Radiating Trade:
Creating Gravity through Spatial Geometry

Seth Kingery *
University of Michigan

November 2016

Abstract

Current models generally create a gravity relationship using trade costs, and are dependent on the functional form of agent preferences. This paper seeks to examine if this relationship might not simply be a consequence of the spatial geometry of agents searching outward from their initial location, indifferent to costs and the choice of utility function. It constructs a simple model of agent search to generalize this process in a plane, and shows how this generates the standard inverse distance relationship in trade between countries. The paper then demonstrates how to easily combine a search-driven extensive margin and a cost-driven intensive margin in a single model. Finally, the paper shows how the search process is altered on the surface of a sphere (or globe), yielding a model that behaves identically to the standard gravity model over short distances, but with reduced trade flows at long distances. The predictions of the spherically modified model are strongly validated empirically. The slight difference between the spherical and standard models implies relatively little difference in predicted trade flows. But the presence of a geometry-driven modification to the standard model supports the notion that a large portion of the gravity relationship may be a latent feature of the geometry of the agent search process. It also suggests that trade flows might not be predominately determined by the structure of trade costs, with implications for both welfare and modeling strategies.

1 Background

Most models of trade produce a "gravity" relationship for bilateral trade between countries. That is, bilateral trade flows are proportional to the product of the GDPs of the two countries divided by distance, as is well documented in the empirical evidence. In these models, this relationship is usually produced by including a cost for transporting goods between countries. Unfortunately, the costs needed to fit these models to the data are far larger than those observed in actual transportation costs. In addition, the gravity relationship has been extremely stable over time (such as in Disdier & Head 2008), despite falling transport costs, lengthening supply chains, and a host of

*I would like to express my deepest thanks to Alan Deardorff, as well as Sebastian Sotelo, Kyle Handley, Kathryn Dominguez, and Javier Cravino for their advice and patience. I would also like to thank Andrei Levchenko, for both his theoretical and technical assistance, together with Linda Tesar and Dominick Bartelme for their helpful discussions.

other changes in the real economy. This paper proposes an alternative to trade costs: that trade falls with distance because trading partners far away are less likely to be found.

Put differently, consider an economic agent standing in the middle of a space uniformly packed with potential trading partners arranged in larger and larger concentric circles. Our agent must select a trading partner from among those at a given distance. As the circles get further out, there are more agents on each circle because the larger circumference will "fit" more potential trading partners. If the probabilities of trading with all the partners on a given circle are the equal (*i.e.*, there is no *ex ante* heterogeneity between trading partners at a given distance), then a larger circle implies a lower probability of our economic agent selecting any *specific* trading partner on a given circle. Therefore, due to geometry, as our agent chooses among partners at a greater distance, the probability of selecting any one partner at that distance decreases because there are more options. And it happens that the decrease in selection probability with distance implied by this process is exactly consistent with the gravity relationship.

In order to examine this phenomenon more deeply, this paper will construct a simple model in which agents using a uniform random radial search pattern try to locate buyers. (The search process was chosen for simplicity, to the point of being somewhat abstract, but the paper shows why this is an underlying feature of any probabilistic search going outward from some starting point.) This search process, combined with a very simple model of an economy, produces gravity exactly. And because this model derives gravity from the geometry of search alone, this result is relatively insensitive to other modeling decisions, such as the structure of preferences, production, or the size of trade costs. This makes the geometric process easy to combine with existing models of preferences and trade costs, as will be demonstrated below. But more important, it implies that this geometric process *could explain why the gravity relationship has proven so stable across time and region despite changes in the global economy and trade technology*.

This insensitivity to economic changes also has substantial implications for how trade counterfactuals and welfare gains are modeled. But before developing the geometric process and examining those features of the model, it is helpful to first discuss key findings in the literature and some of the shortcomings of the current prevailing technique for achieving gravity: "implied" trade costs.

1.1 Gravity and Costs

The gravity relationship in trade was first noted more than 50 years ago (Tinbergen 1962). In its simplest form, it states that trade (for simplicity we will simply say exports) from county i to country j is of the form

$$X_{ij} = \zeta \frac{Y_i Y_j}{\delta_{ij}^{\epsilon}}$$

where Y_i is the GDP of country i, δ_{ij} is the distance between the two countries, and ϵ is expected to be equal to one.[1] This basic relationship has been observed across sectors, in final and intermediate goods (Miroudot et al 2009), and even for services (Kimura & Lee 2006, Walsh 2008). The equation is so powerfully present in the data that it has taken on a privileged status in the literature. Now, when trade flows deviate from gravity it is usually taken as evidence of trade distortion. But, while every popular contemporary model of trade reduces to gravity, the observation of gravity in trade ante-dated a theoretical motivation for its existence (Anderson 2010).

While the fact of the gravity relationship is well understood, the explanation for its existence is more contentious. As was pointed out by Obstfeld & Rogoff (2001) it is hard to establish a justification for gravity in a frictionless world. The friction almost universally employed to create a gravity relationship has been trade costs. The story is simple; traded goods accrue some marginal cost per unit of distance that they are transported. For modeling convenience this cost is assumed to come in the form of an iceberg cost, in which more than one unit of a good ($\tau(\delta) > 1$) must be shipped in order to provide one unit at distance (δ). A quantity of the good ($\tau - 1$) "melts" in transit, so the cost of producing that extra good represents the "trade cost". But the exact structure of how distance is transformed into trade costs often is elided over (Anderson & van Wincoop 2003 being a rare exception).

One reason could be because of the necessary constraint that faces the gravity modeling exercise. Three essential features form a trilemma: (i) the gravity relationship, (ii) the preference (or supply) structure in the model, and (iii) trade costs. These three interact such that choosing any two pins down the third.[2] For instance, Novy (2013) has elegantly demonstrated how changing

[1] Recent surveys on this topic include Anderson & van Wincoup (2003) and Anderson (2010). Also, note that ϵ has been deeply studied, and in recent years there have been several papers estimating this value at something less than one, but it is presented as it is here because this was the original conception, and the paper shows a strong theoretical reason this should be the case.

[2] It is helpful to note how most models, even supply-driven ones, simplify to an Armington (1968) style gravity structure.

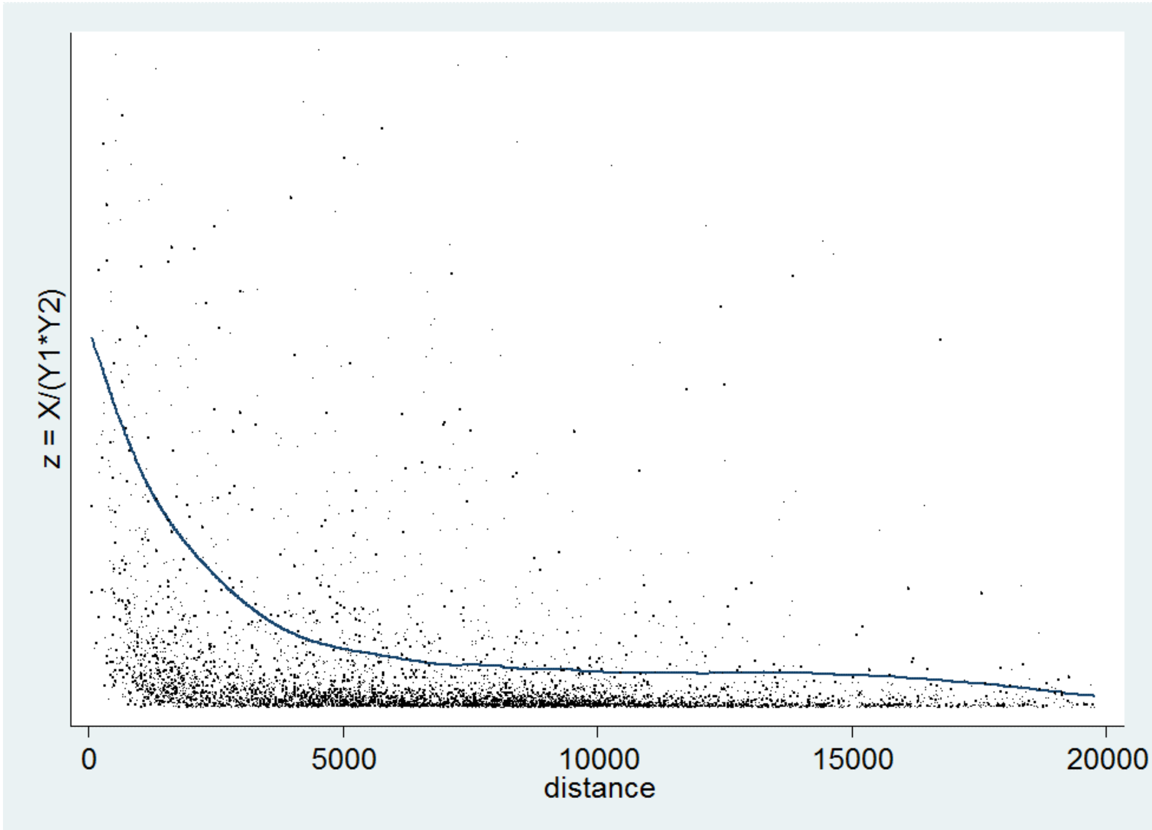

Figure 1: This is a local linear smoothing over the data of $\gamma_{ij} = \frac{X_{ij}}{Y_i Y_j} \approx \frac{\zeta}{\delta_{ij}^{\epsilon}}$. It is superimposed on a scatter plot of the raw data (some extreme values of which, on the left hand side below distance = 500, have been truncated). Note that the Earth is 40,075 km in circumference at its largest, so 20,000km is the approximately the maximum achievable distance.

the choice of utility transforms the implied trade costs. This paper will not move beyond a standard CES-Armington setting, but the implication of this trillema holds. Using Anderson & van Wincoup's (2004) notation, define the price of a traded good as $\tau = \boldsymbol{\beta} * \delta^{\rho}$ where $\tau - 1$ is trade cost, δ is distance, ρ is the elasticity of cost with respect to distance, and $\tau = \boldsymbol{\beta} = (1 + \sum \beta_i)$ is some scalar that includes β_i cost effects (typically including borders, language, et cetera). An example of this trilemma is the well noted relationship in the Armington model between price growth (ρ), the trade elasticity (σ), and the gravity model distance exponent (ϵ): $\rho(1 - \sigma) = \epsilon \approx 1$.

For derivations of this, see Arkolakis et al (2012), Deardorff (1998), Dixon et al (2016).

1.2 Observed Trade Costs are Small and Linear

Before proceeding, it is important to pause here and discuss two things. First, it is helpful to note that observed trade costs in the data on freight rates, et cetera, are actually quite small relative to the values estimated from gravity equations. Second they are increasing at roughly a constant marginal rate per unit of distance when one examines shipments to non-bordering countries. In contrast, for purposes of gravity modeling, trade costs are usually modeled as increasing logrithmically at a rate consistent with a $\rho \approx 0.3$, as in Hummels (1999) and Anderson & van Wincoop (2003). However, when looking at countries that are not close, the data suggest that marginal cost of distance over the ocean is nearly constant (which justifies how costs will be modeled later in this paper). To show both the structure and the small size of observed trade costs, consider the following figures. Using sector-specific data of US imports that include a detailed summary of trade costs, it is possible to measure the observable physical trade costs experienced by US importers. The data cover 13 years, 1991 to 2003, and include imports from 188 countries divided into sectors by SIC code, altogether over 3,000,000 observations. The data also include transaction values, so it is easy to produce a percentage trade cost of transporting the physical good. This ad valorem trade cost ($\tau_D - 1$) derived from the data implies a data motivated traded-good price (τ_D). Figure 2 displays these plots for several of the most well represented sector categories.

In order to first adjust for endogeneity in purchasing decisions, a semi-parametric local linear smoothing is estimated for each sector (*see* Figure 2). Because the data is from the US, there are important considerations with regard to distance. As demonstrated by Coughlin and Novy (2016), there is a notable spatial component to border effects, and being a large country between two oceans the US has few nearby trading partners. Once one looks further from the borders, the increase in marginal cost is very nearly linear. To tell a simple story, this is about the distance at which all trade is by ocean and the marginal cost of transporting goods one more kilometer on the ocean is more or less a very small fuel cost. But the most important insight from these is to note how remarkable small the costs are. At the most extreme distance, the estimated trade cost is about 10% of the value of the good. This is extremely small relative to that implied by the standard models.[3]

Finally, crude regressions are run on the data to provide rough values of the growth parameters

[3]Note that there are many other trade costs in used in these models, such as tariffs, but those are not correlated with distance and therefore will not contribute to generating gravity.

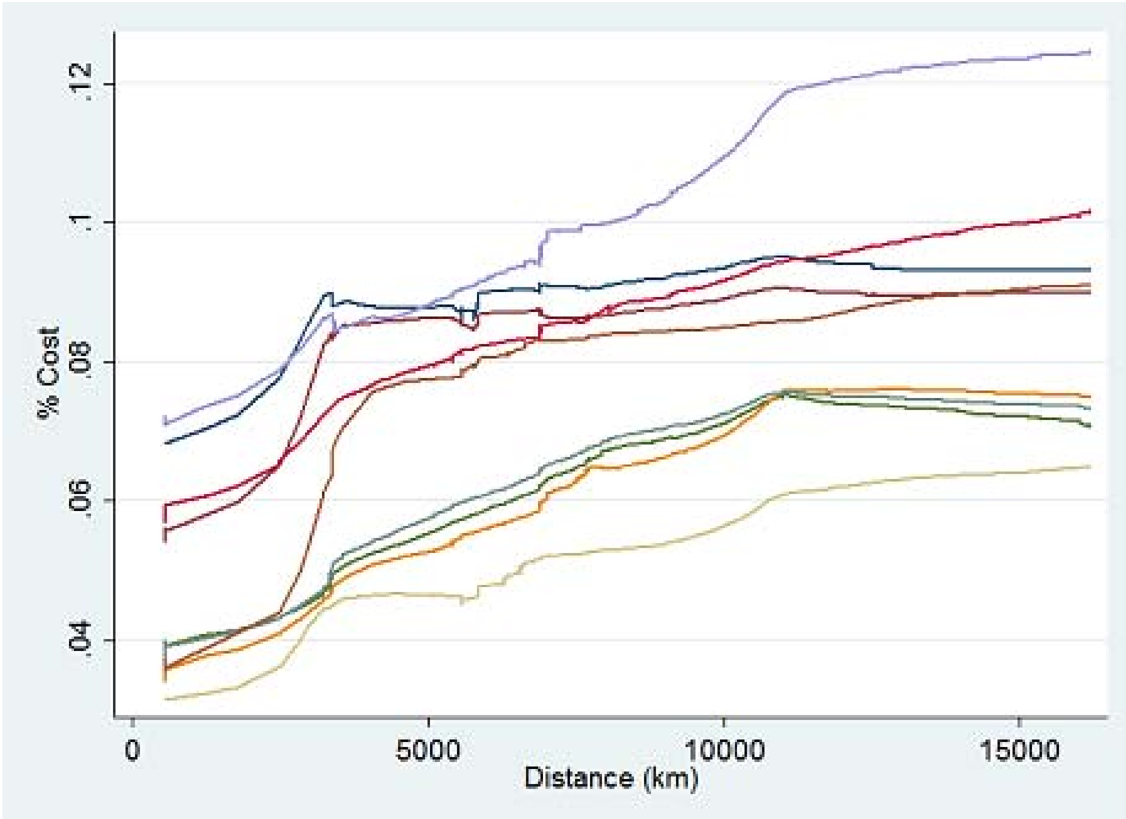

Figure 2: a local linear smoothing of observed trade costs in several sectors. The cost values are given as percentage of the total cost of the transported goods. Sectors represented were those most prevalent in the data, including fruit, sea food, textiles, children's toys, and auto-parts. Note that costs over short distances are harder to fit to a simple trend.

of trade costs (*see* Table 1):

$$T_{tki} = \beta_\delta * d_i + \sum \beta_t + \varepsilon_{tki}$$

Here, k enumerates each good, i the country from which the US imports, and t the year where β_t is a fixed effect for each year. In the first column of Table 1, T_{tki} is the ad valorem trade cost (also referred to as $\tau_D - 1$), and d_i is the distance (δ) in kilometers of country i. In this setting, $\beta_\delta = 0.00000289$ represents the constant marginal cost as a percentage of a goods value for shipping it one kilometer further. In the second column of Table 1, T_{tki} is logged ad valorem trade cost, $ln(\tau_D - 1)$, and d_i is logged distance, $ln(\delta)$, implying this $\beta_\delta = 0.319$ is the growth elasticity of trade costs. This is consistent with what has been found in the literature (Hummels 1999, among others).

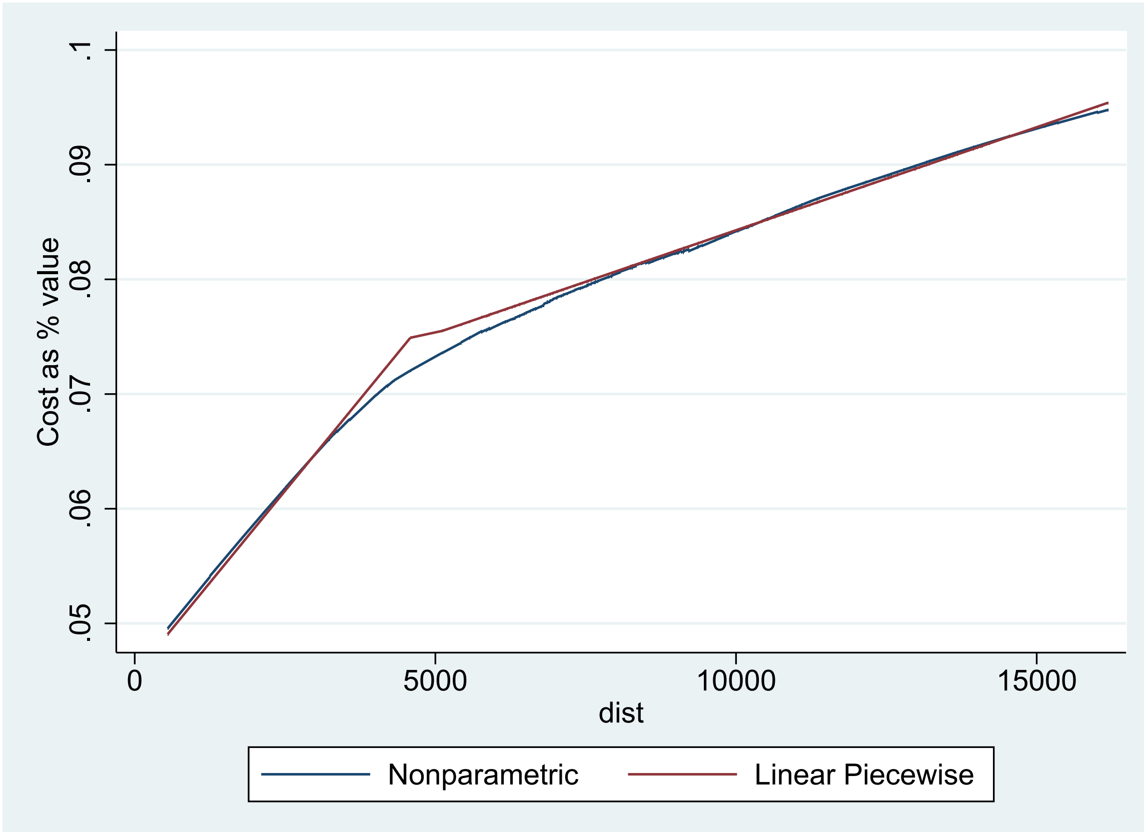

Figure 3: a local linear smoothing of the smoothed trends in all the sectors examined in the data. Again, the cost values are given as percentage of the total cost of the transported goods. The piecewise model imposes a "kink" at 5,000 km justified on the basis that the continental US 4,400 km wide, so only goods traveling from a foreign country to the US market over greater distances are likely to all be transported by the same method to the same port of entry. The fit is at least consistent with the notion that marginal costs for sea transport are lower and reasonably constant.

Both coefficients are similarly and highly significant, though the second specification has a more favorable R^2.[4]

1.3 "Implied" Costs Grow with Distance

The gravity relationship is very well studied and easily observed and the choice of functional form for preferences or production is constrained by questions of tractability. Therefore, the structure of trade costs is usually the feature of the trilemma that has to give. So it is unsurprising that existing

[4]Though not explored here, it is interesting to consider how running the model in logs changes the importance of different observations. In a variable like $\tau_D - 1$ which are usually less than one, values approaching zero are mapped to values of larger and larger absolute (in this case negative) size, so fitting observations near the origin is far more important. But observations near the origin are those least reliable in the data. So note how much the R^2 changes in the third specification when the same logged regression is run on the same logged regression *away* from the origin, $ln(\tau_D)$.

trade models tend to predict costs that are not consistent with observed measurements, but rather larger ones "implied" by the data. For instance Balistreri & Hillberry (2006) found that, for typical parameter estimates, this logic would imply that 50% of traded goods (or equivalent value) melt in transit.[5]

There are two ways in which these implied trade costs are troublesome. First is the fact, that multiplicative "implied" costs grow with distance in the Armington setting. Using the trends observed in the data, consider a form for trade costs $\tau = 1 + \beta_D * \delta^\rho$ where $\beta_D = 0.0044$ and $\rho = 0.319$ (from Specification 2 in Table 1). This trade cost ($\tau - 1$) is plotted in Figure 4 with a **blue** line over the relevant distances ($0 - 20,000km$). In most specifications of fitted models, a nontrivial "implied" cost (β_I) must be added into $\boldsymbol{\beta}$ in order to explain home-bias, among other things, because β_D is far too small to explain the effect. This discussion will use simulated values built on a small estimate of the cost leap for goods crossing any border: $\beta_I = 0.1$ (note that this value is actually smaller than typical border effect estimates).

There is no particularly intuitive story why this border effect implied cost should combine with our observed cost in a *multiplicative* way. The casual, intuitive explanation of this cost leap is that it is paid once, at the border, and should not vary beyond the border. This would suggest the implied cost is added to the observed cost, giving $\tau = 1 + \beta_D * \delta^\rho + \beta_I$ (this $\tau - 1$ is shown in Figure 1 with the **red** line). But of course, in the standard Anderson & van Wincoup setting, this value *is* included multiplicatively, meaning trade costs are $\tau = \boldsymbol{\beta} * \delta^\rho \approx 1 + (\beta_D + \beta_I) * \delta^\rho$ (this $\tau - 1$ is shown with the **green** line).[6]

Stated roughly, there is no obvious justification why the distance-growth elasticity ($\rho = 0.319$)[7] of observed costs in the data (β_D) should describe the distance-growth of implied costs (β_I). In fact, it's not clear that β_I should grow with distance at all. If we take the assumptions of the Armington model as given and the implied costs as correctly estimated, the area between the red and green lines represents "implied" costs that are *growing* in distance that are *not* justified. At least not without a compelling story about why, for instance, the home-bias effect must increase

[5]For a few papers that have estimated strikingly large trade costs, see Hummels (1999), Chen & Novy (2011), and Costinot & Clare (2013).

[6]Anderson & van Wincoup's specification is not of this form but I am, at this stage, using a different ρ as will be discussed below. The graphed equation motivates the same issue in the modeling while being easier to look at.

[7]This is the growth rate of costs, not τ_D as will be discussed below.

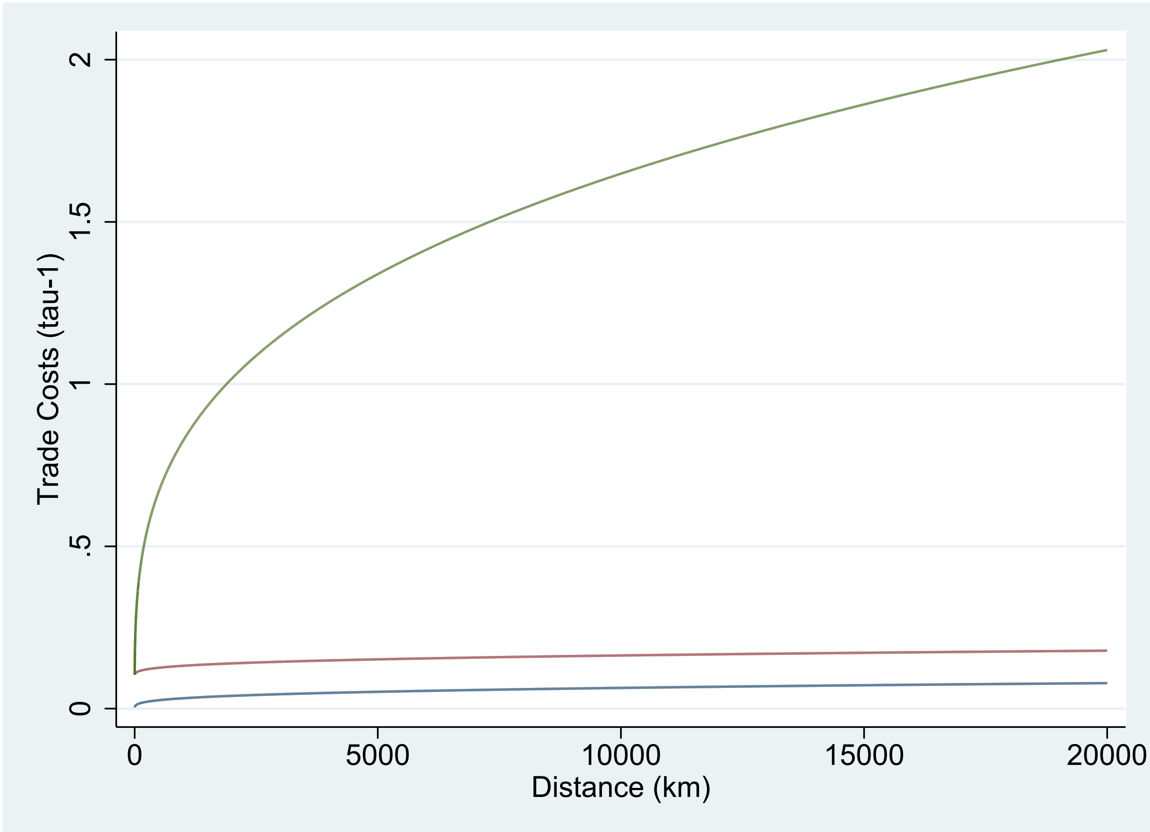

Figure 4: comparing multiplicative versus additive model-"implied" trade costs ($\tau - 1$). The **blue** line gives the actual growth trend of shipping costs in the data. The **red** line gives the same costs with a small implied cost added as a level. The **green** line gives the same additional implied cost multiplied into $\tau - 1$. The area between the red and green lines represents model-implied costs that *increase* with distance, as typical in most Armington-related models.

with distance at the same rate as shipping costs.[8]

In order to explain these implied distance-variant costs that are not easily observed, there have been several alternative strategies. One is to introduce information frictions like in Rauch (1998) and Allen (2014). And a search-style effect is corroborated by the observation of network effects, such as in Rauch & Trindade (2002) or Egger et al (2015). One striking result, which will be referred to again later, is that the distance coefficient falls enormously (65%) for eBay purchases, in a setting where search is not performed in physical space (Lendle et al (2015)).

[8] If more intuitive, unvarying "implied" costs were instead used, there would still be tricky implications. If implied costs were included additively, then the elasticity of trade cost growth would have a different exponent (ρ'): $\tau - 1 = \beta_D * \delta^\rho + \beta_I = \beta' * \delta^{\rho'}$. The larger implied costs β_I grow, the smaller ρ' would become. For instance, using the values of β_D, ρ, and β_I from our simulation, that would imply $\rho' = 0.09$. So even "invariant" model-generated costs would produce parameters that cannot be taken directly from cost data.

1.4 "Implied" Costs are Very Large Compared to Observed Costs

The second way in which implied trade costs are troublesome is that they are strikingly enormous. Consider the equation relating cost growth to gravity: $\rho(1-\sigma) = \epsilon \approx 1$. This elegant relationship derives from how prices feed into the Armington CES demand function. But note that this is a statement about price, not trade costs. So when considering the parameter values derived from the data, the correct value of ρ is not that of observed trade costs ($\tau_D - 1$).[9] Instead it the relevant growth elasticity is that of total observed costs (τ_D), which is estimated in Specification 3 of Table 1 as $\rho = 0.015$. Substituting this value into the gravity-cost relation, if $\epsilon = 1$ and $\rho = 0.015$ this implies $\sigma \approx 68$, which is far outside the realm of estimated values of trade elasticity. In order to achieve even an upper-bound plausible estimate of trade elasticity from the literature, say $\sigma = 6$, ρ would need to be more than ten times larger than what is found in observed costs.

Stated in less theoretical terms, the fitted value from the data for the trade cost of shipping a good one km is 0.4% of its value. The fitted cost for shipping it 20,000 km is 8.2% of its value. That may be a 20-fold increase in cost over that distance, but it is only a 7.8% maximum increase in price due to shipping costs. The trade costs implied by almost any standard fitted model suggest maximum-distance trade costs many times the value of the good, which is an order of magnitude larger than observed shipping costs. Therefore, the trade cost story of gravity is nearly entirely dependent on unobserved costs implied by trade models. And because the costs are unobserved, it is very hard to produce a test of the validity of the framework at large, let alone the constituent parts of a given model.

2 Motivation

The gravity relationship in trade is simple and powerful, but any explanation of it is nearly perfectly theoretical. The intent of this paper is to produce a novel explanation and provide testable implications to validate or reject that model. A guiding inspiration for this alternate framework will be to look at gravity in another setting. The gravity model of trade is so called because it so resembles the equation describing physical gravity, which states that the force between two bodies

[9] There has been some ambiguity in the literature on this point. The growth elasticity of $\tau_D - 1$ is approximately $\rho = 0.3$ which, if incorrectly substituted into the cost-gravity equation is consistent with a trade elasticity of $\sigma \approx 4$ which looks tantalizingly plausible, though invalid.

is proportional to the product of their masses divided by the *square* of their distance (in contrast to the trade setting in which the divisor is distance to the first power). In subsequent (and unresolved) debates among physicists the inverse square portion of the equation has been noted as interesting because this is the same functional form as the rate at which a light source dims over distance. To see why light dims at this rate, simply consider the surface of a sphere ($A = 4\pi r^2$) around a light source. As the radius of the sphere increases, the same amount of light will be spread over an area increasing quadraticaly, so the amount of light at any point will *decrease* at an inverse square rate. This, however, is in three dimensional space. If we restate the problem in only two dimensions, something radiated in all directions (within a plane) from a central point would dissipate as it traveled outward on the circumference of an expanding circle ($C = 2\pi r$). So while a flashlight in 3-space would dim at an inverse square rate with distance, a flashlight in 2-space would only dim at the inverse of distance. Therefore, the simple observation at the center of this paper is that a gravity-style relationship between objects in a two-dimensional physical space can be generated by *any* model in which the "force"—or in this case trade—*radiates blindly from a source point*.[10]

2.1 Gravity and Distance

To provide intuition for the process to be used in the final model, let us first consider two example search processes. In the first case, consider an individual trying to pick from a finite set of objects of uniform shape (for simplicity, circles) and size (radius $\frac{\varepsilon}{2}$) from the center of a large, flat, dark room using only a laser pointer (see the left hand side Figure 5 below).[11] If the agent is blindly trying to find these objects by casting a laser beam into the dark, this can be thought of as the agent simply choosing a random azimuth (θ) for the beam to be cast in. To consider the likelihood of an object being found by this method, the probability of discovery is merely the likelihood of some azimuth being selected that shines on the object in question. If the azimuth is chosen randomly we can describe it as a uniform random variable ($\theta \sim U[0,2\pi]$).[12] If a given object A, located at distance

[10]It is important here to note an excellent 2016 paper by Ferdinand Rauche who discusses this insight in his recent publication "The Geometry of the Distance Coefficient in Gravity Equations in International Trade". We were unaware of each other's work until recently, but he explores this first geometric notion elegantly and it is recommended as a complimentary discussion of this foundational issue.

[11]At this point, the shape of the room is not important, but for the time being we will think of it as square, as we will later be examining how the model holds for arbitrary numbers of objects drawn randomly from a Cartesian uniform distribution.

[12]Subsequently there will be discussion of the ways in which the uniform distribution differs in polar versus Cartesian coordinates. Please note that this variable is being drawn from one dimension where no such distinction is necessary.

d_A, is not in the "shadow" of any other as seen from the origin, the likelihood of discovery is the portion of the two-dimensional "horizon" it occupies as seen by the searcher. The portion of the horizon occupied is (roughly) the circle's diameter (ε) divided by the measure of the horizon at the circle's distance ($2\pi d_A$).[13] Thus the general likelihood of discovery for an "unobstructed" object in a single search is noted to be $\frac{1}{2\pi}\frac{\varepsilon}{d}$. This only considers the case that the object is unobstructed, but the size of the objects was not specified at the outset, or important to the functional form of the relative likelihood of finding objects. So if the objects were distributed in a uniform random fashion, there will exist some ε sufficiently small so as to make all objects unobstructed. To clarify, consider the case in which N objects are selected from a random uniform distribution in Cartesian space inside the room ($f_C(x,y) = 1/(2D)^2$) where D is the distance from the center to the edge of the room.[14] Because the probability of any two points falling on exactly the same ray from the middle of the room is zero in continuous space, there must exist *some* choice of ε so that all the objects are fully visible. In this setting, the probability of finding any given object continues to take the same form as above, but no assumption about the arrangement of the objects in the room has been necessary. This insight will be used later when we establish an analogous search process for searching over a continuum of points.

The second, ultimately equivalent example will be to consider the same unfortunate searching in the dark, but now they are lucky enough to be searching with a flashlight instead of a laser pointer (see the right hand side Figure 5). The searcher will again choose a random azimuth to shine the flashlight ($\theta \sim U[0,2\pi]$), and for every possible azimuth choice the light cast in this direction has a fixed aperture (θ_F). Now it is possible to illuminate multiple objects at once, so the individual will select between the set of illuminated objects in proportion to how brightly they are illuminated. Consider a choice of azimuth so that only objects B and C located at distances d_B and d_C are fully illuminated at the same time. The brightness of the objects, as seen by the searcher, will be a function of the fraction of the flashlight beam that is striking them. This is a function of how much of the two-dimensional "horizon" they occupy ($\frac{1}{2\pi}\frac{\varepsilon}{d}$).[15] This measure will be referred

[13] It should be noted that the portion of the horizon occupied by the circle is slightly less than the diameter. But as distance increases and object size decreases this difference approaches zero. Since these are the exact circumstances for which the model will be considered in going forward, the issue is ignored for simplicity.

[14] Specifically $(x,y) \sim U[(-D,-D),(D,D)]$.

[15] To clarify, this search is taking place in two dimensional space, so all the light "rays" exist only in the plane. As mentioned above this is not how a light source actually dissipates in three-space.

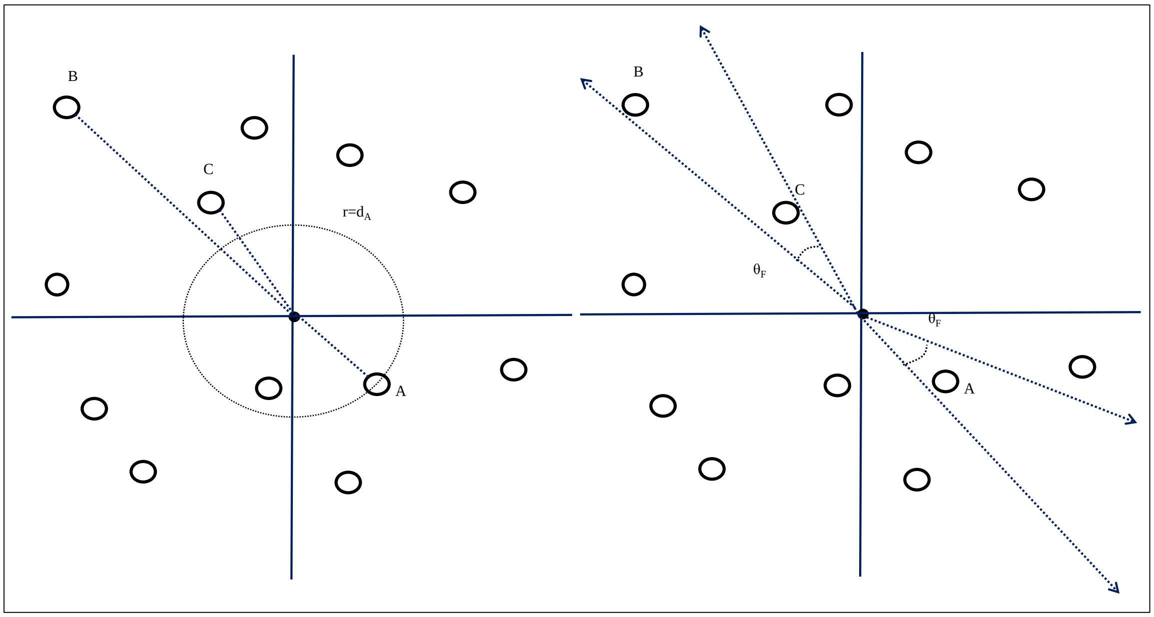

Figure 5: On the left hand side is illustrated the "laser pointer" version of the model in which an agent searches for objects in a dark room by firing it along random azimuths. The two dimensional "horizon" that A sits upon is illustrated. The right hand side illustrates the "flashlight" version of this same problem.

to as an object's "arcwidth" ($A_x = \frac{1}{2\pi}\frac{\varepsilon_x}{d_x}$) going forward. Therefore, conditional on the two objects being illuminated, the selection probability of selecting a given object is simply the ratio of that object's arcwidth and the total arcwidth of the illuminated objects, in this case $p^I_{BC} = \frac{A_B}{A_B+A_C}$. (Empty portions of the searchers horizon are ignored, because the searcher is not seeking them.) An object's likelihood of being fully illuminated by the searcher's choice of azimuth is a function of its size and the aperture of the flashlight given by $\frac{\theta_F - A_B}{2\pi} = p^{\theta}_B$.[16] Generalizing these processes to an arbitrary number of objects, we can increase the number (N) of randomly distributed objects as in the previous example. In this case it is possible to show that the probability of selecting random object x will converge towards $p_x = \frac{1}{2\pi g(N)}\frac{\varepsilon_x}{2\pi d_x}$ for some function $g(N)$. The key result is that the functional form follows gravity.

This second model is much more complex than the first, but it allows us to visualize how "rays" or "beams" of search expand and dissipate over distance. This will be useful when we extend these results to the surface of a sphere. But first we will take the intuition of these simple models, and

[16]In this example, the case of "partial" illuminations will be ignored, because as the number of objects increases and the size of the objects decreases this likelihood will approach zero.

extend it to searches over all the points of continuous space ($\mathbb{R}^2$) rather than discrete objects.

2.2 Gravity and GDP

The toy models of search only explain the denominator of the gravity equation. Just as important is explaining the relationship in the numerator: why should trade be driven by a product of GDPs? In a Sept. 1, 2015 column, Paul Krugman was discussed his general thoughts on the gravity equation. In describing his own intuition he said:

> Think about two cities with the same per capita GDP. They will trade if residents of city A find things being sold by residents of city B that they want, and vice versa.
> So whats the probability that an A resident will find a B resident with something he or she wants? Applying what one of my old teachers used to call the principle of insignificant reason, a good first guess would be that this probability is proportional to the number of potential sellers Bs population.
> And how many such desirous buyers will there be? Again applying insignificant reason, a good guess is that its proportional to the number of potential buyers As population.
> So other things equal we would expect exports from B to A to be proportional to the product of their populations.

This is, in fact, nearly identical to the rationale that will be used in this paper. The basic assumption will be that the number of buyers and sellers, populations, GDP, *and* area of a country will first-order be the same.

3 The Planar Model

3.1 Model Economy

We hope to construct a model that draws attention to the general insight that the gravity equation can be a characteristic of physical space perceived in radial coordinates, rather than something constructed ad hoc for the search model in this paper. In order to do this, all of the choices (production, utility, et cetera) going forward in this model will be out of a desire for simplicity. More complicated variations will be alluded to, but this paper attempts to prove the concept in only the most essential form.[17]

[17] I intend to examine more complex variations in subsequent work.

In this model, each point in the two-space continuum is an agent (x,y). These sellers search for buyers by making randomized search in polar coordinates. (It is not critical to the results if buyers search for sellers or the reverse, as long as only one side of the transaction searches. For the purposes of this paper, searchers are sellers.) Each point on the continuum is both a buyer and a seller, and each has an identical endowment of one unit of their good type, unique to each point on the continuum. The utility for each consumer is also extremely simple:

$$\mathbb{U}_{(x,y)}\left\{c_{(x_i,y_j)}^{(x,y)}\right\} = max\left\{c_{(x_i,y_j)}^{(x,y)}\right\}$$

where $c_{(x_i,y_j)}^{(x,y)}$ is the consumption by agent (x,y) of the good produced by agent (x_i,y_j)[18] and agents cannot consume their own good type. Transport costs are nil, and in equilibrium all points should be paired; therefore the model achieves equilibrium when all sellers part with their good at the identical global price (because all goods are perfect substitutes and there are no trade costs) and use their revenue to buy one unit of the good sold by the seller who paired with them.

The search process itself is similarly simple. The seller searches in polar coordinates, and their pairing is created probabilistically based on an independently uniformly distributed azimuth ($\theta \sim U[0,2\pi]$) and distance ($\delta \sim U[0,D]$). In other words, and in keeping with the notion of "radiating" trade, the agent picks a direction and distance at random. If the target agent is already paired then the seller searches again. It is important to note that all searchers search over identical spaces, and that the draws of azimuth and distance are not serially correlated.[19] Therefore the points removed from the search space on each iteration do not alter the shape or descriptive statistics of the probability space for subsequent iterations, and the search can be repeated however many times is needed to provide all sellers with a match.[20]

[18] The utility and market clearing can be made much more complicated and the model will still hold. But such additions are not necessary for the key results here and will be examined in subsequent work.

[19] The timing of the search process can produce an interesting math problem, but one that is also ultimately uninformative. A simple variation to consider in lieu of simultaneous search is to have each agent make a separate random uniform draw at the start of each round of search to determine if they will be "active" (search as a seller only) or "passive" (wait to be found as a buyer only). Because search efficiency is not relevant to this paper, the extra stage in the search process is excluded. However, it is worthwhile to note how this could be useful for more complex utilities, sales, and market clearing.

[20] It should be noted here that, while it is obvious that this process should be 1-to-1 (all sellers find buyers), in an uncountable space it is not as clear why the process should be onto (all buyers are paired with sellers). In fact, it is true that an infinitesimally small (Lebesgue measure zero) set of buyers will go unpaired because they are sought stochastically. If we desired we could exclude these agents' sales from the model because they now lack the purchase necessary for market clearing, then their buyers' sales could be excluded and so on. But we can ignore this for three reasons: (i) after some number of "chasing market failure" iterations the set will close and a dense coverage of points with the desired characteristics would exist, (ii) the model could, with minor alteration, be infinitely repeated so that agents could "save" money across

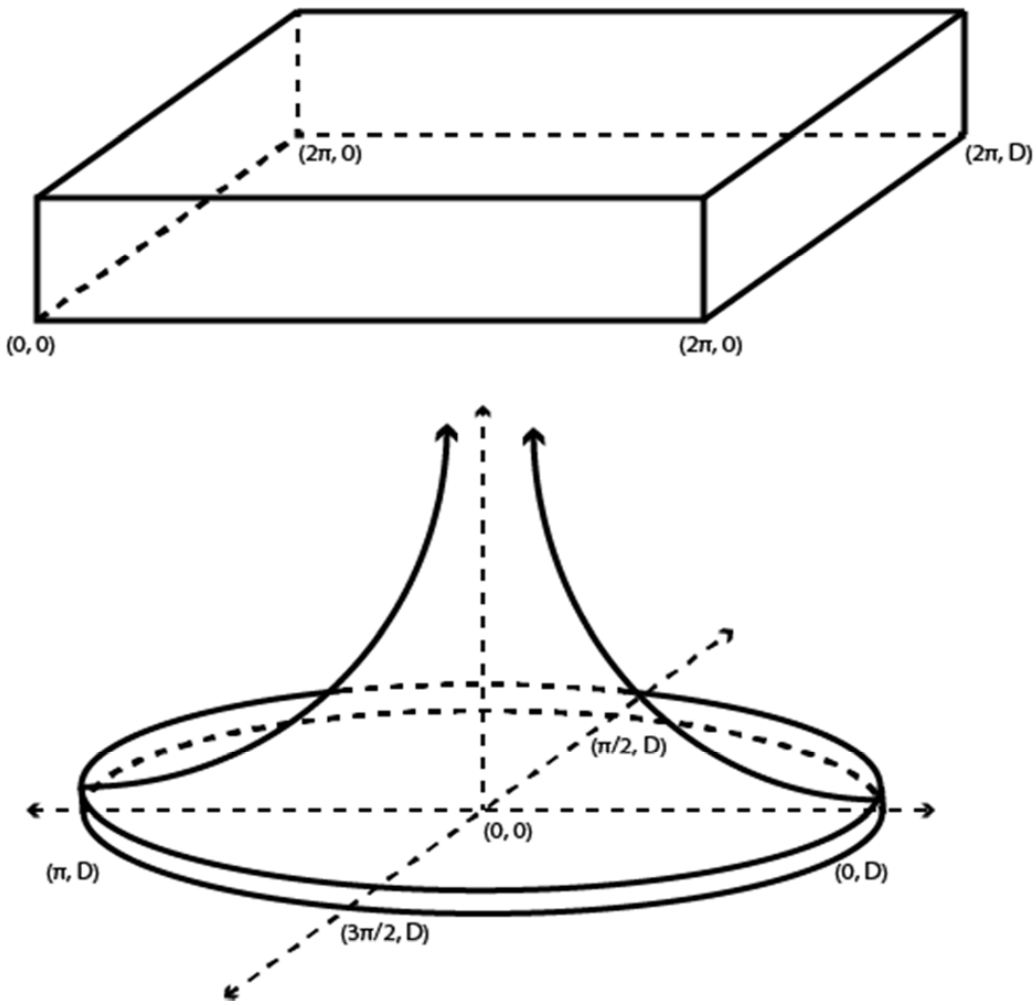

Figure 6: the visualization of the transformation of the pdf of a Cartesian uniform distribution into polar coordinates. Note that the z-axis is not to the same scale in both graphs.

This process produces a network of sales between agents across the continuum. The structure of these pairings, as the search space is (in probability) identical for all agents from their location, is distributed the same for each agent. To intuit the pairing probability that this implies, consider the visualization of a uniform two variable pdf in Cartesian space: a rectangular prism of uniform

periods using fiat money, or (iii) if market clearing was defined in terms of integration over ε-neighborhoods the missing buyers would become irrelevant. In all three cases we will end up with the same structure for the pairing function and relevant characteristics to the equilibrium while imposing stronger assumptions and increasing mathematical complexity. Therefore, we will simply ignore the "dust" of unpaired buyers.Also note that while the distance variable has a finite bound, but the search plane as we have articulated it at this point does not. Therefore, in the planar case agents do not search over the entire space. This will change in the spherical case. In the meantime, note that even if every agent does not search the full space, every point in space is searched by the same number of agents. This means the assumptions about all points having equal likelihood of selection on each iteration hold.

height (see Figure 6). When translated into polar coordinates, this prism is transformed by compressing all the points on the edge corresponding to distance zero into a single point at the origin. This produces a mass point, and an instantaneous pdf value of infinity at the origin.[21] Similarly, the opposite edge of the prism corresponding to a distance of D is stretched around the origin reducing its height. From this simple picture, it is easy to see that a uniform Cartesian distribution is *not* uniform in polar coordinates.[22]

More formally, consider a random variable for polar coordinates ($R = (\delta, \theta)$) that is distributed random uniform when plotted in Cartesian space ($f_P(R) = \frac{1}{2\pi D}$, which is the pre-transformation rectangular prism in Figure 6). Now we consider what this variable would look like seen in polar space. But, since in basic definitions and concepts are not preserved in polar coordinates[23], we will instead plot this random variable in polar space and transform it into Cartesian coordinates ($X = (x, y)$) with a new pdf ($f_C(X)$). To do this, we need only make a standard transformation of random variable (the general form for this is $f_C(X) = f_P(R) \left|J_{g^{-1}}\right|$). Let $g : R \to X$ be the function that changes polar coordinates into Cartesian ones.[24] In order to perform a change of random variable, we will then need to consider $g^{-1} : X \to R$ where $g^{-1}(X) = (\sqrt{x^2 + y^2}, arctan(\frac{x}{y})) = R$. The determinant of the Jacobean (the matrix of partial derivatives) with respect to X is

$$\left|J_{g^{-1}}\right| = \left|\frac{\partial g^{-1}}{\partial X}\right| = \begin{vmatrix} \frac{x}{\sqrt{x^2+y^2}} & \frac{1}{1+(\frac{x}{y})^2}\left(\frac{-y}{x^2}\right) \\ \frac{y}{\sqrt{x^2+y^2}} & \frac{1}{1+(\frac{x}{y})^2}\left(\frac{1}{x}\right) \end{vmatrix} = \frac{1 + (\frac{x}{y})^2}{\sqrt{x^2+y^2}(1 + (\frac{x}{y})^2)} = \frac{1}{\sqrt{x^2+y^2}}$$

Note that this is simply $1/\delta$. This is a key observation of the paper. The fact that pairing likelihood falls with distance in this case is not a clever or obscure artifact of the choice of $f_P(R)$, but is in

[21]While this is not intuitive, it does not change the well-defined nature of the transformed pdf for the same reason that the integral of $1/x$ is $ln(x)$. Also, none of the results in this paper examine the case that distance equals zero, because agents do not pair with themselves.

[22]Probability in polar coordinates lacks several intuitive characteristics of probability on the plane, so it is necessary to define what a "uniform" distribution is in this setting. For our purposes, a distribution is "uniform" if the cumulative probability over any two regions of equal size in the domain of equal area are equal to each other. The underlying issue is that points in polar space are more "densely" packed around the origin (which is also the driving observation of this paper). So even if the function we are evaluating ($f_P(R)$) has a constant value across all coordinates, its integral over different regions of equal area could be dramatically different.

[23]Among other things, integrating with respect to both variables and integrating over area are different operations in polar coordinates. This is because in two dimensional Cartesian integration, $\int f(x,y)\ dA = \int f(x,y) dx dy$. But in polar coordinates the area term is $dA = r\ dr\ d\theta$. As a consequence, in this setting a pdf and a cdf do not have the usual mathematical relationships to one another.

[24]The transformation of a random variable does not require us to use this function, only its inverse. But the function in question is $g(R) = (r\ cos(\theta), r\ sin(\theta)) = (X)$ which itself has a Jacobean whose determinant is simply $\left|J_g\right| = r$.

fact derived from viewing the search space in polar coordinates. So any choice of pdf for R will have to be transformed by a gravity-style expression, and a many choices of $f_P(R)$ will produce a gravity-style pairing outcome.[25] In our specific example, the transformed pdf ($f_C(X)$) is of the form

$$f_C(X) = f_P(R)\left|J_{g^{-1}}\right| = \frac{1}{2\pi D}\frac{1}{\sqrt{x^2+y^2}} = \frac{1}{2\pi D}\frac{1}{\delta}$$

in Cartesian space. From this, we can see that, conditional on any choice of θ, the transformed pairing probability in the plane is of the form

$$p_P(\delta) = \frac{1}{2\pi D}\frac{1}{\delta}\ for\ 0 \leq \delta \leq D$$

which holds for *any* choice of θ.

At this juncture, it is necessary to address the issue of the leading coefficient for the pairing likelihood equation. It seems appropriate that the integral of this pdf, over the full surface of the planar disk, should equal one. Unfortunately, the actual globe (and the data) are not uniformly covered with potential agents. In fact, most of the globe is in fact "empty" (at least for trade purposes) space. Therefore the exact leading coefficient of the pairing likelihood has the problem of being (i) very convoluted to compute empirically and (ii) different for every country on the planet (consider Fiji versus Austria). To deal with this issue, rather than encumber the model with false precision, the exact coefficient will not be discussed going forward. As will be seen, none of the following analysis will depend on this value (aside from noting that it is in all cases positive) and the reference (gravity through trade cost) model makes no predictions on this point for comparison.[26] This generality will be explored in more detail when the model is taken to the data.

This pairing probability function, in the planar case, produces trading behavior among countries nearly identical to the standard gravity model. As each point in the plane is an agent that

[25] The realm of options for $f_P(R)$ that still produce an ultimate gravity-style pairing likelihood is an interesting question I have not yet answered. While it is obvious that any $f_P(\delta,\theta)$ that is constant with respect to δ will produce a gravity style equation, it seems that there are many more functional forms that will produce a probability in distance that is "first-order" equivalent to gravity.

[26] To be more precise, the way in which an agent searches over a subset of space does not affect the gravity result as long as underlying search space of δ,θ is uniform when plotted in Cartesian coordinates, or "piecewise uniform" in that it could permissibly be several different rectangular prisms instead of one.

produces and consumes perfectly substitutable, fungible goods, a country would simply be some contiguous shape drawn on the plane. The GDP of that country would be the integral of output across all agents, which is simply area in this context. For simplicity, in this discussion we will treat all countries as circular. Therefore, a country can be described as some epsilon neighborhood of area z (ε_z), in which GDP is proportional to area.

Now consider the case that a country is vanishingly small (ε_A), which is reasonable considering that countries are very small as a portion of the surface of the Earth.[27] In this small country setting the pairing likelihoods from all points within the country to all other points approach being identical. This allows us to treat the entirety of such a country as being a continuum of agents located at an individual point on the globe. In this case, the expected number of a country's pairings in terms of distance becomes $f_A(\delta) = \beta_A \frac{\mathrm{A}}{\delta}$ or simply a mass of size A pairing identically and independently with probability $1/\delta$.[28] Therefore, given two small countries, ε_A and ε_B, which lie some distance apart d, the trade flow from ε_A to ε_B should be the pairing likelihood of a mass of size A pairing at distance d with a mass of size B, which is simply $f_{AB}(\delta) = Bf_A(\delta) = \beta_A \frac{\mathrm{AB}}{\delta} = \beta_A X_{AB}$, where X_{AB} is exports from A to B. This demonstrates that, in the context of this model, as country size approaches zero and distance increases away from zero, the trade flows are described by the standard gravity equation.

3.2 Planar Model with Intensive Margin

Having created a simple search model that achieves gravity that is driven solely by the search process, trade costs are conspicuous in their absence. Trade costs that increase in distance do exist whatever their magnitude, and must have some impact on trade flows under any set of preferences. So the next step is to create a model that allows for trade costs to affect demand conditional on a successful search pairing. In order to do this, consumers will need to pick multiple goods and allocate their expenditure among them. Each consumer will now consume a home good (x_H) and a foreign good (x_F). A countrys income will be distributed equally among all citizens to simplify the model. This simplification is reasonable because all of the good specific distance effect is created

[27]The (by far) geographically largest modern country, Russia, occupies a mere 3.2% of the planet's surface. For the 194 countries currently recognized by the United States the average country size is less than 0.15% of the planet's surface.

[28]In a plane that has no vacancies or voids (i.e., a planet with no oceans), $\beta_A = \frac{1}{2\pi D}$, but for the reasons mentioned above we will now start to transition to a more general statement of β.

through the demand function, whereas income effects of distance are spread across all goods. (It is important to note here that in any model that is isomorphic to Armington this is the case, so very little of the result relies on the simplification.) The price of the home good is set as the numeraire and trade costs are modeled as being linear in distance subject to a constant marginal cost (α) and a fixed cost of trade (T). So measured in home price, the price of x_F is $C_F = 1 + \alpha\delta + T$.

For simplicity in this specific case we will impose CES preferences, though for reasons that will be apparent most any selection could be made. So agents maximize

$$\mathbb{U} = \left(x_F^{\frac{\sigma-1}{\sigma}} + x_H^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}$$

The first order conditions yield the requirement that $x_F = C_F^{-\sigma} x_H$, so we are left with a partial equilibrium demand curve $x_F = C_H^{-\sigma} Y_H$ where Y_H is the GDP of the home country.[29] But one infinitesimal agents consumption of a one good type has no discernible impact on shared GDP of the country, so this will ultimately be the optimization for all agents in the country. So the demand function for all agents, conditional on a pairing, is $x_F = (C_F)^{-\sigma} Y$. So now it is possible to create a combined aggregate demand function for the country as a function of distance that combines both the pairing likelihood for a given distance (p_δ) and trade cost effect, scaled by the number of eligible partners in the foreign country (Y_F). This function is:

$$D(\delta, Y_H, Y_F) = p_\delta Y_F (C_F)^{-\sigma} Y_H = \underbrace{\frac{\beta_H}{\delta} Y_F}_{\text{Extensive}} \underbrace{(1 + \alpha\delta + T)^{-\sigma}\, Y_H}_{\text{Intensive}}$$

It is interesting to pause here and note that the two parts of this function actually have an interpretation that can be related to the existing literature. The pairing likelihood component describes the decline in varieties consumed with respect to distance, which would be empirically but not theoretically similar to the extensive margin as it is observed in the data. The demand-driven component would describe the decline in intensity of trade flows conditional on pairing, akin to the intensive margin. Usually when these terms are discussed in the literature they refer to different characteristics of the distribution of firms or consumers, which is not at all how they are

[29]The standard general equilibrium solution of this system usually produces a demand curve with an exponent is $1-\sigma$, as discussed by Head & Mayer (2015). The partial equilibrium setting closes off all wealth effects leading to a simpler solution. However, when estimated the specific form will prove irrelevant, so $1-\sigma$ could just as easily be used.

being mentioned here. But in data analysis, they would look more or less the same; the number of firms participating in markets would decline with distance at one rate, and the participating firms would export less to markets they are present in at another rate.

One interesting implication of this model is that it provides a totally different motive for the existence of this margin that could be studied in greater detail. If this new geometrically motivated extensive margin exists, it would be unaffected by the size of trade costs. Using almost any preferences, this new hybrid framework would, when taken to the data, provide smaller estimates of trade costs in a way more consistent with observed data. But most intriguingly, the observed importance of the extensive margin in this model will in most cases be larger than that of the intensive margin (because the extensive margin is already gravity shaped). This would be consistent with trends that have been found in the data (Hummels & Klenow 2005).[30]

4 The Spherical Model

4.1 Motivation: Planar vs. Spherical

This search behavior produces a model that looks very similar to the iceberg cost model, without any iceberg cost. However, there is a pointed difference that occurs when one considers placing the model on a globe rather than a plane: the radii converge on the opposite side of the world. To refer back to the two intuitive, discrete models, consider the situation that the dark room has in fact become the surface of a sphere, and that the light from the laser pen or flashlight must now follow the two dimensional space along the surface of the sphere. Consider the laser pointer example first. If we assume the searcher is standing at the north pole of the globe, then any light rays from their position would follow lines of longitude over the surface and converge at the South Pole (see Figure 8). This would produce an increase in finding probability for objects located at the South Pole analogous to the increase near the North Pole. The same would occur in the flashlight example if flashlight rays also followed lines of longitude (the beam would get brighter again at the South Pole). In both cases we would be left with the absurd result that the "easiest" object to find would be one located at the opposite pole of the planet. We could instead make flashlight

[30]This role for the distance component of the extensive margin in also deep in most relevant structural models, as discussed by Chaney (2008), Helpman et al (2008), and Bernard et al (2011).

rays that were not bound to lines of longitude and simply spread away from the aperture as if on a plane, so that the arcwidth of the beam was the same as a function of distance.[31] This would allow us to describe a model in which things far away were more “dimly lit”, but it would create new problems because the flashlight “beams” from different azimuths would cover different points until they approached the South Pole and where they could begin to cover some of the same points. So even in the expanding and diming flashlight beam case, the issue again arises near the South Pole due to duplicate coverage across choices of azimuth. In any case, transferring the search processes of the motivating models directly onto a sphere violates their underlying intuition which is that objects “shrink” and are harder to “find” as they are further away.

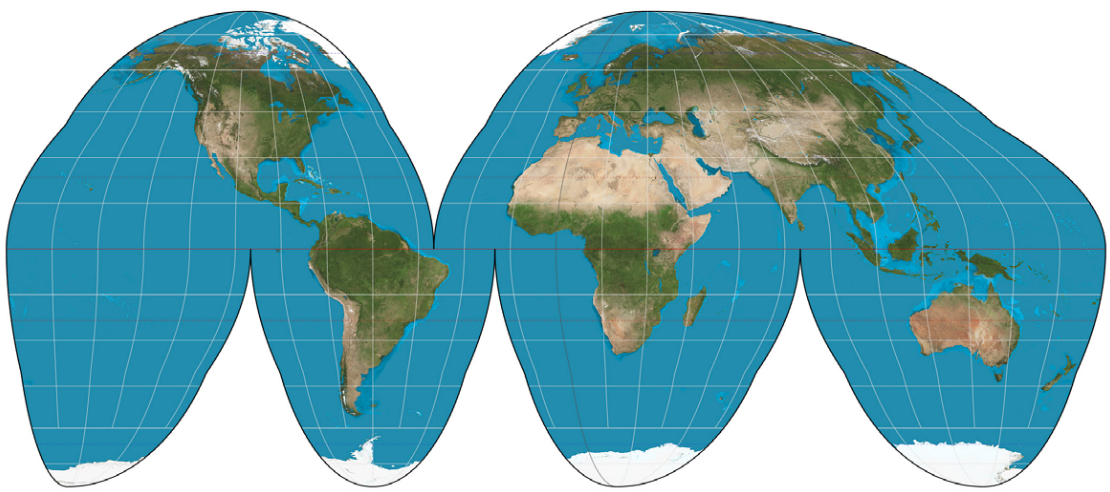

Figure 7: an example of a “Goode Homolosine” projection map. This projection preserves landmass areas better than most, and shows the portion of plane that must be removed in order to represent the globe in two dimensions. *Source*: Wikimedia Commons.

The basic problem with this conversion is the paucity of points to locate an object on a sphere for long distances when compared to a plane. It is the same reason why all map projections of a globe must be achieved by stretching objects near the edges of the projection rather than shrinking them. To consider why this is, plot the circumference of circles (centered at the North Pole) on a globe as a function of distance from the Pole, versus the circumference of concentric circles drawn at the same distance on a plane (see Figure 9).

[31] This is not intuitive, as “rays” no longer travel in straight lines. One way to formalize this process would be to measure the arcwidth of the flashlight beam in the plane for every distance, and then create a beam on the surface of the globe so that the width is the same at every distance. This would mean that the “brightness” would fall in distance, but only the ray in the middle of the beam would travel in a straight line. Those at the edges would need to curve outward.

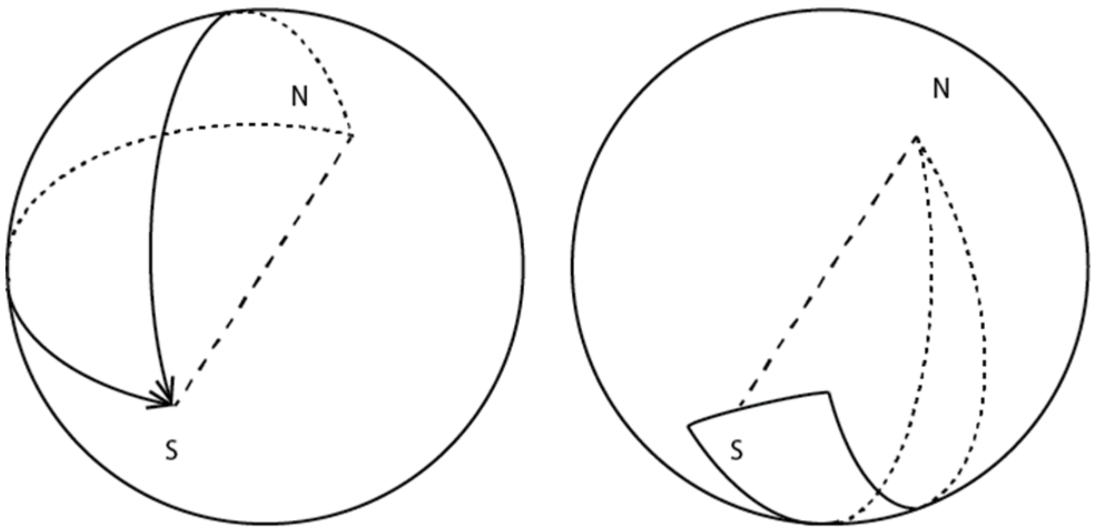

Figure 8: a picture illustrating the two example models transformed into spherical space without adjustment. The "laser pen" example is given at left, the "flashlight" example in which the light dims and disperses over distance as it would on a plane is given at right.

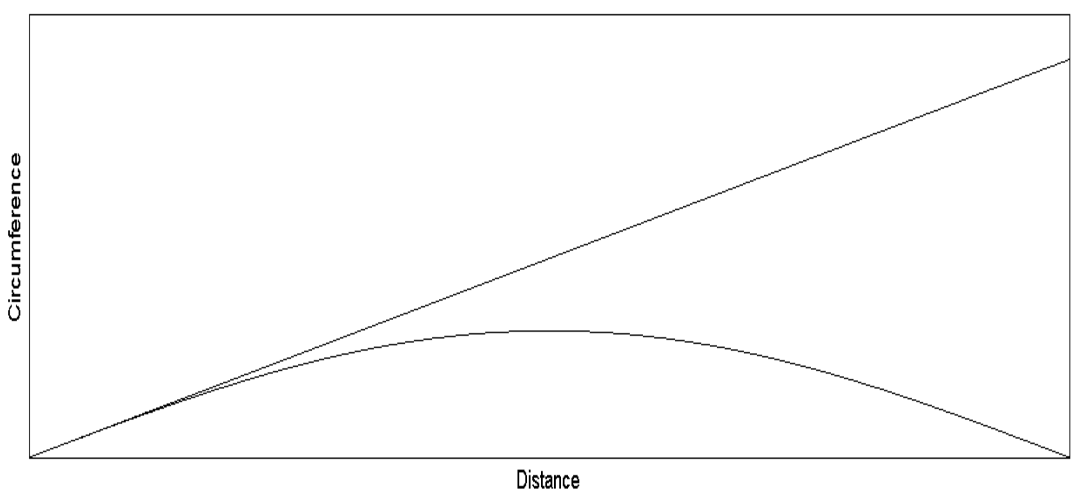

Figure 9: the circumference of a circle in a planar disk and on a sphere as a function of distance from the North Pole. The upper line is for the disk ($C = 2\pi d$) and the lower line is a sine function that will be derived in detail below.

Going back to the motivating models, it is obvious that our problems arise from the way in which lines of longitude return to meet one another on a globe. If we could perform the search in the planar, light dispersing setting, and then translate that result in way isomorphic to searching on a globe, it would be possible to preserve the general intuition of the models. In order to make a plane isomorphic to a globe, objects near the edge of a planar disk would need to grow smaller, in the intuitive inverse of the distortion created by globe to plane projections. So, if we made the objects in the dark room shrink with distance, we could preserve the desired characteristics of the models.[32]

4.2 The Spherical Adjustment

In order to make sense of the continuous model as applied to the sphere, we will make an analogous alteration. As all objects in the continuous case are of the same infinitesimal size, instead of altering their size we will remove surplus points from an agent's search plane so that the search field will be isomorphic to a sphere (see Figure 10). This removal will be made random with respect to azimuth (see the right-hand diagram in Figure 10). Specifically, the integrated "width" of all the candidate search points in the plane for the agent at a given distance δ_0 should be the same as the width of the points falling at δ_0 on the surface of a sphere (the proportion of the circumference that is not black in the left-hand diagram in Figure 10).[33] This reduction in points requires a normalization of the pairing probability to adjust to the reduction in candidates. We will achieve this simply by normalizing for surface area on a globe as a function of distance from a point.

Let $A_P(\delta)$ be the instantaneous change in surface area of a circle drawn on a plane at distance δ from the point O. Let $A_G(\delta)$ be the change in area of a circle on a sphere (or globe). Define the function $p^G(\delta) = A_G(\delta)/A_P(\delta)$ to be the probability of a random point lying at distance δ being included in the search space after being adjusted to match the characteristics of the sphere. The function $p^G(\delta)$ is equal to the ratio of the circumferences of these two circles at distance (δ), which we will define as $C_G(\delta)/C_P(\delta)$. The denominator is obvious enough to compute, but $C_G(\delta)$ requires

[32] As this process is fairly convoluted to do separately for the discrete-object example models, it will not be derived in detail.

[33] There is a minor technical slight-of-hand taking place here. If an agent draws truly random points to be in or out of the search space at every possible distance based on the probability we will establish below, the Lebesgue measure of both the included and excluded sets for a given distance will be equal and equal to the entire 2-dimensional horizon at that distance. This is because the selected space of uncountably many points allows for no countable subcover of less than the total search space. A technically precise alternative would be to divide the horizon into countably many sets of vanishing size, and assign them randomly to be in or out of the search space based on the probability function. As the "integration" definition of the modified search space is only presented for intuitive understanding and we never ultimately are called upon to integrate over this space, the issue is ignored because it only adds complexity.Also note that this random removal of points does not affect the properties that made solving the search model so simple in the planar case. The resulting search space differs for every agent because it is stochastic, but the ex ante search probability functions are identical for every agent. The random removal does not alter that the search process (i) is not serially correlated, (ii) is 1-1, and (iii) is close enough to onto to make unpaired buyers not an issue.

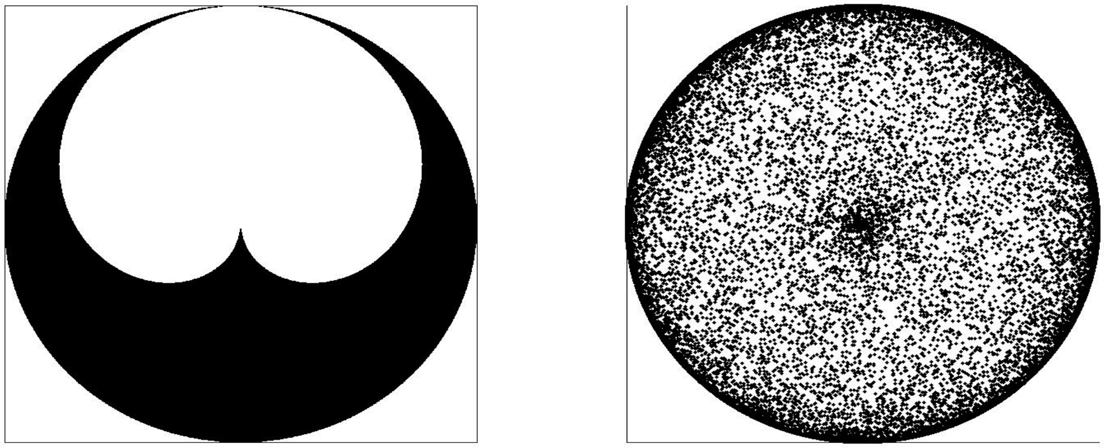

Figure 10: The image at left shows the surface of a sphere with a circumference of 2D laid flat on a disk with diameter 2D. The area in black represents all the points that exist only in the planar disk. The image at right displays approximately the same coverage (subject to graphing software limitations, where the non-zero size of "points" seems to indicate unneeded "darkening" near the origin). In this case the "missing" points on the sphere occur at the same frequency with respect to distance, but are randomized with respect to azimuth.

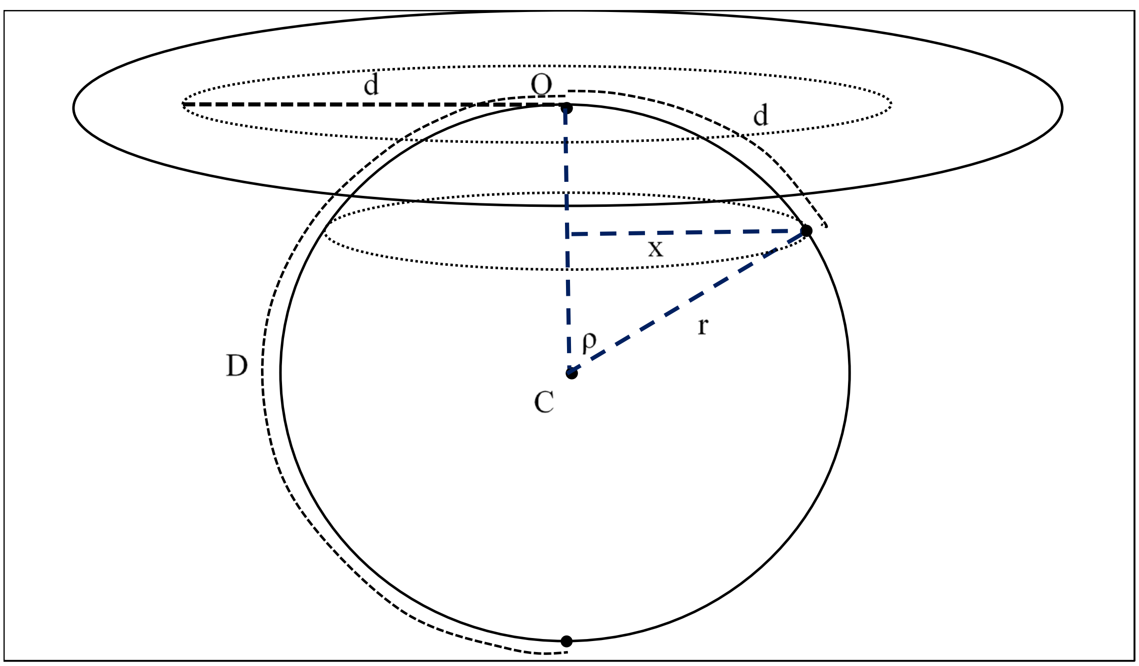

Figure 11: this diagrams the problem of determining circumference on a sphere versus on a disk. The labels and definitions used in the derivations are provided.

some simple trigonometry. As the circumference of the sphere we are examining is 2D (because D is defined to be the maximum distance achievable on the globe), we can deduce that the radius of this sphere is $r = \frac{D}{\pi}$. Define $\rho(d)$ to be the angle between the origin and the edge of the circle of size d as seen from the middle of the sphere, which we can see (in radians) is $\rho(\delta) = \frac{\pi\delta}{D}$. We can construct a right triangle (seen in Figure 11 as having edges x and r and angle ρ) from the middle of the sphere to the edge of the circle at distance δ to the line between the origin and the middle of the sphere. By the definition of the sine function, we can determine that the opposite edge (between the ray and circle) is of length $x = r\ sin(\rho) = \frac{D}{\pi} sin(\frac{\pi\delta}{D})$. Therefore

$$C_G(\delta) = 2\pi x = 2D\ sin(\pi\frac{\delta}{D})$$

and

$$p^G(\delta) = A_G(\delta)/A_P(\delta) = 2D\ sin(\frac{\delta}{2D})/2\pi\delta = \frac{D}{\pi}\frac{sin(\frac{\pi}{D}\delta)}{\delta}$$

The new pairing probability of the model is the joint likelihood of (i) a point being chosen using the planar search model and (ii) the likelihood that the point in question is included in the sphere-adjusted search space. Because these two probabilities are independent, we see that the pairing probability in the sphere adjusted space as a function of distance, $p_S(\delta)$, becomes

$$p_S(\delta) = p_P(\delta)p^G(\delta) = \frac{\beta}{\delta}\frac{D}{\pi}\frac{\sin\left(\pi\frac{\delta}{D}\right)}{\delta} = \beta\frac{\sin\left(\pi\frac{\delta}{D}\right)}{\delta^2}$$

(Note that the leading coefficient has again been generalized for the same reasons as in the planar case.) This pairing likelihood will produce matching behavior very much like that of the behavior found in the planar model near the origin, which can be intuited in two ways: (i) nearer the origin the sphere becomes approximately flat, and (ii) the Maclaurin first order approximation of the function sin(x) is simply x, yielding a first order approximation of the full function of $1/x$ in the neighborhood of zero.

Returning to the reasoning set forth above, consider two small countries, ε_A and ε_B, which lie separated by some distance δ. As distance increases and country-size decreases, the trade flow between these two countries will approach $f_{AB}(\delta) = \alpha\ \mathrm{AB}\frac{sin(\pi\frac{\delta}{D})}{\delta^2} = X_{AB}$. This equation is the spherical-search modified statement of the gravity equation, which is nearly identical over short distances, but deviates in its predictions at extreme distance (see Figure 12). This is the other startling result

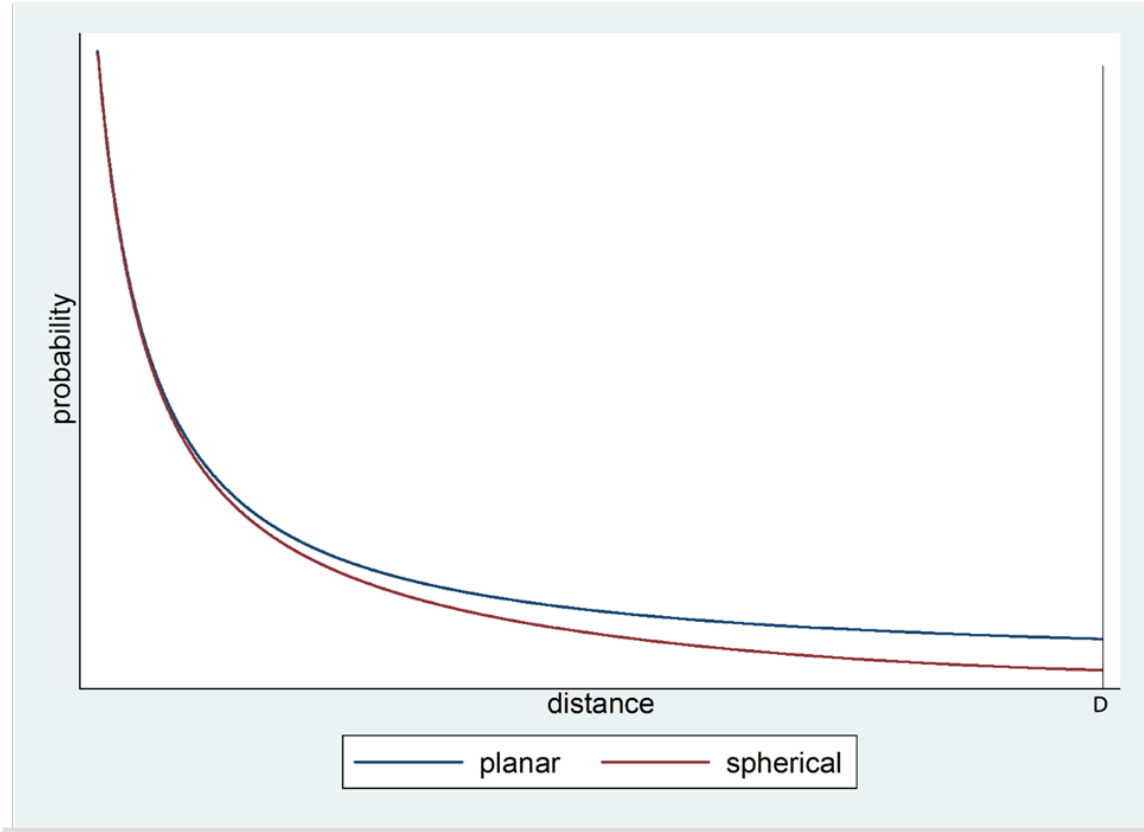

Figure 12: this is a comparison of the pairing pdfs for the original planar version of the model (blue) versus the spherical model (red). Note that they are identical near the origin, but that the pairing probability of the spherical model goes to zero as distance approaches the maximum on the globe.

of the paper: an alternative statement of the gravity equation derived from the most basic model of search and a simple adjustment for the differences between a plane and a sphere. The way in which gravity derives from search in polar space was shown above taking care to seek the weakest assumptions and greatest generality possible. The adjustment and result in the spherical case does not have a claim to generality as strong as the determinant of the Jacobean for converting from polar to Cartesian coordinates; there are other ways of distorting the planar pdf to match with a sphere and the one selected here might be the simplest but does not necessarily encompass or relate to the alternatives. However, our spherically-adjusted form for the gravity model differs from the standard model, so now it is possible to test their diverging predictions (and see whether the alternative structure is born out in the data).

4.3 Implications of Spherical Adjustment

This radiant model produces two noteworthy predictions, one minor and one significant. The first is that a continuum of agents searching in two dimensional space provides an additional reason why we might expect small countries to export a larger portion of GDP.[34] The gravity equation above was constructed by generalizing the results of the model to shrinking countries at increasing distances, in which case the search probability of all agents in a given country approach being identical. As distances shrink (or country shapes become irregular) this is less and less the case. The shortest distances describe countries' trade with themselves. The effect of country size on trade derives from the simple observation that if all agents search independently, an agent at the center of a large country is more likely to have their successful pairing "caught" by another agent in their own country when compared to an agent in a small country. So this gives a motivation for the fact that small countries export more without invoking more complex arguments about comparative advantage (factor endowments, increasing returns, et cetera). Unfortunately, the math underlying this discussion grows intractable very quickly, so will not be explored in depth here.

The second prediction is the deviation from standard gravity due to the spherical adjustment. Generally, any claim of deviation from standard gravity should met with skepticism due to the depth of the validation of the relationship. However, in this case the adjustment is so small it is extremely hard to observe in the data. To illustrate this, consider the simple regressions given in Table 2. The data are bilateral trade flows from the year 2000. The data cover 165 countries, and each observation is a bilateral trade link. Observations are only used for the purpose of this analysis if there are positive trade flows in both directions (exporter and importer) and GDP information is available for both countries, leaving us with 5887 useful observations.

In order to allow us to examine the functional form implied by the data, we will define a simple statistic (γ_{AB}) for each bilateral trade relationship

$$\gamma_{AB} = \frac{X_{AB} + X_{BA}}{Y_A Y_B} \approx \beta \frac{1}{\delta} \; or \; \beta \frac{sin(\frac{\pi}{D} d)}{d^2} \; for\ some\ \gamma > 0$$

This analysis uses the sum of bilateral exports in order to (i) remove trade imbalance issues from

[34] This is not clearly an implication unique to this model, as discussed in Lashkaripour (2006)

the analysis and (ii) reduce the impact of observation error and prevent selection bias[35].

Both the standard and modified gravity equations suggest a functional form but not a coefficient (β) for the relationship between γ and distance. Therefore, a simple method of examining how well the spherical model fits the data and comparing the performance of the models is linear regression of the data on generated values of the planar model ($x_p = \frac{1}{\delta}$) and the spherical model ($x_s = \frac{sin(\frac{\pi}{D}\delta)}{\delta^2}$). In both cases the coefficient values are uninformative, but the coefficients' significance and the model fit (adjusted R^2) can tell us how well the models are performing. Looking at the regression results (Table 2), both models fit nearly identically well both in size of the coefficient and the adjusted R^2. This similarity, and the powerful multicollinearity it implies, means that it will be extremely hard to run a comparison of these two models without more theoretical apparatus. But it also explains how the difference in functional form might have gone unnoticed before.

4.4 Spherical Model with an Intensive Margin

Above, a model was developed combining the search process with a cost-driven intensive margin was developed. It is now possible to move it into a spherical search context. If we combine the same constant-marginal cost intensive margin ($(C_\delta)^{-\sigma} Y_H$) motivated by the same home-versus-foreign CES choice with the new spherically adjusted extensive margin ($p_S(\delta)Y_F$), we get:

$$D_S(\delta, Y_H, Y_F) = p_S(\delta)Y_F(C_\delta)^{-\sigma} Y_H = \underbrace{\beta_H \frac{\sin\left(\pi\frac{\delta}{D}\right)}{\delta^2} Y_F}_{\text{Extensive}} \underbrace{(1+\alpha\delta+T)^{-\sigma}\, Y_H}_{\text{Intensive}}$$

This new demand function is a modest transformation of the extensive-margin-only spherically adjusted demand curve, owing to the fact that the effects of linearly increasing costs diminish quickly in a CES setting. But regardless, by separating out intensive and extensive margins there is now more hope of reducing the multicollinearity problem for comparing the spherical and planar versions of the extensive margin. Consider the new spherical demand equation in logs.

$$ln(\gamma_{FH}) = ln\left(2\frac{D_S(\delta, Y_H, Y_F)}{Y_H Y_F}\right) = \ln\left(\sin\left(\pi\frac{\delta}{D}\right)\right) - 2\ln(\delta) - \sigma ln(1+\ \alpha\delta + T) + ln(\beta_H) + ln(\beta_F) + C + \varepsilon$$

[35] Small countries tend to have less and worse data. Therefore, when looking at long distances, unidirectional data tend to be dominated by large rich countries as the point of origin. By requiring the availability of data in both directions it lessens the possibility of systematic bias arising from this.

The proposed model actually makes very strong assertions about the estimated parameter values of two of the coefficients in any linear regression in logs: the coefficients of the log-sin term should be one and of the log-distance term should be negative two. However, the linear cost term is unknown, and so it would be difficult to examine this term using linear regression techniques. Fortunately, the Maclaurin-series expansion of this term is fairly straightforward: [36]

$$f(\delta) = ln(1+\ \alpha\delta + T) = ln(1+T) + \left(\frac{\alpha}{1+T}\right)\delta - \left(\frac{\alpha}{1+T}\right)^2\delta^2 + \left(\frac{\alpha}{1+T}\right)^3\delta^3 - \left(\frac{\alpha}{1+T}\right)^4\delta^4 + \ldots$$

Therefore, we can examine the relationship between this model and the data by running the regression

$$ln(\gamma_{FH}) = \rho_1 \ln\left(\sin\left(\pi\frac{\delta}{D}\right)\right) + \rho_2 \ln(\delta) + \eta_1\delta + \eta_2\delta^2 + \eta_3\delta^3 + \eta_4\delta^4 + D_H + D_F + C + \varepsilon$$

where the model predicts $\rho_1 = 1$, $\rho_2 = -2$, and $\eta_1{=}{-}\sigma\frac{\alpha}{1+T}$, $\eta_2{=}\sigma\left(\frac{\alpha}{1+T}\right)^2$, $\eta_3{=}{-}\sigma\left(\frac{\alpha}{1+T}\right)^3$ and so on with alternating sine. It is important to here note that, because log of costs is multiplied by *negative* σ the coefficients should alternate sine in the opposite pattern to the Maclaurin Series expansion.[37] The strongest rejections of the model would be $\rho_1 = 0$, $\rho_2 = -1$, and $\eta_i = 0$.

But before discussing the results of this regression, there is also a simpler variation that will be informative. Consider the model without *without* the extensive margin, wherein:

$$\begin{aligned} ln(\gamma_{FH}) &= -\sigma ln(1+\ \alpha\delta + T) + ln(\beta_H) + ln(\beta_F) + C + \varepsilon \\ &\approx \eta_1\delta + \eta_2\delta^2 + \eta_3\delta^3 + \eta_4\delta^4 + D_H + D_F + C + \varepsilon \\ &\approx (1-\sigma)ln(1+\alpha\delta+T) + ln(\mathbb{P}_H) + ln(\mathbb{P}_F) + C + \varepsilon \end{aligned}$$

This regression examines only how constant marginal cost would effect demand in our crude partial-equilibrium setting. With no more search parameters to cause country-level idiosyncrasies it is not clear why country-level fixed effects should be present. However, running the regression with country-level fixed effects would be isomorphic to a standard Armington model with linear

[36] For an example of using Taylor series analysis, see Baier & Bergstrand (2009)

[37] Estimates of the elasticity of substitution vary widely, but a common Armington estimate is an elasticity of substation of about 3.5.

costs, wherein the fixed effects would now be interpreted as price level effects ($\mathbb{P}_i$) or multilateral resistance terms. This is helpful for our discussion, because it gives us an "implied" constant-marginal cost trade cost (which is not often estimated) for the data. It shows us how well the Taylor expansion fits the data. And most importantly, it allows us to examine if and how much implied trade costs fall by adding the extensive margin to the model.

The results of these regressions are given in Table 3. In the first two columns the models are estimated without fixed effects. The first column is the intensive-margin-only specification. Working under the assumptions that T is fairly small and that the first term in an alternative Taylor series is the most important and best estimated, we can hold up the estimate of $\eta_1 = -9.62e-04$ as a rough estimate for marginal trade cost (α). Using an estimated trade elasticity of $\sigma = 3.5$ implies $\alpha = 2.75e-04$ which is two orders of magnitude larger than our estimate from trade cost data ($\alpha = 2.89e-06$). So this is consistent with the assertion that *marginal* "implied" costs are very large in the trade-cost framework. In the second column, the model is run with the intensive margin included. As we can see, the estimated values of ρ_1 and ρ_2 are near what the model predicts. The Taylor series coefficients are no longer significant, but their sine and relative amplitude is consistent with what the model require. Note that the amplitude of every coefficient is smaller than that of the cost-only regression, suggesting implied costs have indeed fallen. Also, the point estimate for $\eta_1 = -2.49e-06$, though not significant, would be consistent with trade costs the correct order of magnitude when compared to what we see in the data.

Adding country level fixed effects only makes the success of the model grow more clear. First, the cost-only model is run and the implied marginal trade cost grows even larger. However it is worth noting that, if country shipping costs are idiosyncratic ($\tau_i = 1 + \alpha_i\delta + T_i$) then the inclusion of fixed effects might capture some of this heterogeneity. Unfortunately, the presence of such heterogeneity makes the Taylor expansion somewhat misspecified. This should not effect the sign or amplitude of the η_i estimates, but it does make the coefficient values hard to interpret independent of the fixed effect values, which would explain why adding fixed effects causes η's to grow larger. In either case, the addition of fixed effects causes the errors around the η estimates to fall in every specification, with or without an extensive margin.

Looking at Specification 2 and Specification 3, we again see the estimates of ρ_1 and ρ_2 are consistent with the spherically adjusted model. They again reduce the size of the marginal cost estimates (η_i) in a way consistent with premise of the model. Specifications 1, 2, and 3 together provide strong evidence of a spherically adjusted extensive margin in the data. But this evidence grows still more compelling when considering Specification 4: a "planar" extensive margin that does not include the sine function. When the model is estimated without the spherical adjustment, there are a few striking implications. First, the η_i estimates do not change substantially. This suggests that (i) the intensive margin is being identified, and (ii) switching between spherical and planar extensive margins is not changing how much of the variation is explained by the extensive margin. Second, the planar model (δ^{-1}) of the extensive margin is very nearly rejected and fits far worse than the other three specifications. And third, if we presumed that spherically adjusted model was correct, ρ_2 in Specification 4 would be akin to estimating a constant elasticity approximation of the extensive margin. Intriguingly, this value is consistent with extensive margin distance elasticities observed in the data for the decline of varieties over distance (Hummels 2005).

5 Conclusion

This paper has shown that a very simple model that generates gravity in a fundamentally different way than through trade costs. It utilizes a very basic search process, and creates gravity as a latent feature of search. This effect is independent of the structure of preferences or production. It then shows how to easily integrate this framework with more familiar models to generate a trade-cost driven intensive margin and a search-driven extensive margin. Before proceeding to greater refinement, it is helpful to pause here and note that this "planar" model is nearly as consistent with the data as the trade cost model in that it relies on a simplification of some true and obscure search process whereas the trade cost model relies on obscure, unmeasured trade costs of a specific structure.

But by examining the deeper characteristics of the search process, it is possible to find an actual falsifiable claim for the search model: the spherical adjustment. So next the paper develops the inevitable spherical search correction for the model. It also shows how to integrate this with more conventional models and takes the result to the data. The spherical adjustment is well validated in the data. It reduces the size of implied trade costs. And it produces a motive for the extensive margin that is independent of the model, and at a magnitude consistent with estimates in the literature.

This draws attention to a major open question: there has been no deep justification for the trade cost motivation for gravity. This is a very simple alternative architecture that achieves the same ends. Both have advantages and disadvantages, but the effort to distinguish them will be complex and deep.

Table 1: Shipping Cost Regressions

	Specification 1 $\tau_D - 1$	Specification 2 $ln(\tau_D - 1)$	Specification 3 $ln(\tau_D)$
Distance (α)	**2.89e-06 (2.56e-09)**		
Constant (T)	**.0464 (2.24e-05)**		
Log Distance		**0.319 (1.61e-04)**	**0.015 (4.63e-5)**
Log Constant		**-5.515 (0.001)**	**-0.067 (4.07e-4)**
R^2	0.0296	0.5654	0.033
N	3,029,547	3,029,547	3,029,547

Coefficients reported with (standard errors). All variables significant at 99% confidence or better are given in **bold**.

Table 2: Similarity of Planar & Spherical Models

	Specifcation 1 γ_{ij}	Specifcation 2 γ_{ij}
Planar Model (x_p)	1.01e-13 (16.50)	No
Spherical Model (x_s)	No	1.00e-13 (16.51)
Adj-R^2	0.0477	0.0477

Regressions of the implied functional form of $\gamma_{ij}(\delta)$ on the actual data. The t-statistics are given in parenthesis due to extreme small size of standard errors. All coefficients are highly significant. The values of γ in the data are very small, which is the reason for the small size of the coefficients.

Table 3: The Intensive and Extensive Model

Distance Regressor	Cost Only Without FE	Specification 1	Cost Only With FE	Specification 2	Specification 3	Specification 4 Planar Model
$\ln\left(\sin\left(\pi\frac{\delta}{D}\right)\right)$	No	**1.285** **(0.653)**	No	**1.200** **(0.528)**	**1.182** **(0.527)**	No
$\ln(\delta)$	No	**-2.377** **(0.606)**	No	**-2.016** **(0.503)**	**-1.680** **(0.506)**	**-0.643** **(0.207)**
δ	**-9.62e-04** **(9.05e-05)**	-2.49e-06 (0.00027)	**-1.33e-03** **(9.01e-05)**	**-7.97e-04** **(2.42e-04)**	**-9.57e-04** **(2.44e-04)**	**-7.12e-04** **(2.18e-04)**
δ^2	**1.36e-07** **(1.93e-08)**	3.70e-08 (4.32e-08)	**1.82e-07** **(1.87e-08)**	**1.41e-07** **(3.92e-08)**	**1.55e-07** **(3.92e-08)**	**1.03e-07** **(3.17e-08)**
δ^3	**-9.30e-12** **(1.56e-12)**	-4.10e-12 (3.09e-12)	**-1.16e-11** **(1.49e-12)**	**-1.02e-11** **(2.81e-12)**	**-1.08e-11** **(2.80e-12)**	**-6.79e-12** **(2.16e-12)**
δ^4	**2.30e-16** **(4.18e-17)**	1.52e-16 (8.53e-17)	**2.62e-16** **(3.92e-17)**	**2.70e-16** **(7.66e-17)**	**2.79e-16** **(7.65e-17)**	**1.54e-16** **(5.24e-17)**
Country FE	No	No	Yes	Yes	Yes	Yes
Border Effects	Yes	Yes	Yes	No	Yes	Yes
R^2	0.229	0.229	0.545	0.545	0.547	0.546

Coefficients reported with (standard errors). All variables at 95% confidence or better are given in **bold**.

References

Anderson, James E., and Eric Van Wincoop. “Gravity with gravitas: a solution to the border puzzle.” *The American Economic Review* 93.1 (2003): 170-192.

James E. Anderson; Eric van Wincoop “Trade Costs” *Journal of Economic Literature*, Vol. 42, No. 3. (Sep., 2004), pp. 691-751

James E. Anderson, 2011. “The Gravity Model,” *Annual Review of Economics*, *Annual Reviews*, vol. 3, pages 133-160

Armington, Paul S. “A theory of demand for products distinguished by place of production.” Staff Papers 16.1 (1969): 159-178.

Baier, Scott L., and Jeffrey H. Bergstrand. “Bonus vetus OLS: A simple method for approximating international trade-cost effects using the gravity equation.” *Journal of International Economics* 77.1 (2009): 77-85.

Baldwin, Richard. “Heterogeneous firms and trade: testable and untestable properties of the Melitz model.” No. w11471. National Bureau of Economic Research, 2005.

Balistreri, Edward J., and Hillberry Russell H. “Trade Frictions and Welfare in the Gravity Model: How Much of the Iceberg Melts?” *The Canadian Journal of Economics / Revue Canadienne D'Economique* 39, no. 1 (2006): 247-65.

Chaney, Thomas. 2008. ”Distorted Gravity: The Intensive and Extensive Margins of International Trade.” *American Economic Review*, 98(4): 1707-21.

Chen, Natalie, and Dennis Novy. “Gravity, trade integration, and heterogeneity across industries.” *Journal of International Economics* 85.2 (2011): 206-221.

Coughlin, Cletus & Novy, Dennis. “Estimating Border Effects: The Impact of Spatial Aggregation.” CEPR Discussion Paper No. DP11226.

Costinot, Arnaud, and Andres Rodriguez-Clare. Trade theory with numbers: Quantifying the consequences of globalization. No. w18896. National Bureau of Economic Research, 2013.

Deardorff, Alan. ”Determinants of bilateral trade: does gravity work in a neoclassical world?.” *The regionalization of the world economy*. University of Chicago Press, 1998. 7-32.

Disdier, Anne-Clia, and Keith Head. “The puzzling persistence of the distance effect on bilateral trade.” *The Review of Economics and Statistics* 90.1 (2008): 37-48.

Dixon, Peter, Michael Jerie, and Maureen Rimmer. ”Modern trade theory for CGE modelling: the Armington, Krugman and Melitz models.” *Journal of Global Economic Analysis* 1.1 (2016): 1-110.

Feenstra, Robert C., James R. Markusen, and Andrew K. Rose. ”Using the gravity equation to differentiate among alternative theories of trade.” *Canadian Journal of Economics/Revue canadienne d'conomique* 34.2 (2001): 430-447.

Helpman, Elhanan, Marc J. Melitz, and Stephen R. Yeaple. ”Export Versus FDI with Heterogeneous Firms.” *American Economic Review* 94, 1 (March 2004): 300-316

Elhanan Helpman & Marc Melitz & Yona Rubinstein, 2008. ”Estimating Trade Flows: Trading Partners and Trading Volumes,” *The Quarterly Journal of Economics*, MIT Press, vol. 123(2), pages 441-487, 05.

Hummels, David, and Peter J. Klenow. ”The variety and quality of a nation’s exports.” *The American Economic Review* 95.3 (2005): 704-723.

Hummels, David L. ”Toward a geography of trade costs.” Available at SSRN 160533 (1999).

Imbs, Jean, and Isabelle Mejean. ”Elasticity optimism.” *American Economic Journal: Macroeconomics* 7.3 (2015): 43-83.

Kimura, F. & Lee, “The Gravity Equation in International Trade in Services”, *H. Rev. World Econ*. April 2006, Volume 142, Issue 1, pp 92–121

Lawless, Martina, and Karl Whelan. ”A note on trade costs and distance.” (2007).

Melitz, Marc J. and Stephen J. Redding. 2015. ”New Trade Models, New Welfare Implications.” *American Economic Review*, 105(3): 1105-46.

Melitz, Marc J., and Gianmarco IP Ottaviano. ”Market size, trade, and productivity.” *The review of economic studies* 75.1 (2008): 295-316.

Melitz, Marc J. ”The Impact Of Trade On Intra-Industry Reallocations And Aggregate Industry Productivity,” *Econometrica*, 2003, v71(6,Nov), 1695-1725.

Miroudot, S., R. Lanz and A. Ragoussis (2009), “Trade in Intermediate Goods and Services”, *OECD Trade Policy Papers*, No. 93, OECD Publishing.

Miroudot, Sbastien, Jehan Sauvage, and Ben Shepherd. ”Measuring the cost of international trade in services.” *World Trade Review* 12.04 (2013): 719-735.

Novy, Dennis. ”International trade without CES: Estimating translog gravity.” *Journal of International Economics* 89.2 (2013): 271-282.

Obtsfeld & Rogoff ”The Six Major Puzzles of International Macroeconomics: Is there a Common Cause?”. Chapter in *NBER Macroeconomics Annual 2000*, Volume 15. 2001 (p.339-412).

Rauch, F. (2016), The Geometry of the Distance Coefficient in Gravity Equations in International Trade. *Review of International Economics*. doi:10.1111/roie.12252

Simonovska, Ina & Waugh, Michael E., 2014. ”The elasticity of trade: Estimates and evidence,” *Journal of International Economics*, Elsevier, vol. 92(1), pages 34-50

Walsh, Keith. ”Trade in services: does gravity hold?.” *Journal of World Trade* 42.2 (2008): 315-334.

Tinbergen, Jan. 1962. “Shaping the World Economy: Suggestions for an International Economic Policy.” New York: Twentieth Century Fund.