The Gravity Equation in International Trade: An Explanation

Thomas Chaney

Sciences Po

The gravity equation in international trade states that bilateral exports are proportional to economic size and inversely proportional to geographic distance. While the role of size is well understood, that of distance remains mysterious. I offer an explanation for the role of distance: If (i) the distribution of firm sizes is Pareto, (ii) the average squared distance of a firm's exports is an increasing power function of its size, and (iii) a parameter restriction holds, then the distance elasticity of trade is constant for long distances. When the firm size distribution follows Zipf's law, trade is inversely proportional to distance.

I. Introduction

Half a century ago, Tinbergen (1962) used an analogy with Newton's universal law of gravitation to describe the patterns of bilateral aggregate trade flows between any two countries *A* and *B* as "proportional to the gross na-

I want to thank Fernando Alvarez, Marianne Andries, Michal Fabinger, Xavier Gabaix, Sam Kortum, Bob Lucas, Nour Meddahi, Marc Melitz, Jim Tybout, Stephane Villeneuve, Jon Vogel, and seminar participants in Berkeley, Bilkent, Bocconi, Boston University, Chicago, Columbia, Erasmus, Harvard, Hitotsubashi, KU Leuven, London Business School, Louvain-CORE, London School of Economics, Mannheim, the New York Fed, Oxford, Princeton, Rochester, Sciences Po, Toulouse, University of British Columbia Vancouver, University College London, Yale, and Zurich for helpful discussions and National Science Foundation grant SES-1061622 and European Research Council grant 337272—FiNet for financial support. I am indebted to Jong Hyun Chung for his research assistance. The first drafts of this paper were written when I held positions at the University of Chicago and Toulouse School of Economics. Data are provided as supplementary material online.

Electronically published January 16, 2018
[*Journal of Political Economy*, 2018, vol. 126, no. 1]

tional products of those countries and inversely proportional to the distance between them" (65):[1]

$$\text{Trade}_{A,B} \propto \frac{(\text{GDP}_A)^{\alpha}(\text{GDP}_B)^{\beta}}{(\text{Distance}_{AB})^{\zeta}},$$

with α, β, and $\zeta \approx 1$. The so-called "gravity equation" in international trade has proven surprisingly stable over time and across different samples of countries and methodologies. Head and Mayer (2014) use a meta-analysis of 1,835 estimates of the distance coefficient ζ in gravity type regressions in 161 published papers; 328 of those estimates are "structural," in the sense that they use either importer/exporter fixed effects or a ratio-type method. The mean distance elasticity is -1.1 (median -1.14 and standard deviation [SD] 0.41) among the "structural" estimates and -0.93 (median -0.89 and SD 0.4) among all estimates. The distance elasticity is remarkably stable, hovering around -1 over a century and a half of data. The size coefficients α and β are also stable and close to one.[2]

While the role of economic size (α, $\beta \approx 1$) is well understood in a variety of theoretical settings, to this day no explanation for the role of distance ($-\zeta \approx -1$) has been found. This paper offers an explanation for the role of distance.

The first contribution of this paper, proposition 1, is to identify three sufficient conditions under which the gravity equation holds. If (i) the distribution of firm sizes is Pareto, (ii) the average squared distance of a firm's exports is an increasing power function of its size, and (iii) a parameter restriction holds, then the distance elasticity of trade converges to a constant for long distances, equal to -1 in the special case of Zipf's law.

The reason why the shape of the distribution of firm sizes matters is as follows. If long-distance trade is mostly the prerogative of large firms and short-distance trade that of small ones, then whether distance has a weak or a strong negative impact on aggregate trade depends on whether there are many or few large firms relative to small ones. The fewer large firms there are, the stronger the negative impact of distance on trade; the faster the distance of exports rises with firm size, the milder the negative impact of distance on trade.

Formally, when firm sizes are Pareto distributed and the average squared distance of exports is an increasing power function of distance, the dis-

[1] Since then, the empirical trade literature has typically used GDP as a measure of size rather than GNP. Both measures give similar results.

[2] Anderson and van Wincoop (2003) extend Bergstrand's (1985) early contribution and show how to structurally estimate gravity equations consistent with a simple Armington model and how to deal with differences in country sizes. Santos Silva and Tenreyro (2006), Helpman, Melitz, and Rubinstein (2008), and Eaton, Kortum, and Sotelo (2012) show how to accommodate zeros in bilateral trade when estimating gravity equations.

tance elasticity of trade is asymptotically negative and constant for long distances. One of the strengths of this result is that, beyond condition (ii) on the average squared distance of exports, I need not impose any restriction on how distance affects firm-level exports. I do not have to take a stance on the specific nature of trade frictions at the firm level. In the special case in which the distribution of firm sizes has a very thick upper tail and converges to Zipf's law, how much further away large firms export compared to small ones becomes negligible, and the distance elasticity of aggregate trade is approximately -1.

Using data on French exporters, I show that Zipf's law is a good approximation of the distribution of firm sizes among large firms; larger firms export over longer distances than small ones in a such way that the average squared distance of exports is approximately a power function of firm size, the parameter restriction I need to impose is satisfied, and the distance elasticity of aggregate trade is close to -1, as the theory predicts. Using data disaggregated into industrial sectors, I show that systematic variations in the distribution of firm sizes and in the power function linking firm size and average squared distance of exports are associated with systematic variations in the distance elasticity of trade, as the theory predicts.

The second contribution of this paper is to build a micro-founded model in which conditions (i) and (ii) are derived endogenously and (iii) has a structural interpretation. In this model, firms are distributed over a geographic space and trade intermediate inputs with each other. I assume that information about potential suppliers and customers is costly. Firms gradually acquire this information over time and build a network of suppliers and customers spanning increasingly long distances, becoming larger and more efficient. This process of firm growth leads to (i) an invariant Pareto distribution of firm sizes and (ii) an average squared distance of exports that is an increasing power function of firm size, so that the gravity equation for aggregate trade holds.

In contrast, within the strict confines of existing trade theories, to explain the -1 distance elasticity of trade, some knife-edge alignment of all deep parameters is required, with no a priori justification.[3] Even if that condition were coincidentally satisfied in a particular year, for a particu-

[3] For example, in the Anderson (1979) or the Krugman (1980) models, the product of the demand elasticity and of the distance elasticity of iceberg trade costs must equal -1. In the Eaton and Kortum (2002) or Bernard et al. (2003) models, the product of the Frechet shape parameter of sectors' or firms' productivity and of the distance elasticity of iceberg trade costs must equal -1. Arkolakis, Costinot, and Rodriguez-Clare (2012) show that similar predictions can be derived in many settings with or without heterogeneous firms. Across all those models, a knife-edge condition on deep parameters must be imposed for the distance elasticity to equal -1.

lar sector and a particular country, it is hard to understand how it could survive more than a century of changes in the technology of transportation, the political impediments to trade, the nature of the goods traded, or the relative importance of the countries trading these goods. I show instead that as long as the distribution of firm sizes is relatively stable, as long as larger firms export over longer distances than smaller ones, and even if the geography of firm-level trade undergoes radical changes, the gravity equation for aggregate trade remains stable. Of course, within my model, I must impose conditions such that the distribution of firm sizes remains stable and approximately Zipf; but given evidence gathered in advanced economies over several decades in which the firm size distribution is indeed stable, the requirement for a distance elasticity of trade stable and close to -1 in my model is milder than in other conventional trade models.

Recent developments in trade theory have already shifted the attention of trade economists toward properties of the size distribution of firms. This is the case of the seminal models of Bernard et al. (2003) or Melitz (2003), where the distribution of firm sizes does matter for the patterns of aggregate trade, as demonstrated in Bernard et al. (2003) and in Chaney's (2008) extension of the Melitz model. Those theories, however, have remained attached to a rudimentary notion of trade frictions (iceberg costs and, in the case of the Melitz model, fixed export costs). Interpreted literally, such trade frictions impose too narrow a structure on the patterns of firm-level trade, and the conditions for the -1 distance elasticity of trade are generically not satisfied.

Beyond generating endogenously the gravity equation in international trade, my model also offers tractable tools to analyze an economy composed of a complex network of input-output linkages between firms related to the model of Oberfield (forthcoming) and to my earlier model in Chaney (2014). The main difference compared to Oberfield's model is that I allow firms to buy inputs from more than one firm in equilibrium, but I impose a strong symmetry condition to keep the model tractable. The main substantive difference compared to that in Chaney (2014) is that I characterize the aggregate properties of the model, while I restricted my analysis to firm-level outcomes in my earlier work. The main technical difference is that I do away with "direct search" and take a more literal network view of information transmission (firms meet new trading partners only through their existing partners). To characterize the evolution of firm-level outcomes and derive aggregate predictions in such a complex network of firm linkages, I do not solve for the characteristics of any particular firm, but rather for the probability distribution of those characteristics. Using Fourier transforms offers solutions for those distributions in a dynamic system. These tools are general purpose and can be used beyond the specific trade application in this paper, for instance, to analyze first-, second-, or even higher-order moments of aggregate variables in a network economy.[4]

The remainder of the paper is organized as follows. In Section II, I show that three conditions are sufficient for the gravity equation in international trade to emerge. In Section III, I present empirical evidence that those conditions are approximately satisfied in the data and can explain key features of the geography of aggregate and industry-level trade. In Section IV, I present a stylized dynamic model of firm-to-firm trade in which those conditions emerge endogenously.

II. A Simple Explanation for the Gravity Equation in Trade

In this section, I give three conditions sufficient for the gravity equation in international trade to hold. This explanation is purely mathematical: I do not propose, yet, any justification for those conditions. Section IV presents a micro-founded model in which they are derived endogenously.

The role of economic size (typically measured as GDP) is well understood since at least Krugman (1980): When the size of a country doubles, its production doubles so that its exports approximately double; its consumption doubles so that its imports approximately double; that is, the elasticity of aggregate trade with respect to importer and exporter size is expected to be close to one, as it is in the data. Departures from the unit elasticity with respect to size when countries of asymmetric size trade together are also well understood, as shown by Anderson and van Wincoop (2003). As I have nothing new to say about the role of economic size, I focus my attention entirely on the role of distance. For the remainder of this paper, consider trade normalized by origin and destination country size. The next proposition states my main result formally.

Proposition 1. If the following three conditions hold:

(i) firms sizes follow a Pareto distribution over $[K_{\min}, +\infty)$ with shape parameter $\lambda \geq 1$;
(ii) the average squared distance of exports is an increasing power function of firm size, meaning that the fraction of exports shipped at a distance x by firms of size K is given by a function $f_K(x)$ such that

[4] Similar tools are used to study the percolation of information about some underlying payoff-relevant information in a network of traders (see, e.g., Duffie and Manso 2007). While the application is different, diffusion of information about the geographic location of trading partners in my case, diffusion of information about the value of an asset in the other, the underlying mathematical structure of the problem is similar.

$$\int_0^\infty x^2 f_K(x)\,dx = K^\mu \left[\int_0^\infty x^2 f_{K_{\min}}(x)\,dx\right] \quad \text{with } \mu > 0,$$

where I further impose that $f_K(x)$ and $f'_K(x)$ are bounded from above and that $f_K(x)$ is weakly decreasing above some threshold $\bar{x}$; and

(iii) $\lambda < 1 + \mu$,

then $-\zeta$, the elasticity of aggregate trade between two countries A and B normalized by country size ($\text{Trade}_{A,B}$) with respect to distance, is asymptotically constant:

$$\text{Trade}_{A,B}(\text{Distance}_{A,B} = x) \underset{x\to+\infty}{\propto} \frac{1}{x^\zeta} \quad \text{with } \zeta = 1 + 2(\lambda - 1)/\mu.$$

Furthermore, if the distribution of firm sizes is close to Zipf's law ($\lambda \approx 1$), then aggregate trade is inversely proportional to distance ($\zeta \approx 1$).

Proof. To save on notation, normalize size and distance units so that $K_{\min} = 1$ and $\int_0^\infty x^2 f_1(x)\,dx = 1$. Aggregate exports at a distance x, denoted by $\varphi(x)$, are the sum of all firm-level exports at x. From condition (i), firm sizes are Pareto distributed with parameter λ:

$$\varphi(x) \propto \int_1^\infty [K f_K(x)] K^{-\lambda-1}\,dK.$$

Introduce the scaled function $g_K(x) \equiv K^{\mu/2} f_K(K^{\mu/2}x)$. Note that from condition (ii), all the g_K's have the exact same second moment:

$$\int_0^\infty x^2 g_K(x)\,dx = \int_0^\infty x^2 f_1(x)\,dx = 1 \qquad \forall K \geq 1.$$

Plugging the g_K's in the expression for φ and with the change of variable $u = K^{-\mu/2}x$,

$$\begin{aligned}\varphi(x) &\propto \int_1^\infty K^{-\lambda-\mu/2} g_K(K^{-\mu/2}x)\,dK \\ &\propto x^{-[1+2(\lambda-1)/\mu]} \int_0^x u^{2(\lambda-1)/\mu} g_{(x/u)^{2/\mu}}(u)\,du.\end{aligned}$$

From lemma 1 in the appendix,

$$\lim_{x\to\infty} \int_0^x u^{2(\lambda-1)/\mu} g_{(x/u)^{2/\mu}}(u)\,du < \infty.$$

This is to be expected as all the g_K's have the same finite second moment, and $2(\lambda - 1)/\mu < 2$ from condition (iii). As proposed,

$$\varphi(x) \underset{x\to+\infty}{\propto} \frac{1}{x^{1+2(\lambda-1)/\mu}}.$$

Trivially, when the distribution of firm sizes is close to Zipf's law ($\lambda \approx 1$), then $1 + 2(\lambda - 1)/\mu \approx 1$, and aggregate trade is inversely proportional to distance, $1/x^{1+2(\lambda-1)/\mu} \approx 1/x$. QED

There is ample empirical evidence that the distribution of firm sizes can be approximated by a Pareto distribution, as in condition (i), and that among Pareto distributions, Zipf's law ($\lambda \approx 1$) is a good candidate (see, e.g., Axtell 2001). If Zipf's law holds approximately, then condition (iii) is easily satisfied, for any $\mu > 0$. So in practice, condition (iii) is unlikely to be violated.[5] Section III presents empirical evidence using French firm-level trade data showing that conditions (i), (ii), and (iii) hold approximately, with $\lambda \approx 1.005$ and $\mu \approx 0.11$. Section IV offers a theoretical model in which both conditions (i) and (ii) emerge endogenously and condition (iii) has a structural interpretation.

To gain further intuition for the mathematics behind proposition 1, I consider three special cases for the geography of firm-level exports in which aggregate trade can be calculated analytically.

Firms export to a single location.—Let us start with the extreme simplifying assumption that an exporter of size K exports toward a single location. Condition (ii) imposes that the average squared distance of export of a size K firm be proportional to K^{μ}, so the firm exports only at a distance x proportional to $K^{\mu/2}$ and nowhere else. Condition (i) further imposes that the distribution of firm sizes is Pareto with shape parameter λ: There are $dF(K) \propto K^{-\lambda-1}dK$ firms of size K, with total sales proportional to $K^{-\lambda}dK$. Only those firms export at a distance $x \propto K^{\mu/2}$, so $K^{-\lambda}dK$ is the volume of aggregate exports at that distance. With the simple change of variable $x \propto K^{\mu/2}$, exports at a distance x are inversely proportional to $x^{1+2(\lambda-1)/\mu}$, as in proposition 1.

Firm-level trade independent of distance.—Now let us assume that the distribution of firm-level exports is a uniform function of distance. With firm-level exports uniform in distance, condition (ii) imposes that a firm of size K exports the same amount, $K^{1-\mu/2}/\sqrt{3}$, toward each distance x in the interval $[0, \sqrt{3}K^{\mu/2}]$, and nothing beyond. Only those firms with a size K larger than $3^{1/\mu}x^{2/\mu}$ export at a distance x. Aggregate exports at a dis-

[5] Even if condition (iii) were violated, trade is still proportional to $1/x^{1+2(\lambda-1)/\mu}$ as long as f_1 admits a finite $\lceil 2(\lambda - 1)/\mu \rceil$th moment. If $\lambda > 1 + \mu$, this is a moment of order higher than 2, so that f_1's finite second moment in condition (ii) is no longer sufficient. For many conventional distributions, all moments exist and are finite, or at least several are (including moments above the second one).

tance x, $\varphi(x)$, are then the sum of exports of all firms larger than $3^{1/\mu}x^{2/\mu}$. With Pareto distributed firms as in condition (i),

$$\varphi(x) \propto \int_{K=3^{1/\mu}x^{2/\mu}}^{\infty} \frac{1}{\sqrt{3}} K^{1-\mu/2} K^{-\lambda-1} dK \propto \frac{1}{x^{1+2(\lambda-1)/\mu}} \qquad \forall x \geq \sqrt{3}.$$

Aggregate trade is initially constant for distances shorter than $\sqrt{3}$ and then falls off with distance with an elasticity exactly equal to $-[1 + 2(\lambda - 1)/\mu]$, as in proposition 1.

Zipf's law and firm-level trade decays exponentially with distance.—Finally, let us assume that the distribution of firm sizes follows Zipf's law and firm-level trade decays as an exponential function of distance. With exponential decay, condition (ii) imposes that a firm of size K exports $K^{1-\mu/2}\exp(-K^{-\mu/2}x)$ at a distance x. Aggregate exports at a distance x, $\varphi(x)$, are the sum of all firm-level exports. With Pareto distributed firm sizes as in condition (i),

$$\begin{aligned}\varphi(x) &\propto \int_{1}^{\infty} K^{1-\mu/2} \exp\left(-K^{-\mu/2}x\right) K^{-\lambda-1} dK \\ &= \int_{0}^{\infty} \frac{2}{\mu} \exp\left[-u\left(1 + 2\frac{\lambda - 1}{\mu}\right) - e^{-u}x\right] du,\end{aligned}$$

where I use the change of variable $K^{-\mu/2} = e^{-u}$ to get the last equation. If the distribution of firm sizes is exactly Zipf ($\lambda = 1$), the last integral can be solved in closed form:

$$\varphi(x) \propto \int_{0}^{\infty} \exp(-u - e^{-u}x) du = \frac{1 - e^{-x}}{x}.$$

As the distance x grows large, aggregate exports behave like $1/x$, as in proposition 1.

Those three special cases show analytically that under conditions (i) and (ii), the distance elasticity of aggregate trade is either constant or asymptotically constant and equal to -1 under Zipf's law. Those examples also make it obvious that the details of the geography of firm-level trade beyond condition (ii) are not very important. In particular, I do not need to assume that firm-level trade follows a gravity type equation: Firm-level trade can be a mass point, a uniform distribution, or an exponential distribution, each of which differs fundamentally from the constant distance elasticity of the gravity equation (trade as a power function of distance). It is also perfectly fine if not all firms of a given size export to the very same locations. Some may export over short distances and some over long distances. Condition (ii) imposes only that among a set of firms of size K, the average squared distance of exports is a power function of K.

The intuition for why the distribution of firm sizes (the parameter λ) and how far large firms export relative to small ones (the parameter μ)

matter for the geography of aggregate trade is straightforward. If small firms mostly export over short distances, at least on average, and large firms export over long distances, on average, then aggregate exports toward remote locations are mostly coming from large firms. The more large firms there are relative to small ones (the smaller λ) or the faster the distance of firm-level trade increases with firm size (the larger μ), the milder the negative impact of distance on aggregate trade (the smaller $1 + 2(\lambda - 1)/\mu$). Furthermore, if the distribution of firm sizes is very thick tailed and is close to Zipf's law ($\lambda \approx 1$), then surprisingly, whether the average distance of exports increases fast or slowly with firm size ($\mu > 0$ large or small) no longer matters, and aggregate exports are inversely proportional to distance $-\zeta = 1 - 2(\lambda - 1) \approx -1$.

This is not a tautological proposition. Zipf's law is a statement about how much different firms sell. The gravity equation is a statement about where different countries export.

One strength of proposition 1 is to formally link one power law (the distribution of firm sizes) with another (the geographic distribution of aggregate exports). It is as if the power law property of the size distribution of firms were "transferred" to the geographic distribution of aggregate exports. Proposition 1 is of general interest: It says that for any population of agents with Pareto distributed sizes such that larger agents behave differently from small ones in one alternative dimension, this alternative dimension will also be Pareto distributed with a specific exponent. For instance, if firms' (or plants') employment sizes are Pareto distributed and firms (or plants) with more employees pay higher wages in such a way that average squared wages are a power function of firm or plant size, then according to proposition 1, this implies that the income distribution should also be Pareto distributed. Many economic variables are Pareto distributed (see Gabaix 2008), so proposition 1 may prove useful in various settings. I leave such applications for future research.

III. Empirical Evidence

I now turn to an empirical exploration of proposition 1 using French firm-level trade data. First, I present evidence that conditions (i)–(iii) and proposition 1 hold in the data on all firms. Second, I explore a more subtle prediction emanating from proposition 1: Variations across sectors in the distribution of firm sizes and in how firm size affects the average squared distance of exports ought to be associated with variations in the distance elasticity of sectoral trade.

A. Data

Firm-level data.—To calibrate the distribution of firm sizes and the squared distance of exports, I use data on French exporters in 1992. The data are collected by the French customs and are similar to the data I used in Chaney (2014). For each exporter, I know how much (in French francs) and where it exports. I restrict my sample to firms that export over 1 million French francs (US$200,000 in 1992). There are around 28,000 such firms in my sample. When needed, data are disaggregated into 82 industrial sectors (three-digit).

Aggregate data.—Aggregate bilateral exports from France are constructed by summing up firm-level exports, either across all firms or within each industrial sector. I match this data set with data on population-weighted distances between countries from the Centre d'Etudes Prospectives et d'Informations Internationales (CEPII) and with data on bilateral trade flows from the National Bureau of Economic Research.[6]

B. *Variable Definitions and Parameter Estimation*

Following Axtell (2001), I order all French exporters in increasing order of size, where a firm's size is the value of its worldwide exports. I construct 50 bins of equal log width, $b = 1, \ldots, 50$, ranging from 1 million French francs to the actual largest amount exported by a single firm.

The average size of firms in bin b is given by

$$K_b = \frac{\sum_i \sum_c \text{exports}_{i,c} \mathbf{1}[i \in b]}{\sum_i \mathbf{1}[i \in b]}, \tag{1}$$

where $\text{exports}_{i,c}$ are total exports of firm i to country c and $\mathbf{1}[\cdot]$ is the indicator function.

The fraction of firms of size larger than K_b is given by

$$1 - F(K_b) = \frac{\sum_{b'=b}^{50} \sum_i \mathbf{1}[i \in b']}{\sum_{b''=1}^{50} \sum_i \mathbf{1}[i \in b'']}. \tag{2}$$

The average squared distance of exports among firms in bin b is given by

$$\Delta(K_b) = \sum_c (\text{Distance}_{\text{France},c})^2 \left(\frac{\sum_{i \in b} \mathbf{1}[\text{exports}_{i,c} > 0]}{\sum_{c'} \sum_{i \in b} \mathbf{1}[\text{exports}_{i,c'} > 0]} \right), \tag{3}$$

where $\text{Distance}_{\text{France},c}$ is the distance between France and country c. In each bin b, there are many firms. From this large number of firms, the fre-

[6] See Mayer and Zignago (2006) for further details on the construction of the distance measures. See Feenstra et al. (2004) for further details on the construction of bilateral trade data from the UN Comtrade data set. I am grateful to Thierry Mayer for providing me with a concordance table between the pre-1993 French nomenclature of industrial sectors (Nomenclature d'Activites et de Produits) and the international nomenclature (International Standard Industrial Classification).

quency at which firms in bin b export to country c is an empirical proxy for the function $f_{K_b}(\text{Distance}_{\text{France},c})$.[7]

Proposition 1 relates the shape parameter of the distribution of firm sizes (λ) and the size elasticity of the average squared distance of exports (μ) to the distance elasticity of trade (ζ).

The shape parameter of the Pareto distribution of firm sizes, λ, is estimated by ordinary least squares (OLS) from

$$\ln[1 - F(K_b)] = a - \lambda \ln(K_b) + \epsilon_b. \tag{4}$$

The size elasticity of the average squared distance of exports, μ, is estimated by OLS from

$$\ln \Delta(K_b) = a + \mu \ln(K_b) + \epsilon_b. \tag{5}$$

The distance elasticity of aggregate exports, ζ, is estimated by OLS from

$$\ln\left(\text{Exports}_{\text{France},c}\right) = a + b\ln(\text{Imports}_c) - \zeta \ln(\text{Distance}_{\text{France},c}) + \epsilon_c, \tag{6}$$

with $\text{Exports}_{\text{France},c}$ the sum of exports toward country c of all French firms and Imports_c the total imports by country c.[8] As proposition 1 is an asymptotic result, I consider two specifications of (6): Either with all distances or with only distances above 2,000 kilometers (approximately 1,200 miles, about 80 percent of the sample).

The measures K_b, $1 - F(K_b)$, and $\Delta(K_b)$ in equations (1)–(3) and the parameters λ, μ, and ζ in equations (4)–(6) are computed using either data on all firms or data disaggregated into sectors.

C. *Testing Proposition 1 across All Firms*

Figure 1 shows that conditions (i)–(iii) in proposition 1 are approximately satisfied. The top panel shows that the distribution of firm sizes is well

[7] I could instead use a value-weighted measure of this frequency:

$$\Delta(K_b) = \sum_c (\text{Distance}_{\text{France},c})^2 \left(\frac{\sum_{i\in b}\text{exports}_{i,c}\mathbf{1}\left[\text{exports}_{i,c} > 0\right]}{\sum_{c'}\sum_{i\in b}\text{exports}_{i,c'}\mathbf{1}\left[\text{exports}_{i,c'} > 0\right]}\right).$$

Using value weights introduces some additional amount of residual noise, but the statistical tests in this section are all robust to using this alternative specification, as shown in the online appendix.

[8] I use data only on French exports, so that I cannot include both importer and exporter fixed effects, which Head and Mayer (2014) recommend as good practice for estimating gravity equations. This specification with a control for aggregate imports rather than GDP is, however, immune to the omitted variable critique of Anderson and van Wincoop (2003): Aggregate imports by country c control for any unobserved differences in prices between c and other exporters (the inward resistance term of Anderson and van Wincoop); the constant a controls for any difference in prices between France and other exporters (the outward resistance term).

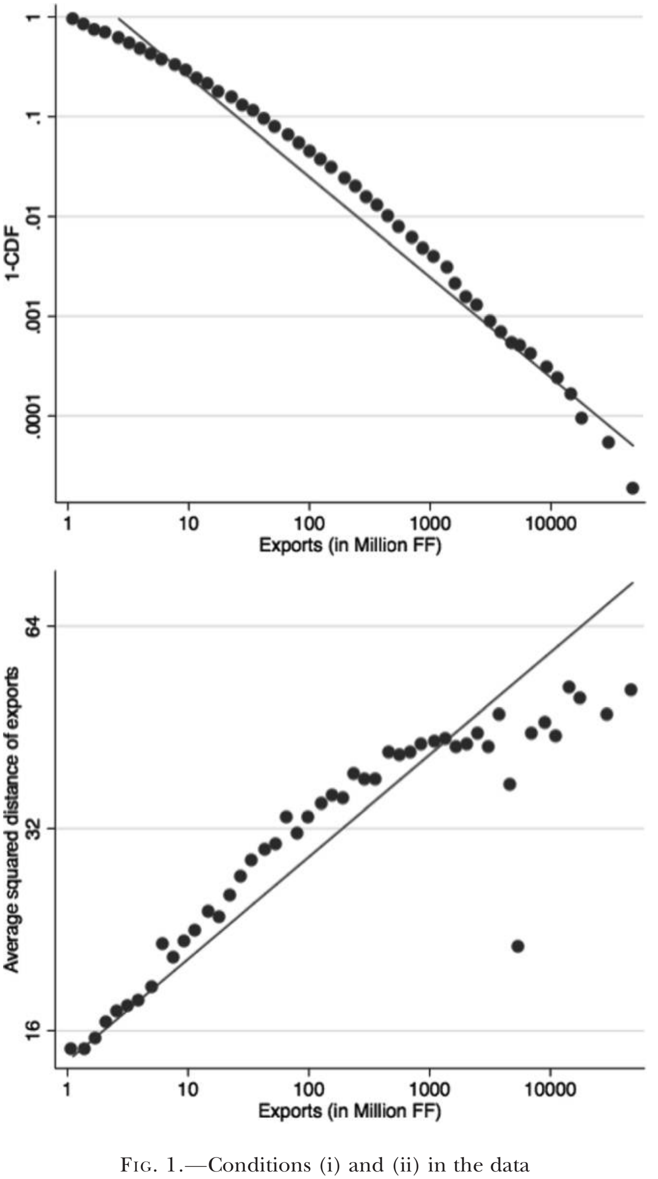

FIG. 1.—Conditions (i) and (ii) in the data

TABLE 1
Testing Proposition 1 across All Firms

Condition (i): distribution of firm sizes	$\lambda = 1.0048$ (SE = 0.0213, $R^2 = .981$)
Condition (ii): average squared distance of exports	$\mu = 0.113$ (SE = 0.0078, $R^2 = .817$)
Condition (iii): parameter restriction ($\lambda < 1 + \mu$)	$\Pr(\lambda \geq 1 + \mu) = 0.02\%$
Distance elasticity of trade:	
All distances	$\zeta_{\text{all}} = 0.767$ (SE = 0.111, $R^2 = .810$)
Long distances (>2,000 km)	$\zeta_{\text{long}} = 1.090$ (SE = 0.215, $R^2 = .720$)
Predicted distance elasticity of trade	$1 + 2(\lambda - 1)/\mu = 1.086$ (SE = 0.520)
Proposition 1:	
Wald test for $\zeta_{\text{all}} = 1 + 2(\lambda - 1)/\mu$	p-value of χ^2 test = 99.3%
Wald test for $\zeta_{\text{long}} = 1 + 2(\lambda - 1)/\mu$	p-value of χ^2 test = 99.4%

Source.—French customs, CEPII, and NBER.

Note.—This table tests proposition 1 using data on all 27,903 French firms that export more than 1 million French francs in 1992 (≈US$200,000). The variable λ is the Pareto shape coefficient for the distribution of firm sizes, estimated from (4); μ is the size elasticity of the average squared distance of exports, estimated from (5); $\Pr(\lambda \geq 1 + \mu)$ is calculated using 10,000 bootstrapped estimates of λ and μ; ζ is the distance elasticity of aggregate exports, estimated from eq. (6). The standard error of the predicted distance elasticity of trade, $1 + 2(\lambda - 1)/\mu$, is computed using 10,000 bootstrapped estimates of λ and μ. The p-value for the Wald test of $\zeta = 1 + 2(\lambda - 1)/\mu$ is computed by comparing the Wald statistic $W = \{\zeta - [1 + 2(\lambda - 1)/\mu]\}^2/\{\mathrm{Var}(\zeta) + \mathrm{Var}[1 + 2(\lambda - 1)/\mu]\}$ to a χ^2, where $\mathrm{Var}[1 + 2(\lambda - 1)/\mu]$ is calculated using 10,000 bootstrapped estimates of λ and μ. Robust standard errors and adjusted R^2 are in parentheses.

approximated by Zipf's law, $1 - F(K) \propto K^{-\lambda}$ with $\lambda = 1.0048$, as in condition (i). The bottom panel shows that the average squared distance of exports is approximately a power function of firm size, $\Delta(K) \propto K^{\mu}$ with $\mu = 0.113$, as in condition (ii).[9] The parameter restriction of condition (iii) is satisfied for the point estimates ($1.005 < 1.113$).

Table 1 presents formal statistical tests for conditions (i)–(iii) and for proposition 1.

The Pareto distribution in condition (i) offers a precise approximation of the distribution of firm sizes. The R^2 from estimating equation (4) is 98.1 percent. The shape parameter λ is precisely estimated and is close to the unity Zipf's law benchmark ($\lambda = 1.0048$, standard error [SE] 0.0213). The relationship between firm size and the average squared distance of exports is close to the log-linear relation of condition (ii). The R^2 from estimating equation (5) is 81.7 percent. The elasticity parameter μ is precisely estimated ($\mu = 0.113$, SE 0.0078). To test the parameter restriction $\lambda < 1 + \mu$ of condition (iii), I form the probability $\Pr(\lambda \geq 1 + \mu)$ by estimating the parameters λ and μ from 10,000 bootstrapped samples. Condition (iii) is violated for 0.02 percent of the estimates only.

[9] The straight lines in the top and bottom panels of fig. 1 correspond to fitted regression lines from estimating eqq. (4) and (5), respectively.

The distance elasticity of aggregate trade is $\zeta_{\text{long}} = 1.090$ (SE = 0.215, $R^2 = .720$) when using "long" distances only (>2,000 km) and $\zeta_{\text{all}} = 0.767$ (SE = 0.111, $R^2 = .810$) when using all distances. The predicted distance elasticity of trade is $1 + 2(\lambda - 1)/\mu = 1.086$ (SE = 0.520), where I use 10,000 bootstrapped estimates to compute the standard error. It is very close to the actual distance elasticity for long distances but larger than the elasticity for all distances. To formally test the equality between the actual and predicted distance elasticities of trade, I form a Wald test for the equality $\zeta = 1 + 2(\lambda - 1)/\mu$. The p-value for this χ^2 test is high (99.4 percent for long distances and 99.3 percent for all distances), so that the actual and predicted distance elasticities of trade are statistically indistinguishable from each other. The data do not reject proposition 1.

D. *Testing Proposition 1 across Sectors*

Proposition 1 also predicts that variations in the Pareto shape parameter for the distribution of firm sizes (λ) and in the size elasticity of the average squared distance of firms' exports (μ) should be associated with systematic variations in the distance elasticity of trade (ζ). To test this prediction, I exploit variations in those parameters across sectors. I estimate the parameters λ_s, μ_s, and ζ_s from equations (1)–(6) separately for each sector s, where ζ_s is estimated using trade over long distances only (>2,000 km), and estimate the following equation:

$$\zeta_s = \alpha + \beta\left(2\frac{\lambda_s - 1}{\mu_s}\right) + \epsilon_s. \tag{7}$$

To get unbiased estimates for α and β and account for estimation error in my right-hand-side variable, I use a generalized least squares method (GLS) and weigh each observation $2(\lambda_s - 1)/\mu_s$ by the inverse of its bootstrapped variance, where I use 100 bootstraps for each sector. To prevent my results from being driven by precisely estimated but extreme outliers, all coefficients are 90 percent winsorised.

Figure 2 offers a visual representation of equation (7), a scatter plot of the sectoral distance elasticity of trade, ζ_s, as a function of $2(\lambda_s - 1)/\mu_s$. The straight line on figure 2 is the fitted line of a GLS estimation of equation (7). The raw correlation between the two variables is 31 percent, and the weighted correlation is 89 percent, where each observation $2(\lambda_s - 1)/\mu_s$ is weighted by the inverse of its estimated (bootstrapped) variance, as in the GLS estimation of equation (7). Variations across sectors in the Pareto distribution of firm sizes (λ_s) and the size elasticity of the average squared distance of exports (μ_s) are systematically associated with variations in the distance elasticity of trade (ζ_s), in the direction the theory predicts.

Formally, the parameter estimates from a GLS regression of equation (7) are $\alpha = 1.113$ (SE = 0.0073) and $\beta = 0.113$ (SE = 0.007), and

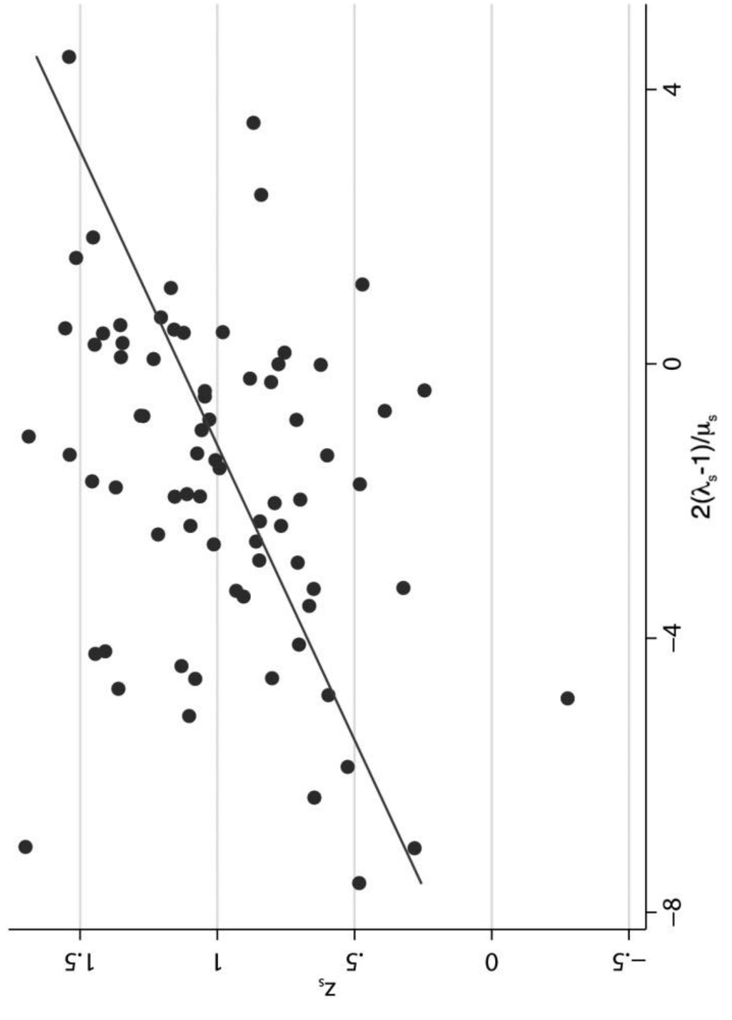

FIG. 2.—Proposition 1 across sectors

the R^2 is 79 percent. Both estimates α and β are positive and significant at the 1 percent level, as the theory predicts. A literal interpretation of proposition 1 predicts $\alpha, \beta = 1$. This quantitative implication of proposition 1 is only partially confirmed empirically. While the estimated α is close to one, β is close to 1/9, substantially below the predicted $\beta = 1$. There are several possible explanations for this discrepancy between the theory and the data. First, estimates of the parameters λ and μ are only imprecisely estimated when the data are disaggregated into sectors, as many sectors contain only a small number of firms. This may cause a downward bias in the estimate of β. The GLS method corrects only partially for this attenuation bias. Second, the theory behind proposition 1 is extremely stylized. It abstracts from many other factors that may affect the patterns of international trade. In trying to stay as close as possible to the theory, I have estimated the most minimalist gravity equation, controlling only for size and distance. It is possible that other elements I have purposefully left out of proposition 1 attenuate its sharp prediction and bias the estimates of β downward.

IV. A Dynamic Model of Firm-to-Firm Trade

In this section, I present a dynamic model of firm-to-firm trade in which conditions (i) and (ii) in proposition 1 emerge endogenously, and the parameter restriction (iii) has a structural interpretation. This model is stylized. It is meant as an illustration of a mechanism that generates sufficient conditions for the gravity equation in international trade.

A. *Setup*

The world is made of a continuum of locations. Within each location, there is a continuum of firms, born at a constant rate. Once born, a firm gradually acquires new trading partners in increasingly remote locations. A firm trades intermediate inputs with all its trading partners (buys from suppliers and sells to customers). Older firms have more trading partners so that they are larger and export more and over longer distances. I take all the parameters governing the dynamic evolution of firms as well as the determinants of the amounts traded as exogenous. In the online appendix, I offer a micro-founded model in which those parameters are determined endogenously. As this micro-founded model is not the primary focus of this paper, I have decided to relegate it to the online appendix for the benefit of the curious reader only.[10]

[10] The model is one in which firms combine equipped labor with differentiated intermediate inputs. As in Romer (1990), the diversity of a firm's suppliers acts in a way similar to physical capital. As firms "invest" into acquiring information about new suppliers, they be-

Space.—Firms are uniformly distributed over an infinite one-dimensional continuous space represented by $\mathbb{R}$. While this topology is not a good description of the actual world, I nonetheless generate rich predictions regarding the geography of firm-level and aggregate trade. As the model will be symmetric, I focus my attention on a firm located at the origin, that is, coordinate $x = 0$.

Time.—Time is continuous, with new firms born at a rate γ in each location, so at time t, there is the same density of firms $e^{\gamma t}$ in every location. This parameter γ is endogenously determined from the free entry of new firms in the online appendix.

Birth and death of a firm.—When a firm is born, it samples a mass K_0 of contacts among other newborn firms only. This symmetry assumption is strong but seems a good compromise between tractability and keeping rich predictions for firm heterogeneity.

The K_0 contacts of a newborn firm are distributed geographically according to the density function $k_0(\cdot)$: the mass of contacts in the interval $[a, b]$ is $\int_a^b k_0(x)dx$. I assume that k_0 is symmetric and has a finite second moment but can take any arbitrary shape otherwise.

Firms are infinitely lived. All results carry through with exogenous Poisson death shocks.

Birth and death of contacts.—New contacts are continuously created. At any point in time, each existing contact may reveal one of its own contacts according to a Poisson process with arrival rate β. In other words, a firm directly learns about new contacts from the contacts of its existing contacts. This corresponds to what I call remote search in Chaney (2014). This parameter β is endogenously determined from the optimal decision to acquire new contacts in the online appendix.

Existing contacts are continuously lost to an exogenous Poisson shock with rate δ.

Firm-to-firm trade.—As a result of informational frictions, a firm sells its output only to its existing contacts. I normalize the value of individual shipments to one for all firms for simplicity. The number of contacts of a firm, K, is therefore also a measure of its size. The main conclusions of the model are robust to allowing an active intensive margin of shipments, for example, with larger firms sending not only more but also larger shipments. I derive the value of shipments endogenously and the conditions under which all shipments have equal size in the online appendix.

come larger and more efficient. I show how to use the analogy between physical capital and diversity of suppliers to rephrase the dynamic problem of the firm as a classical Lucas (1967) model of investment. The model delivers endogenous entry of new firms at a constant rate and Gibrat's law, a growth rate of individual firms independent of size. I characterize along a steady-state growth path the equilibrium of an economy made up of increasingly complex vertical production chains.

I assume $\gamma > \beta - \delta > 0$. The assumption $\beta - \delta > 0$ is not required, but it rules out the counterfactual prediction of infinitesimally small firms and firm sizes shrinking over time. The assumption $\gamma > \beta - \delta$ rules out a degenerate equilibrium in which older firms become "too" large.

I now define two variables: The function k_a describes the geographic distribution of the contacts of a firm of age a, and K_a describes the total mass of contacts of this firm,

$$k_a : \mathbb{R} \to \mathbb{R}^+, \quad K_a \equiv \int_{\mathbb{R}} k_a(x)\,dx,$$

where $k_a(x)$ is the density of contacts a firm of age a has in location x: The mass of contacts in the interval $[a, b]$ is $\int_a^b k_a(x)\,dx$. The total mass of contacts a firm of age a has worldwide is K_a.

The distribution of contacts evolves recursively according to the partial differential equation (PDE)

$$\frac{\partial}{\partial a} k_a(x) = \beta \int_{\mathbb{R}} \frac{k_a(x-y)}{K_a} k_a(y)\,dy - \delta k_a(x), \tag{8}$$

with the initial condition $k_0(x)$. Multiplying both sides by $dxda$, (8) describes the net creation of new contacts in a neighborhood dx of location x over a short time interval da. Any existing contact of the firm (there are $k_a(y)\,dy$ of them in each neighborhood dy of y) reveals with probability βda one of their contacts. This contact happens to be in a neighborhood dx of x with probability $[k_a(x-y)/K_a]\,dx$.[11] To count all newly acquired contacts, I add the names coming from all possible sources (I integrate y over $\mathbb{R}$), and I remove the $\delta k_a(x)\,dxda$ old contacts exogenously lost.

With continuous time, space, and a continuous measure of contacts, I have removed all sources of randomness from the model. Age fully determines size and the geography of trade. A discrete and stochastic version of this model would break the tight link between age and firm characteristics.

I show in the next section that the law of motion (8) for the distribution of a firm's contacts implies that conditions (i) and (ii) of proposition 1 are exactly satisfied. But before doing so, I present one intermediate result that sheds light on the mechanics of the model.

[11] The distribution of contacts for a firm of age a located in y, call it $k_{y,a}$, is the same as that of a firm of the same age a located in the origin $(y = 0)$, $k_{0,a}$, except all coordinates are simply shifted by the constant $-y$: $k_{0,a}(x) = k_{y,a}(x-y) = k_a(x)$. I rely here on the simplifying assumption that a firm of age a meets only other firms in the same cohort, which themselves have the same distribution k_a. I show in the online appendix how this strong assumption can be substantially relaxed.

Proposition 2. For any initial distribution k_0 that is symmetric and admits a finite second moment, the normalized distribution of contacts of a firm of age a, $f_a = k_a/K_a$, converges when age a grows large to a Laplace distribution (a two-sided exponential),

$$f_a(x) \underset{a\to\infty}{\sim} \frac{1}{2\sqrt{\Delta_0/2}e^{\beta a/2}} \exp\left(-\frac{|x|}{\sqrt{\Delta_0/2}e^{\beta a/2}}\right).$$

This property holds exactly for all a's if

$$f_0(x) = \frac{1}{2\sqrt{\Delta_0/2}} \exp\left(-\frac{|x|}{\sqrt{\Delta_0/2}}\right).$$

Proof. To save on notation, normalize units so that $\sqrt{\Delta_0/2} = 1$. Integrate over $\mathbb{R}$ the law of motion (8) for a firm's contacts to get a simple ordinary differential equation (ODE) for K_a, $(\partial/\partial a)K_a = (\beta - \delta)K_a$ with initial condition K_0, which admits the solution $K_a = K_0 e^{(\beta-\delta)a}$. Use this solution for K_a and simple manipulations to derive from (8) a PDE for the normalized distribution of contacts $f_a = k_a/K_a$:

$$\frac{\partial}{\partial a} f_a(x) = \beta\left[\int_{\mathbb{R}} f_a(x-y) f_a(y)\,dy - f_a(x)\right], \tag{9}$$

with initial condition $f_0 = k_0/K_0$. Taking a Fourier transform of (9), with the notation $\hat{f}$ for the transform of f, recognizing a convolution product in the integral on the right-hand side of (9), and using the convolution theorem gives a simple ODE for $\hat{f}_a$,[12]

$$\frac{\partial}{\partial a}\hat{f}_a = \beta \hat{f}_a(\hat{f}_a - 1), \tag{10}$$

with initial condition $\hat{f}_0$. Introducing $h(e^{\beta a/2} is, a) = \hat{f}_a(s)$, simple algebra gives

$$\frac{\partial}{\partial a} h(y, a) = \beta\{[h(y,a)]^2 - h(y,a)\} - \frac{\beta}{2} y \frac{\partial}{\partial y} h(y, a).$$

[12] The Fourier transform of a function f is defined as $\hat{f}(s) = \int_{\mathbb{R}} e^{-isx} f(x)\,dx$. If f is the probability density function (pdf) of a random variable X, it is intimately related to the characteristic function: $\varphi_X(s) = \mathbb{E}[e^{isX}] = \hat{f}(-s)$. The convolution $f * g$ of two functions f and g is defined by $f * g(x) \equiv \int_{\mathbb{R}} f(x) g(y - x)\,dy$. Remember also that the pdf of the sum of two random variables is the convolution of their pdfs. The convolution theorem states that the Fourier transform of the convolution product of two functions is the pointwise product of their Fourier transforms, $\widehat{f * g}(s) = \hat{f}(s) \cdot \hat{g}(s)$.

From lemma 2 in the appendix, $\lim_{a\to+\infty}(\partial/\partial a)h(y, a) = 0$. So as $a \to +\infty$, $h(y, a)$ is defined by

$$[h(y, a)]^2 - h(y, a) = \frac{1}{2} y \frac{\partial}{\partial y} h(y, a).$$

This ODE admits the solution $h(y, a) = 1/(1 - y^2)$. Recognizing the transform of a Laplace distribution,

$$\hat{f}_a(s) \underset{a\to\infty}{\sim} \frac{1}{1 - (e^{\beta a/2} is)^2} = \frac{1}{1 + (e^{\beta a/2} s)^2} \qquad \forall s$$
$$\Leftrightarrow f_a(x) \underset{a\to\infty}{\sim} \frac{1}{2e^{\beta a/2}} \exp\left(-\frac{|x|}{e^{\beta a/2}}\right) \qquad \forall x.$$

In the special case in which the initial distribution of contacts is exactly a two-sided exponential, $f_0(x) = 1/2 \exp(-|x|)$, a simpler guess and verify proof is in order. Guess $f_a(x) = 1/2\sigma(a) \exp(-\sigma(a)|x|)$ for some $\sigma(a)$ to be determined; insert this guess into (9); calculating some easy integral results in the ODE $\sigma'(a) = -(\beta/2)\sigma(a)$, which is solved by $\sigma(a) = e^{-\beta a/2}$. QED

The intuition for proposition 2 is as follows. As a firm grows larger, it meets the contacts of its contacts. Information about distant contacts diffuses through this network of firm-to-firm trade. Any individual firm gradually escapes gravity. The distribution of its contacts converges to what resembles a uniform distribution over the entire real line: For any two locations x and y, no matter how far apart from each other, the fractions of contacts in x and in y become equal for a large. In other words, the world does become "flat" for individual firms as they grow large. But, as the reader can already guess, this does not mean the world becomes "flat" in the aggregate.

B. *Three Sufficient Conditions for Gravity*

The next proposition shows formally that in my model of firm-to-firm trade, the distribution of firm sizes is exactly Pareto, as in condition (i) of proposition 1, and the average squared distance of exports is exactly a power function of firm size, as in condition (ii) of proposition 1.

Proposition 3. If the population of firms grows at a constant rate γ and the contacts of individual firms evolve according to equation (8), then the distribution of firm sizes is Pareto,

$$F(K) = 1 - \left(\frac{K}{K_0}\right)^{-\gamma/(\beta-\delta)} \qquad \text{for } K \geq K_0, \tag{11}$$

and the average squared distance from a firm's contacts is a power function of its size,

$$\Delta(K) \equiv \int_{\mathbb{R}} x^2 f_K(x)\,dx = \Delta_0\left(\frac{K}{K_0}\right)^{\beta/(\beta-\delta)}, \tag{12}$$

where f_K is the distribution of contacts of a firm with K contacts ($f_K = k_{a(K)}/K$ with $a(K)$ such that $K_{a(K)} = K$) and $\Delta_0 \equiv \int_{\mathbb{R}} x^2 f_0(x)\,dx$ is the average squared distance from initial contacts.

Proof. I derive each equation in turn.

Equation (11): From the proof of proposition 2, $K_a = K_0 e^{(\beta-\delta)a}$. The relation between a firm's number of contacts K and its age a is therefore given by $e^a = (K_a/K_0)^{1/(\beta-\delta)}$. The population grows at an exponential rate γ, so that at any time t, the fraction of firms younger than a is $(1 - e^{-\gamma a})$. Since a firm of age a has a total number of contacts K_a, use the above expression for e^a to get the proposed formula for the fraction of firms with fewer than K contacts:[13]

$$F(K) = 1 - \left(\frac{K}{K_0}\right)^{-\gamma/(\beta-\delta)}.$$

Equation (12): The average squared distance between a firm of age a and its contacts, Δ_a, is the second moment of the normalized density of contacts, $f_a = k_a/K_a$. Using the property of the Fourier transform, this second moment is simply minus the second derivative of $\hat{f}_a(s)$ evaluated at zero. The ODE (10) for $\hat{f}_a$ with initial condition $\hat{f}_0$ admits the explicit solution

$$\hat{f}_a(s) = \frac{\hat{f}_0(s)}{\hat{f}_0(s) + [1 - \hat{f}_0(s)]e^{\beta a}}. \tag{13}$$

Simple algebra gives the second derivative of $\hat{f}_a(s)$,

$$\hat{f}_a''(s) = \frac{e^{\beta a}\left\{\hat{f}_0''(s)\left[(e^{\beta a} - 1)\hat{f}_0 - e^{\beta a}\right] - 2\hat{f}_0'(s)^2(e^{\beta a} - 1)\right\}}{\left[\left(e^{\beta a} - 1\right)\hat{f}_0(s) - e^{\beta a}\right]^3}.$$

Since f_0 is a well-defined symmetric pdf with second moment Δ_0, I use the following properties of its Fourier transform: $\hat{f}_0(0) = \int_{\mathbb{R}} f_0(x)\,dx = 1$ (a pdf sums up to one), $\hat{f}_0'(0) = (-i)\int_{\mathbb{R}} x f_0(x)\,dx = 0$ (f_0 is symmetric), and $\hat{f}_0''(0) = (-i)^2 \int_{\mathbb{R}} x^2 f_0(x)\,dx = -\Delta_0$ (f_0's finite second moment is Δ_0). The previous expression evaluated at zero simplifies into

$$\Delta_a = -\hat{f}_a''(0) = \Delta_0 e^{\beta a}.$$

[13] Trivially in this model, the distribution of firm sizes is time invariant: F does not depend on calendar time t.

Plug the expression $e^a = (K_a/K_0)^{1/(\beta-\delta)}$ into the above formula for Δ_a to derive the proposed

$$\Delta(K) = \Delta_0\left(\frac{K}{K_0}\right)^{\beta/(\beta-\delta)}.$$

QED

As a firm ages, the number (mass) of its contacts grows at a constant rate equal to the net birthrate of contacts (birthrate β minus death date δ), $K_a = K_0 e^{(\beta-\delta)a}$. Both the number of a firm's contacts and the number of firms grow exponentially, and the model predicts that the distribution of the number (mass) of contacts within the population is Pareto with shape parameter $\gamma/(\beta-\delta)$. The upper tail of the distribution of firm sizes is fatter ($\gamma/(\beta-\delta)$ smaller) if there are more old/large firms relative to young/small ones (γ smaller) or if firm size increases faster with age ($\beta-\delta$ larger).

Note there is nothing mysterious or very elaborate about this result: A constant growth rate of existing firms (Gibrat's law) combined with a constant growth rate of the population of firm sizes is probably the simplest way to generate an invariant Pareto distribution of firm sizes. This corresponds exactly to the Steindl (1965) model. Note also that I do not offer a direct justification for why Zipf's law ($\gamma/(\beta-\delta) \approx 1$) is a better candidate than any other Pareto distribution. Several explanations for Zipf's law have already been proposed, and I refer to them for this result.[14]

The model also predicts that as a firm ages and acquires more contacts, those contacts become increasingly dispersed over space. The intuition for this result is as follows. A firm's initial contacts are some distance away. Each wave of new contacts comes from firms that are themselves further away, so each new wave is geographically more dispersed than the previous one. The average squared distance of exports increases faster with size ($\gamma/(\beta-\delta)$ larger) if firms meet new contacts of contacts at a faster rate (β larger) or if their size grows at a slower rate ($\beta-\delta$ smaller).

Formally, each time a firm meets the "contacts of its contacts," the new average squared distance of the firm's exports becomes the sum of the existing squared distance of the firm's exports and the average squared distance of exports of the firm's contacts. The reason why average squared distances are simply added to each other can be seen from the law of mo-

[14] See in particular the stochastic models in Gabaix (1999) or Luttmer (2011), which deliver endogenously an invariant size distribution that is close to Zipf's law. I choose to use simpler tools to derive the distribution of firm sizes while adding substantial complexity on the geographic dimension of the model. But for the addition of geographic space and the removal of many stochastic elements, my model is close to a network interpretation of Luttmer's model, where firms innovate both on their own and by learning from other firms; if the second channel for innovation (learning from each other) dominates, Luttmer's model behaves similarly to mine.

tion of the firm's contacts in the PDE (9). Mathematically, the integral over y of the function $f_a(y)$ multiplied by $f_a(x - y)$ is a convolution product of the function f_a with itself. In probability theory, the convolution product is used to characterize the probability density of the sum of two random variables. In my model, a firm located in 0 meets a contacts in x via its existing contact in y, so that I simply add up vectors: $\overrightarrow{x - 0} = \overrightarrow{y - 0} + \overrightarrow{x - y}$. If one thinks of the signed distance of a firm's exports as being drawn from a random variable, then the average squared distance of the firm's exports is simply the variance of that random variable. Equation (9) says that the average squared distances get added up each period, just as the variance of the sum of two random variables is the sum of their variances.[15] The firm's contacts themselves are each forming trade links with the contacts of their own contacts, the average squared distance of their exports is also the sum of their existing squared distance of exports and that of their contacts, and so on. From one period to the next, the increase in the average squared distance of exports is proportional to the number of a firm's contacts. The average squared distance of exports grows exponentially over time. Since firm size also grows exponentially, the average squared distance of exports is a power function of size.

The key assumption necessary for equation (12) to hold is that a firm forms new trade links with the contacts of its existing contacts or, more generally, that it learns about new trading opportunities from its existing trading partners. Any model that features such a diffusion of information will be such that the distance of exports grows over time, and the world becomes flat for individual firms as they get large. If this information diffusion process follows an exponential growth (firms meet new trading partners at a constant rate), then the firm size distribution is Pareto and the average squared distance of exports is precisely a power function of firm size.

Both conditions (i) and (ii) of proposition 1 are derived endogenously, with $\gamma/(\beta - \delta)$ the solution for the shape parameter of the Pareto distribution of firm sizes (λ in proposition 1) and $\beta/(\beta - \delta)$ the elasticity of the average squared distance of exports with respect to firm size (μ in proposition 1). Furthermore, the parameter restriction (iii) now has a structural interpretation: The entry rate of new firms (γ) should not exceed the sum of the growth rates of individual firms ($\beta - \delta$) and the gross creation of new contacts (β),

[15] This adding up explains why the law of motion for the location of contacts in my model, my eq. (9), is exactly identical to the law of motion for the posterior about the type of a payoff-relevant variable in Duffie and Manso (2007, 205, eq. [3]): In my model, information about locations percolates when firms trade with each other, so that signed distances are added up; in their model, information about an asset percolates when agents meet, so that priors are averaged to form posteriors, i.e., they get added up (and divided by 2).

$$\gamma < (\beta - \delta) + \beta. \tag{14}$$

To satisfy this restriction, it is enough that the process for creating new contacts exhibits a lot of churning: If firms gain and lose contacts often (β and δ are large), then even if the growth rate of individual firms ($\beta - \delta$) does not differ much from the entry rate of new firms (γ), as is the case in the data, the parameter restriction (iii) will be satisfied.

To recap, under the structural parameter restriction (14), conditions (i)–(iii) of proposition 1 are satisfied, and the gravity equation for international trade emerges endogenously.

The model presented above shares many features of modern heterogeneous firm trade models such as Melitz (2003). It is populated by heterogeneous firms of various sizes. Larger firms are more productive and export more toward more countries and toward more remote countries. The distribution of firm sizes is Pareto as in Chaney's (2008) extension of the Melitz model. But on a more conceptual level, this model departs from traditional trade models in its treatment of distance and trade barriers. In existing models, distance captures or proxies physical trade barriers, with a direct mapping from the geography of trade barriers to the geography of trade. Unless the geography of trade barriers is time invariant, such models cannot explain why distance plays the same role today as it did a century ago. In my model, on the other hand, distance captures informational barriers and the network that transmits information.[16] Unlike physical trade barriers, informational barriers can be circumvented indirectly when people interact and share information. Advances in transportation or communication technologies affect the direct cost of information (the function f_0), even the frequency of interactions between firms (the parameters γ, β, and δ). The patterns of trade at the firm level do change with f_0, γ, β, and δ, which I of course expect to happen along with technological progress. But as long as the distribution of firm sizes remains close to Zipf's law, the patterns of aggregate trade flows remain essentially unchanged.

V. Conclusion

This paper offers a theoretical explanation for the gravity equation in international trade and in particular the mysterious -1 distance elasticity of trade. If larger firms export over longer distances than small ones, then the impact of distance on aggregate trade depends on the distribution of firm sizes. If firm sizes are well approximated by Zipf's law and if the average squared distance of firms' exports is a power function of firm size, as the data suggest, then the distance elasticity of trade ought to be close

[16] See Allen (2014) and Dasgupta and Mondria (2014) for recent trade models with information frictions.

to -1. This result holds irrespective of the precise impact of geographic distance on firm-level trade. In contrast to existing models, this explanation is immune to the critique that the impact of distance on trade should evolve with changes in the technology for trading goods, in the types of goods traded, in the political barriers to trade, in the set of countries involved in trade, and so forth. As long as the distribution of firm sizes is stable and larger firms export over longer distances than smaller ones, aggregate trade should be close to inversely proportional to distance.

Mathematical Appendix

LEMMA 1. The function $g_K(x) \equiv K^{\mu/2} f_K(K^{\mu/2} x)$, where f_K satisfies the conditions in proposition 1, is such that

$$\lim_{x \to +\infty} \int_0^x u^{2(\lambda-1)/\mu} g_{(x/u)^{\mu/2}}(u)\, du < \infty.$$

Proof. Defining $g_K(x) = 0$ for $K \in [0, 1)$, the integral of interest can be split into two parts:

$$\begin{aligned}\int_0^x u^{2(\lambda-1)/\mu} g_{(x/u)^{2/\mu}}(u)\, du &= \int_0^\infty u^{2(\lambda-1)/\mu} g_{(x/u)^{2/\mu}}(u)\, du \\ &= \int_0^1 u^{2(\lambda-1)/\mu} g_{(x/u)^{2/\mu}}(u)\, du + \int_1^\infty u^{2(\lambda-1)/\mu} g_{(x/u)^{2/\mu}}(u)\, du.\end{aligned}$$

For $0 \le u \le 1$, from $\lambda \ge 1$ in condition (i) and $f_K(x)$'s boundedness in condition (ii),

$$u^{2(\lambda-1)/\mu} g_{(x/u)^{2/\mu}}(u) \le g_{(x/u)^{2/\mu}}(u) \le \sup_{x,K}\{g_K(x)\}$$

and

$$\int_0^1 \sup_{x,K} g_K(x)\, du < \infty.$$

Integrate $\int_0^x u^2 g_K(u)\, du$ by part to get

$$g_K(x) = \frac{3}{x^3}\left[\int_0^x u^2 g_K(u)\, du + \int_0^x \frac{u^3}{3} g_K'(u)\, du\right],$$

with $\int_0^x u^2 g_K(u)\, du \le 1$ from g_K's second moment equal to one,

$$\int_0^{\bar{x}} \frac{u^3}{3} g_K'(u)\, du \le \frac{\bar{x}^4}{12} \sup_{x,K}\{g_K'(x)\}$$

from $f_K'(x)$ boundedness in condition (ii), and $\int_{\bar{x}}^x (u^3/3) g_K'(u)\, du \le 0$ from $f_K(x)$ weakly decreasing above $\bar{x}$ in condition (ii). So for all $u \ge 1$,

$$u^{2[(\lambda-1)/\mu]} g_{(x/u)^{2/\mu}}(u) \le A u^{2[(\lambda-1)/\mu]-3},$$

where I define

$$A \equiv 3\left(1 + \frac{\bar{x}^4}{12}\sup_{x,K}\{g_K(x)\}\right) < \infty.$$

From $\lambda < 1 + \mu$ in condition (iii),

$$\int_1^\infty Au^{2[(\lambda-1)/\mu]-3}du < \infty.$$

Invoking Lebesgue's dominated convergence theorem,

$$\begin{aligned}&\lim_{x\to+\infty}\int_0^x u^{2(\lambda-1)/\mu}g_{(x/u)^{\mu/2}}(u)du\\ &= \int_0^1 \lim_{x\to+\infty} u^{2(\lambda-1)/\mu}g_{(x/u)^{\mu/2}}(u)du\\ &+ \int_1^\infty \lim_{x\to+\infty} u^{2(\lambda-1)/\mu}g_{(x/u)^{\mu/2}}(u)du < \infty,\end{aligned}$$

as proposed. QED

LEMMA 2. If $\hat{f}_a$, the Fourier transform of the density of contacts f_a, is governed by the ODE (10) with initial condition $\hat{f}_0$; if the pdf f_0 is symmetric and admits a finite second moment; if h is defined as $h(e^{\beta a/2}is, a) = \hat{f}_a(s)$, then

$$\lim_{a\to+\infty}\frac{\partial}{\partial a}h(y, a) = 0.$$

Proof. Using the solution to the ODE (10) with initial condition $\hat{f}_0$ in the proof of proposition 3, equation (13), and using $h(e^{\beta a/2}is, a) = \hat{f}_a(s)$, the solution for $h(y, a)$ is

$$h(y, a) = \frac{\hat{f}_0(ye^{-\beta a/2})}{\hat{f}_0(ye^{-\beta a/2}) + [1 - \hat{f}_0(ye^{-\beta a/2})]e^{\beta a}}.$$

With the change of variable $u = e^{-\beta a/2}$, the object I will take the limit of is given by

$$\frac{\partial}{\partial a}h(y, a) = -\frac{\beta}{2}u\frac{\partial}{\partial u}h\left(y, -\frac{2}{\beta}\ln u\right).$$

From the solution for h, I get

$$h\left(y, -\frac{2}{\beta}\ln u\right) = \frac{\hat{f}_0(uy)}{\hat{f}_0(uy) + [1 - \hat{f}_0(uy)]u^{-2}}$$

so that

$$\begin{aligned}-\frac{\beta}{2}u\frac{\partial}{\partial u}h\left(y, -\frac{2}{\beta}\ln u\right) = -\frac{\beta}{2}\Big(&y\hat{f}_0'(uy)u\{\hat{f}_0(uy) + [1 - \hat{f}_0(uy)]u^{-2}\}\\ &-\hat{f}_0(uy)\{uy\hat{f}_0'(uy) - 2u^{-2}[1 - \hat{f}_0(uy)] - yu^{-1}\hat{f}_0'(uy)\}\Big)\\ &\div\left(\{\hat{f}_0(uy) + [1 - \hat{f}_0(uy)u^{-2}]\}^2\right).\end{aligned}$$

Taking limits for $u \to 0$ and using the following results, $\hat{f}_0(0) = \int_{\mathbb{R}} f_0(x)dx = 1$ because f_0 is a well-defined pdf that sums to one, $\hat{f}_0'(0) = (-i)\int_{\mathbb{R}} xf_0(x)dx = 0$ because f_0 is symmetric, and $\hat{f}_0''(0) = (-i)^2\int_{\mathbb{R}} x^2 f_0(x)dx = -\Delta_0 > -\infty$ because f_0 admits a finite second moment, I have the following:

- $\lim_{u\to 0} y\hat{f}_0'(uy) = y\hat{f}_0'(0) = 0.$
- Using L'Hopital's rule,

$$\begin{aligned}\lim_{u\to 0}\hat{f}_0(uy) + [1-\hat{f}_0(uy)]u^{-2} &= \hat{f}_0(0) + \lim_{u\to 0}\frac{1-\hat{f}_0(uy)}{u^2}\\ &= 1 + y\lim_{u\to 0}\frac{\hat{f}_0'(uy)}{2u}\\ &= 1 + \frac{y^2}{2}\lim_{u\to 0}\frac{\hat{f}_0''(uy)}{1}\\ &= 1 - \frac{y^2}{2}\hat{f}_0''(0)\end{aligned}$$

  positive and finite for y small enough.[17]
- $$\begin{aligned}&\lim_{u\to 0} -\hat{f}_0(uy)\Big\{uy\hat{f}_0'(uy) - 2u^{-2}\big[1-\hat{f}_0(uy)\big] - yu^{-1}\hat{f}_0'(uy)\Big\}\\ &\quad = 0 + 2\lim_{u\to 0}\frac{1-\hat{f}_0(uy)}{u^2} + y\lim_{u\to 0}\frac{\hat{f}_0'(uy)}{u}\\ &\quad = y^2\hat{f}_0''(0) + y^2\hat{f}_0''(0) = 0,\end{aligned}$$

  where I use L'Hopital's rule again.

Collecting all terms, I confirm

$$\lim_{a\to +\infty}\frac{\partial}{\partial a}h(y,a) = \lim_{u\to 0} -\frac{\beta}{2}u\frac{\partial}{\partial u}h\left(y, -\frac{2}{\beta}\ln u\right) = 0.$$

QED

References

Allen, T. 2014. "Information Frictions in Trade." *Econometrica* 82:2041–83.

Anderson, J. E. 1979. "A Theoretical Foundation for the Gravity Equation." *A.E.R.* 69:106–16.

Anderson, J. E., and E. van Wincoop. 2003. "Gravity with Gravitas: A Solution to the Border Puzzle." *A.E.R.* 93:170–92.

Arkolakis, C., A. Costinot, and A. Rodriguez-Clare. 2012. "New Trade Model, Same Old Gains?" *A.E.R.* 102:94–130.

Axtell, R. L. 2001. "Zipf Distribution of U.S. Firm Sizes." *Science* 293:1818–20.

[17] I only need to characterize $\hat{f}_a(s)$ for s in the vicinity of zero to recover the pdf f_a.

Bergstrand, J. 1985. "The Gravity Equation in International Trade: Some Microeconomic Foundations and Empirical Evidence." *Rev. Econ. and Statis.* 67:474–81.

Bernard, A. B., J. Eaton, B. J. Jensen, and S. Kortum. 2003. "Plants and Productivity in International Trade." *A.E.R.* 93:1268–90.

Chaney, T. 2008. "Distorted Gravity: The Intensive and Extensive Margins of International Trade." *A.E.R.* 98:1707–21.

———. 2014. "The Network Structure of International Trade." *A.E.R.* 104:3600–3634.

Dasgupta, K., and J. Mondria. 2014. "Inattentive Importers." Manuscript, Dept. Econ., Univ. Toronto.

Duffie, D., and G. Manso. 2007. "Information Percolation in Large Markets." *A.E.R. Papers and Proc.* 97:203–9.

Eaton, J., and S. Kortum. 2002. "Technology, Geography and Trade." *Econometrica* 70:1741–79.

Eaton, J., S. Kortum, and S. Sotelo. 2012. "International Trade: Linking Micro and Macro." Technical report, NBER, Cambridge, MA.

Feenstra, R., R. E. Lipsey, H. Deng, A. C. Ma, and H. Mo. 2004. "World Trade Flows: 1962–2000." Working Paper no. 11040, NBER, Cambridge, MA.

Gabaix, X. 1999. "Zipf Law for Cities: An Explanation." *Q.J.E.* 114:739–67.

———. 2008. "Power Laws." In *The New Palgrave Dictionary of Economics*, 2nd ed., edited by S. N. Durlauf and L. E. Blume. London: Palgrave Macmillan.

Head, K., and T. Mayer. 2014. "Gravity Equations: Workhorse, Toolkit and Cookbook." In *Handbook of International Economics*, vol. 4, edited by E. H. Gita Gopinath and K. Rogoff. Amsterdam: Elsevier.

Helpman, E., M. J. Melitz, and Y. Rubinstein. 2008. "Estimating Trade Flows: Trading Partners and Trading Volumes." *Q.J.E.* 123:441–87.

Krugman, P. 1980. "Scale Economies, Product Differentiation, and the Patterns of Trade." *A.E.R.* 70:950–59.

Lucas, R. E., Jr. 1967. "Adjustment Costs and the Theory of Supply." *J.P.E.* 75:321–34.

Luttmer, E. G. 2011. "On the Mechanics of Firm Growth." *Rev. Econ. Studies* 78:1042–68.

Mayer, T., and S. Zignago. 2006. "Notes on CEPII's Distances Measures." Manuscript, Centre d'Etudes Prospectives et d'Informations Internationales, Paris.

Melitz, M. J. 2003. "The Impact of Trade on Intra-industry Reallocation and Aggregate Industry Productivity." *Econometrica* 71:1695–1725.

Oberfield, E. Forthcoming. "A Theory of Input-Output Architecture." *Econometrica.*

Romer, P. M. 1990. "Endogenous Technological Change." *J.P.E.* 98, no. 5, pt. 2 (October): S71–S102.

Santos Silva, J. M. C., and S. Tenreyro. 2006. "The Log of Gravity." *Rev. Econ. and Statis.* 88:641–58.

Steindl, J. 1965. *Random Processes and the Growth of Firms.* London: Charles Griffin.

Tinbergen, J. 1962. "An Analysis of World Trade Flows." In *Shaping the World Economy.* New York: Twentieth Century Fund.