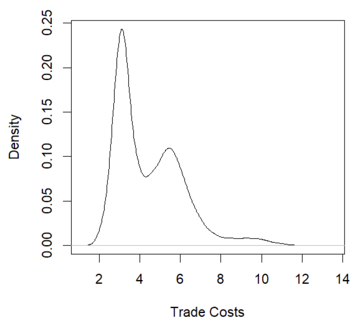
[13] I truncate the long–tailed density plots to preserve image clarity. In all cases, the plots still show over ninety percent of the imputed trade costs, and the omitted trade costs correspond to pairs of regions that were not trading in that sector or mode in the first place.



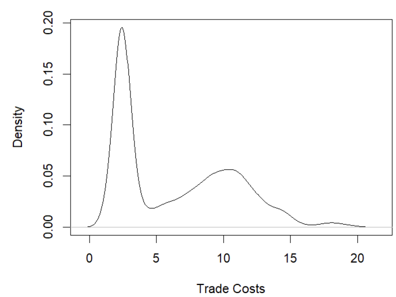
(a) Agricultural Goods



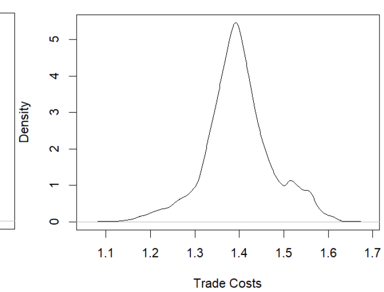
(b) Energy Goods



(c) Chemicals/Pharmaceuticals



(d) Electronics/Machinery



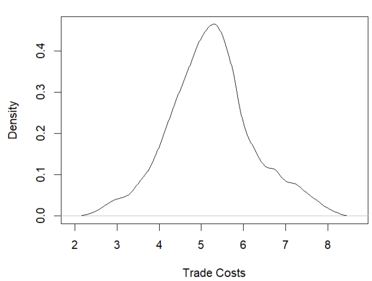
(e) Motor Vehicles

Figure 5: Trade Cost Density Plots for Selected Sectors

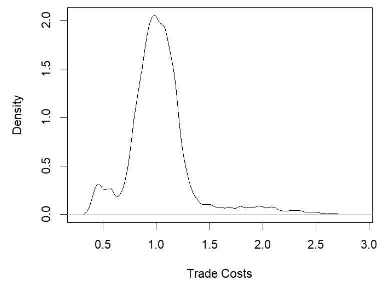
respect to distance discussed in Section 4. Cereal and milled grain products are especially costly to trade, with respective median trade costs of 339 and 136. Energy sector goods have transportation costs lower than agricultural goods but higher than manufactured goods, with crude petroleum having a median trade cost of 17.2 and gasoline having a median trade cost of 6.82.

As shown in Figure 6, goods shipped by airplane have the lowest trade cost, with the median only slightly higher than one. Rail shipping also imposes a fairly low trade cost. Truck shipping is more expensive, but pipeline transportation is the most costly mode by far, imposing a median trade cost of almost 1100 percent compared to a good produced locally.

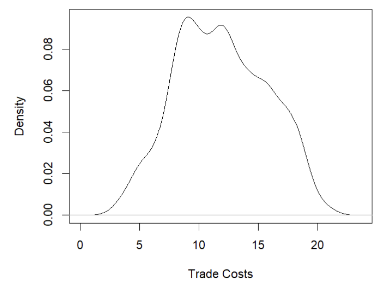
Water and rail transportation trade costs have a longer right tail than truck and pipeline transport, even though the median trade cost of truck and pipeline transportation is higher. This long tail can be explained by the distribution of different types of transportation within the United States. While rail or water may be



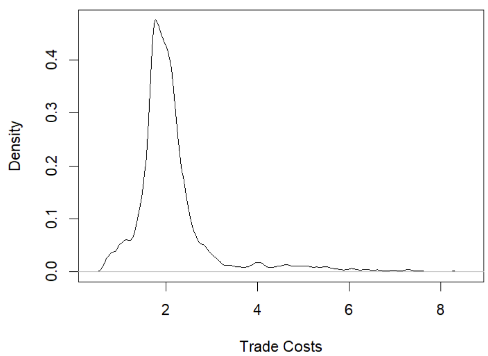
(a) Truck



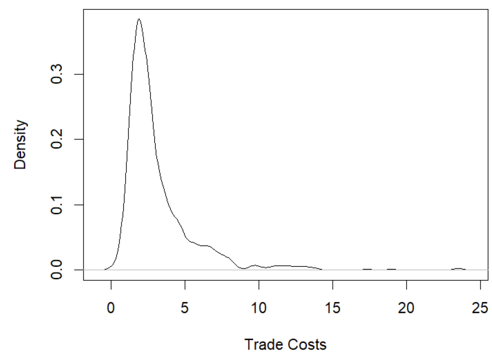
(b) Air



(c) Pipeline



(d) Rail



(e) Water

Figure 6: Trade Cost Density Plots for Different Modes of Transportation

easier for two regions directly connected by these two means of transportation, some regions have no rail or water connections at all, but the Interstate Highway System is ubiquitous and connects every region within the continental United States. The long right tail may also explain why truck shipments are so much more common than rail or water shipments despite having a higher median transportation cost.

The trade cost estimates displayed in Figures 5 and 6, with the overall median of 5.00, are generally higher than trade costs estimated in the international trade literature. Anderson and van Wincoop (2004), which surveys the trade cost literature, determines that trade costs tend to be in the general range of 170%, or $\tau \approx 2.7$. Egger, Larch, Nigai, and Yotov (2020) find a median logged transport cost of .9, implying that the median transport cost would be about 2.46. Anderson and van Wincoop (2003) estimate transport costs for U.S. states and Canadian provinces, finding them to be between 1.3 and 1.5 for all states; these transport cost measures incorporate both subnational and international trade. This contrast suggests that the relative cost of consuming locally–made goods versus goods sent from elsewhere in the United States is therefore higher than the relative cost of consuming American–made goods versus consuming foreign–made goods. A decreasing marginal effect of distance could drive this observed phenomenon; American–made goods already must be transported some distance even without international tariffs, and the extra distance contributed by overseas shipping does not add as much cost. Furthermore, the transport cost measures calculated here could be picking up some of the fixed cost of shipping, since I do not do any precise decomposition. Egger et al. (2020) do perform such a decomposition and find that the median "log–scaled non–tariff transportation cost", a measure that does include fixed costs, is 6.86.

Last, it should be noted that the trade cost computation method used in this section is more useful for comparing trade between two pairs of locales than looking at U.S. trade holistically. If distance were the most important criterion in predicting a region's trade accessibility, then coastal cities would be more expensive to ship to than cities that are farther inland and thus more centrally located within the country. We know this cannot be true, as coastal metropolitan areas such as Los Angeles account for some of the highest shares of subnational shipments.

7 Conclusion

In this paper, I use data from the Freight Analysis Framework to estimate gravity regressions at the subnational level for the United States. The Freight Analysis Framework dataset encompasses 132 subnational regions, forty–two sectors, and eight major transportation categories. I make a novel contribution to the literature by estimating how gravity variables such as distance and state borders affect trade at the subnational level.

Gravity regressions reveal a statistically significant negative relationship between logged distance and logged trade for all forty–two sectors and five major modes of transportation. Agricultural trade is generally more responsive to distance than trade in manufactured goods, with trade in energy goods falling between these two other categories. Among modes of transportation, truck and pipeline are most responsive to distance, with airplane trade not being much affected by distance. The data also show a statistically significant correlation between two regions' trade and their being in the same state, with few exceptions.

Other regional variables such as political leanings, coastline, and demographics are significantly correlated with trade for around a quarter of the sectors included in the FAF data. A similar political lean or concentration of racial minorities predicts higher trade between regions more often than lower trade, while regions that both border the ocean are predicted to do less trade than two regions that do not border the ocean. Two regions that are part of the same Census area are more likely to exchange shipments than two regions that are not, but the correlation is not as strong as it is with two regions that are in the same state.

I then employ the CES framework used to estimate gravity regressions to back out estimates of transportation costs. Trade cost estimates are 7.00 on average but vary widely between sectors, with motor vehicles having the lowest trade cost. Median trade costs are higher for truck shipments than rail, water, or air shipments, even though truck shipments comprise seventy percent of all U.S. subnational trade by value.

Future research on this topic should do further analysis into some of the gravity relationships discussed in section 5, untangling the extent to which observed statistically significant relationships are causal. For example, similar political attitudes could enhance commerce and trade between two regions, but there is no causal reason why regions of a similar political bent would be more likely to trade manufactured goods than agricultural goods, without considering related factors such as population density. Future subnational gravity research should also considering incorporating gravity regressors that vary by sector or mode of transportation as well as region. However, reverse causality could pose a major challenge with incorporating such variables into the regression. Sector–level employment shares, for example, would definitely be influenced by sector–level subnational shipments.

Finally, if such data were available, an interesting exercise would be to extend the analysis presented here to a country where subnational regions can levy tariffs on each other. The Commerce Clause of the U.S. Constitution establishes a free trade zone among U.S. states, but in other countries, such as Brazil and India, regional governing bodies can and do impose regional trade restrictions. Subnational gravity results may differ in the presence of such administrative trade barriers.

8 References

Acemoglu, Daron, David Autor, David Dorn, Gordon H. Hanson, and Brendan Price. "Import Competition and the Great US Employment Sag of the 2000s." *Journal of Labor Economics* 34, no. S1. (January 2016): S141–98.

Agnosteva, Delina, James E. Anderson, and Yoto Yotov. "Intra–national Trade Costs: Measurement and Aggregation." *European Economic Review* 112 (2019): 32–50.

Allen, Treb, and Costas Arkolakis. "Trade and the Topography of the Spatial Economy." *The Quarterly Journal of Economics* 129, no. 3. (August 1, 2014): 1085–1140.

Allen, Treb, and Costas Arkolakis. "The welfare effects of transportation infrastructure improvements." *Review of Economic Studies* 89, no. 6 (2022): 2911–2957.

Anderson, James E. and Eric van Wincoop. "Gravity with Gravitas: A Solution to the Border Puzzle." *The American Economic Review* 93, no. 1 (March 2003): 170–192.

Anderson, James E. and Eric van Wincoop. "Trade Costs." *Journal of Economic Literature* 42, no. 3 (September 2004), 691–751.

Autor, David H, David Dorn, and Gordon H Hanson. "The China Syndrome: Local Labor Market Effects of Import Competition in the United States." *American Economic Review* 103, no. 6. (October 1, 2013): 2121–68.

Caliendo, Lorenzo, Maximiliano Dvorkin, and Fernando Parro. "Trade and Labor Market Dynamics: General Equilibrium Analysis of the China Trade Shock." *Econometrica* 87, no. 3. (2019): 741–835.

Caliendo, Lorenzo, and Fernando Parro. "Trade Policy." NBER Working Paper 29051.

Donaldson, Dave, and Richard Hornbeck. "Railroads and American Economic Growth: a 'market access' Approach." *The Quarterly Journal of Economics* 131, no. 2 (2016): 699–858.

Egger, Peter, Mario Larch, Sergey Nigai, and Yoto Yotov. "Trade costs in the global economy: Measurement, aggregation, and decomposition." WTO Staff Working Paper, 2021.

Feenstra, Robert, and Chang Hong. "Estimating the Regional Impact of Tariff Changes: Application to the United States." Working Paper, 2023.

Fuchs, Simon, and Woan Foong Wong. "Multimodel transport networks." Federal Reserve Bank of Atlanta Working Paper, 2023.

Gurevich, Tamara, and Peter Herman. "The Dynamic Gravity Dataset: 1948–2016." U.S. International Trade Commission Working Paper Series, February 2018.

Hakobyan, Shushanik, and John McLaren. "Looking for Local Labor Market Effects of NAFTA." *Review of Economics and Statistics* 98, no. 4. (October 2016): 728–41.

Heckman, James J. "Sample Selection Bias as a Specification Error." *Econometrica* 47, no. 1. (January 1979): 153.

Heid, Benedikt, Mario Larch, and Yoto Yotov. "Estimating the effects of non–discriminatory trade policies within structural gravity models." *Canadian Journal of Economics* 54, no. 1 (February 2021): 376–409.

Helpman, Elhanan, Marc Melitz, and Yona Rubinstein. "Estimating Trade Flows: Trading Partners and Trading Volumes." *Quartrely Journal of Economics* 123 (2008): 441–487.

Jaworski, Taylor, Carl Kitchens, and Sergey Nigai. "Highways and Globalization." *International Economic Review* 64, no. 4 (2023): 1615–1648.

Lin, Faqin, and Nicholas C.S. Sim. "Death of Distance and the Distance Puzzle." Economics Letters 116, no. 2. (August 2012): 225–228.

Olivero, Maria, and Yoto Yotov. "Dynamic Gravity: endogenous country size and asset accumulation." *Canadian Journal of Economics* 45, no. 1 (February 2012): 64–92.

Novy, Dennis. "International Trade without CES: Estimating Translog Gravity." *Journal of International Economics* 89, no. 2. (March 2013): 271–282.

Riker, David. "Estimating U.S. Import Penetration in Sub-National Regions: International Trade and Economic Development." *Journal of International Trade and Economic Development* 29, no. 7. (2020): 891-906.

Riker, David. "State-Level Import Penetration." U.S. International Trade Commission Working Paper Series, March 2022.

Santos Silva, J.M.C., and Silvana Tenreyro. "The Log of Gravity." *The Review of Economics and Statistics* 88, no. 4 (2006): 641–658.

Yotov, Yoto V., Roberta Piermartini, José-Antonio Monteiro, and Mario Larch. "An Advanced Guide to Trade Policy Analysis: The Structural Gravity Model." WTO, November 15, 2016.