

Singapore Management University

## Institutional Knowledge at Singapore Management University

---

Research Collection School Of Economics

School of Economics

---

1-2020

### Geography, trade and power-law phenomena

Pao-Li CHANG

*Singapore Management University*, [plchang@smu.edu.sg](mailto:plchang@smu.edu.sg)

Wen-tai HSU

*Singapore Management University*, [WENTAIHSU@smu.edu.sg](mailto:WENTAIHSU@smu.edu.sg)

Follow this and additional works at: [https://ink.library.smu.edu.sg/soe\\_research](https://ink.library.smu.edu.sg/soe_research)



Part of the [Finance Commons](#)

---

#### Citation

CHANG, Pao-Li and HSU, Wen-tai. Geography, trade and power-law phenomena. (2020). *Oxford Research Encyclopedia of Economics and Finance*. 1-26. Research Collection School Of Economics.

Available at: [https://ink.library.smu.edu.sg/soe\\_research/2357](https://ink.library.smu.edu.sg/soe_research/2357)

This Encyclopaedia is brought to you for free and open access by the School of Economics at Institutional Knowledge at Singapore Management University. It has been accepted for inclusion in Research Collection School Of Economics by an authorized administrator of Institutional Knowledge at Singapore Management University. For more information, please email [libIR@smu.edu.sg](mailto:libIR@smu.edu.sg).

### Geography, Trade, and Power-Law Phenomena

Pao-Li Chang and Wen-Tai Hsu

Subject: Economic Theory and Mathematical Models , International Economics , Urban, Rural, and Regional Economics

Online Publication Date: Jan 2020 DOI: 10.1093/acrefore/9780190625979.013.296

### Summary and Keywords

This article reviews interrelated power-law phenomena in geography and trade. Given the empirical evidence on the gravity equation in trade flows across countries and regions, its theoretical underpinnings are reviewed. The gravity equation amounts to saying that trade flows follow a power law in distance (or geographic barriers). It is concluded that in the environment with firm heterogeneity, the power law in firm size is the key condition for the gravity equation to arise. A distribution is said to follow a power law if its tail probability follows a power function in the distribution's right tail. The second part of this article reviews the literature that provides the microfoundation for the power law in firm size and reviews how this power law (in firm size) may be related to the power laws in other distributions (in incomes, firm productivity and city size).

Keywords: gravity equation, power law, firm size, city size, geography

---

## Significance of Power-Law Phenomena and Related Reviews

Both gravity equation and power law in firm size are well documented empirical regularities (see Head & Mayer, 2015; Axtell, 2001; Luttmer, 2007; Gabaix, 2009). Because of its seemingly universal applicability, the power-law phenomenon that characterizes trade flows, incomes, firm and city sizes, and network linkages was dubbed by Krugman (1997) as social physics. The work by Arkolakis, Costinot, and Rodríguez-Clare (2012) further highlights that the power coefficient in the gravity equation (i.e., the partial elasticity of trade with respect to variable trade cost) is one of the two sufficient statistics to impute a country's welfare gains from trade for a large set of trade models. As mentioned, the power law in firm size is a key condition for the gravity equation in aggregate trade flows. Moreover, the heavy tail implication of the power law is also consistent with the granular economies phenomenon (Gabaix, 2011) (i.e., the few large firms may be what matters the most for macroeconomic performance). Thus, it is important to understand the plausible explanations for this power law in firm size to arise and how it is connected with other power laws.

For an extensive review of the literature on the gravity equation, see Head and Mayer (2015). This survey draws insights from Head and Mayer (2015) and extends the review to recent developments in theoretical trade modeling. We focus on identifying the critical conditions that lead to a gravity equation in aggregate trade flows and discuss the general applicability of the gravity equation in these alternative models of supply/demand side structures. For an extensive review of power-law phenomena in economics and finance, see Gabaix (2009). Our survey differs from Gabaix (2009) by including recent theoretical explanations for the power law in firm size that do not resort to firm dynamics. In particular, we discuss theories whereby power laws emerge in static environments, due to firm hierarchy, networks, innovation, or geography.

## Trade and Power-Law Phenomena

### Gravity Equation

Since the 1960s, international trade flows (Tinbergen, 1962) have been documented to follow the law of gravity, where the volume of trade increases with the economic size of the trading partners and decreases with their distance:

$$X_{ij} \propto \frac{Y_i E_j}{(\text{Distance}_{ij})^\zeta}$$

(1)

with  $X_{ij}$  denoting the exports from country  $i$  to country  $j$ , and  $Y_i$  and  $E_j$  the gross output and expenditure of the exporting and importing country, respectively. The distance term is typically interpreted as capturing all trade cost  $\tau_{ij}$  created by geographical distance as well as cultural, institutional, and policy barriers (such as similarity in language, religion, legal origin, colonial history, regional trade agreements, currency union, and membership in GATT/WTO). The empirical evidence on the gravity equation is not only abundant for international trade but also regional trade (e.g., Duranton, Morrow, & Turner, 2014; Monte, Redding, & Rossi-Hansberg, 2018; McCallum, 1995; Anderson & van Wincoop, 2003).

The empirical specification, initially thought to be an ad hoc invention until the early 2000, has subsequently been found to be consistent with almost all canonical trade models. These include: (i) Anderson and van Wincoop (2003)—with perfect competition and Dixit-Stiglitz constant elasticity of substitution (CES) preference for goods differentiated by countries of origin; (ii) Eaton and Kortum (2002)—with the Ricardian structure and CES preferences where countries' productivity ( $z$ ) for producing each good is characterized by an *i.i.d.* Fréchet distribution:  $F_i(z) = e^{-T_i z^{-\kappa}}$ ,  $\kappa > 1$ ,  $T_i \gg 0$ ; (iii) Krugman (1980)—with monopolistic competition and CES preference; and (iv) Melitz (2003) and Chaney (2008)—with the Krugman structure along with firm heterogeneity in productivity, characterized by a Pareto distribution:  $G_i(z) = 1 - z^{-\kappa}$ ,  $\kappa > 1$ . By closing these models with

goods market-clearing conditions (as first suggested by Anderson & van Wincoop, 2003), they generally imply a structural gravity equation,

$$X_{ij} \propto Y_i E_j \left( \frac{\tau_{ij}}{\Pi_i P_j} \right)^{-\zeta}$$

(2)

$$(\Pi_i)^{-\zeta} = \sum_{j=1}^N \left( \frac{\tau_{ij}}{P_j} \right)^{-\zeta} E_j \quad \text{for } i = 1, \dots, N$$

(3)

$$(P_j)^{-\zeta} = \sum_{i=1}^N \left( \frac{\tau_{ij}}{\Pi_i} \right)^{-\zeta} Y_i \quad \text{for } j = 1, \dots, N$$

(4)

where  $\Pi_i$  can be regarded as country  $i$ 's multilateral outward resistance to exports, as it is an average of country  $i$ 's bilateral trade cost  $\tau_{ij}$  to reach each destination market  $j$  (relative to market  $j$ 's overall resistance to imports,  $P_j$ ), weighted by destination market size  $E_j$ . Similarly,  $P_j$  can be regarded as country  $j$ 's multilateral inward resistance to imports. The structural gravity equation in (2) extends equation (1) by down-weighting the absolute bilateral trade cost with these two multilateral resistance (MR) terms and linking these MR terms and bilateral trade costs across countries by the structural conditions (3)-(4). See Head and Mayer (2015) for a detailed synthesis of this literature.

Interestingly, the gravity specification is robust to several generalizations of the above literature. For example, Bernard et al. (2003) generalized Eaton and Kortum (2002) by adopting Bertrand competition (instead of perfect competition). This leads to variable markups but nonetheless the same expression for aggregate bilateral trade (under a Fréchet-like productivity distributional assumption). Melitz and Ottaviano (2008) extended Melitz (2003) by adopting linear demand (instead of CES) for differentiated varieties. The model can also accommodate variable markups, and although the implied bilateral trade flow departs from the structural gravity equation (2), it follows a generalized gravity equation:

$$X_{ij} \propto \left( \frac{\tau_{ij}^{-\zeta}}{\Omega_i \Xi_j} \right)$$

(5)

## Geography, Trade, and Power-Law Phenomena

---

where  $\Omega_i$  and  $\Xi_j$  are exporter and importer-specific deep parameters (not limited to gross outputs and aggregate expenditures).

Most importantly, the trade cost factor  $\tau_{ij}$  in the above literature often enters the gravity equation by a decreasing power function, where the power  $\zeta$  corresponds to either the demand-side parameter (the degree of the elasticity of substitution across goods,  $\sigma$ , net of 1) or the supply-side parameter (the inverse dispersion measure of the productivity distribution  $\kappa$ ). Strictly speaking, the Melitz-type model implies a gravity equation that depends not only on variable trade cost  $\tau_{ij}$  but also fixed trade cost  $f_{ij}$ , and inclusive of both the extensive margin (the mass of firms that export to a destination weighted by market share) and the intensive margin of trade (the volume of trade at firm level). Nonetheless, both margins ultimately depend on variable and fixed trade costs by a power law (with different power functions) under a Pareto distribution for firm size, as formally shown by Chaney (2008).

The power  $\zeta$  for the trade cost factor also corresponds to the (partial) trade elasticity, which has attracted much attention in recent quantitative trade models. As first suggested by Arkolakis, Costinot, and Rodríguez-Clare (2012) (ACR), under certain restrictions, the trade elasticity and changes in domestic trade share (an inverse measure of trade openness) are sufficient statistics of a country's welfare gain from trade (or more generally, a country's welfare change across two levels of trade openness), regardless of the underlying trade models:

$$\widehat{W} = (\widehat{\lambda})^{-1/\zeta}$$

(6)

where  $\widehat{W} \equiv W'/W$  indicates the change in real income and  $\widehat{\lambda} \equiv \lambda'/\lambda$  the change in domestic trade share (e.g.,  $\lambda_j = X_{jj}/E_j$ ). Equation (6) holds in most of the canonical trade models we mentioned above that generate a gravity equation. Thus, there is a close link between the gravity equation and the generality of the ACR welfare formula. Once the condition of constant trade elasticity fails, the simple welfare-change formula of ACR in (6) also breaks down. Thus, whether the power law holds globally (across all levels of trade cost and openness) in the gravity equation is a necessary condition for using the ACR formula in (6) for quantitative welfare evaluation.<sup>1</sup>

In the remainder of this section, recent trade models that bear on the gravity equation are reviewed, with special focus on the underlying modeling assumptions that imply a power-law phenomenon in trade flows.

### Network, Geography, and Gravity

In the canonical model of Chaney (2008), bilateral trade flows are a power function in variable trade cost  $\tau_{ij}^{-\zeta}$  under the Melitz structure with Pareto productivity distribution, where the power  $\zeta$  equals exactly the inverse dispersion parameter of the productivity

## Geography, Trade, and Power-Law Phenomena

---

distribution  $\kappa$ . Thus, a power law in firm size implies a power law in aggregate bilateral trade flows under the Melitz setup with CES preferences.

Chaney (2018) instead starts with a general setup without regard to the preferences and proposes three sufficient conditions under which the distance elasticity of trade is constant (at least for long distances). This requires more than a Pareto distribution for firm size (with shape parameter  $\kappa$ ), but also a line-up of firm size and the distance of firms' exports such that the average squared distance of exports is an increasing power function of firm size (with a power  $\mu > 0$ ) and that  $\kappa < 1 + \mu$  (such that the right tail of firm size is sufficiently thick and trade over long distance is dominated by exports of large firms). When these conditions hold, the distance elasticity tends toward  $\zeta = 1 + 2(\kappa - 1) / \mu$ . Thus, if the firm size follows approximately Zipf's law ( $\kappa = 1$ ), the distance elasticity is also approximately equal to one. Chaney (2018) documents that these sufficient conditions indeed hold for French firm-level data, and the distance elasticity  $\zeta$  for aggregate trade is not statistically different from that predicted by the theory:  $1 + 2(\kappa - 1) / \mu$ .

Chaney (2018) then micro-founds the first two conditions by a network theory of firm-to-firm trade, where a firm grows bigger as its network of contacts grows along a real line (with the new contacts arriving at a constant rate through its existing network of contacts), and as a result, exports to contacts that are further away. Firm size is measured by the number of contacts that a firm has. Since firms are born at a constant rate and establish new contacts at a constant rate (among their cohorts), firm size corresponds to firm age, and the fraction of firms above a certain age (and size) is, as a result, a power function (in the section "Micro-Foundation for Power Law in Firm Size," we will discuss further the mechanism under which the above assumptions give rise to the power law in firm size). He then uses partial differential equations to characterize how the probability density function of a firm's contacts (indexed by their distance from the firm) evolves over time to calculate the second moment of the distance distribution. Using Fourier-transform techniques to convert the density function into scalars, and the properties of convolution products, he shows that it is indeed an increasing power function in firm age (size).

This paper thus provides a network micro-foundation for a power-law phenomenon in the firm size distribution, the average squared distance of firm-level exports, and in the geography distribution of aggregate trade flows.

### Endogenous Market Access Cost

The Melitz-Chaney model with fixed export cost and CES preferences implies a uniform elasticity of substitution between varieties, and hence (i) a power-law distribution of firm export sales, and (ii) equal growth rates of trade (in response to trade cost reduction) for all previously traded goods. Eaton, Kortum, and Kramarz (2011) demonstrated—using French firm-level data on exports to each destination market—that firm export sales approach Pareto distribution only at the tail of the largest firms. Meanwhile, Arkolakis (2010) showed that during the NAFTA liberalization episode in the 1990s, the growth rate of trade is larger if the initial sales of goods are lower—using data on disaggregated prod-

## Geography, Trade, and Power-Law Phenomena

uct categories. Both papers reconcile the first stylized fact by allowing for endogenous market access cost (instead of fixed export cost). The market access cost to reach each additional consumer increases (i.e., it is an increasing convex function in the total number of consumers reached). As a result, relatively unproductive firms choose to reach only a few consumers in a given market.

More productive firms sell more in a market at the conventional intensive margin (sales per consumer) but also at the new consumer margin (they endogenously choose to reach more consumers):

$$r_{ij}(z) \equiv p_{ij}(z) q_{ij}(z) = \begin{cases} \sigma \frac{L_j^\alpha w_i^{1-\gamma} w_j^\gamma}{\psi} \left(\frac{z}{z_{ij}^*}\right)^{\sigma-1} \left[1 - \left(\frac{z_{ij}^*}{z}\right)^{(\sigma-1)/\beta}\right] & \text{if } z \geq z_{ij}^* \\ 0 & \text{otherwise} \end{cases}$$

(7)

where  $r_{ij}(z)$  is the export sales of a firm with productivity level  $z$ ,  $L_j$  is the destination market size,  $w_i$  and  $w_j$  are wage levels in country  $i$  and  $j$ , respectively,  $z_{ij}^*$  is the productivity cutoff to enter market  $j$  for firms from country  $i$ , and  $(\alpha, \beta, \psi)$  are parameters in the marketing technology function. For example, a higher  $\beta$  implies a higher degree of convexity in the marketing cost function. When  $\beta = 0$ , the costs to reach additional consumers remain constant and the setup becomes observationally equivalent to the case of fixed export cost. For positive  $\beta$ , by (7), export sales approach the Pareto distribution only at the upper tail (where the intensive margin dominates the new consumer margin for firms of sufficiently high productivities), while the export sales toward the lower tail are increasingly less than what is predicted by Melitz-type models with fixed export cost. These observations are consistent with the sales distribution of French firms in foreign markets documented by Eaton et al. (2011).

Given this structure, Arkolakis (2010) further demonstrates that the partial elasticity of firm-level trade  $r_{ij}(z)$  with respect to a change in  $\tau_{ij}$  is decreasing in firm productivity levels and approaches that of the Melitz-type model ( $\sigma-1$ ) as  $z$  tends to infinity:

$$\zeta(z) = (\sigma - 1) + \frac{\sigma - 1}{\beta} \left[ \left( \frac{z}{z_{ij}^*} \right)^{(\sigma-1)/\beta} \right]^{-1}$$

(8)

Thus, trade cost changes have proportionally bigger impacts on the smaller exporters in a given market.

Interestingly, even though trade elasticity at the firm level is not constant, the paper shows that to a first-order approximation, the aggregate trade flows,  $X_{ij}/X_{jj}$ , are still de-

scribed by a power-law function in the variable trade cost  $\tau_{ij}$  with  $\zeta = \kappa$  (taking into account the intensive margin, the new consumer margin, and the new firm margin), consistent with Chaney (2008). This set of papers thus provides theoretical support for the hypothesis that aggregate trade flows can be approximated reasonably well by a power law, regardless of the micro-structures for firm-level trade, so long as the firm size follows a power law.

### Non-Homothetic Preferences

The above literatures leave the impression that variations in the supply-side micro-structures do not substantively invalidate the power-law phenomenon of aggregate trade flows. The work of Melitz and Ottaviano (2008) also suggested that variable markups and non-homothetic preferences may leave this conclusion intact as well. The work of Arkolakis et al. (2019) (ACDR) confirmed that this is indeed the case, even in environments with non-CES homothetic preferences, or with non-homothetic (but directly additive) preferences. Bertolletti, Etro, and Simonovska (2018) (BES) further show that a setup with indirectly additive preferences (indirect utility functions that are additive in prices), but with the supply-side structure of Melitz and Ottaviano (2008), will arrive at a gravity equation that observes the power law as well (with  $\zeta = \kappa$ , the shape parameter of the Pareto productivity distribution). Thus, in sum, whether the preferences are CES (as in ACR), non-CES homothetic or directly additive non-homothetic (as in ACDR), or indirectly additive non-homothetic (as in BES), so long as the firm productivity follows a power law (i.e., Pareto distribution), the normalized aggregate trade flows ( $X_{ij}/X_{jj}$ ) are still expressed by a power function in  $\tau_{ij}$ .

However, an important point highlighted by Arkolakis et al. (2019) and Bertolletti et al. (2018) is that even if the partial trade elasticity of trade cost is constant, which can in principle be estimated using the gravity equation, the welfare implications are not equivalent across ACR, ACDR-non-homothetic, and BES-non-homothetic. In particular,

$$d \ln W_j^{ACR} = \frac{-d \ln \lambda_{jj}}{\kappa}$$

(9)

$$d \ln W_j^{ACDR-nhomo} = \left(1 - \frac{\rho}{\kappa + 1}\right) \frac{-d \ln \lambda_{jj}}{\kappa}$$

(10)

$$d \ln W_j^{BES-nhomo} = \bar{\epsilon}^c \frac{-d \ln \lambda_{jj}}{\kappa}$$

(11)



## Geography, Trade, and Power-Law Phenomena

---

where  $\rho$  is a sales-weighted average of the elasticity of markups to productivity. For example, it is zero under constant markups as in ACR, and positive in Krugman (1979) and Melitz and Ottaviano (2008). Since in models with non-homothetic preferences, the elasticity of markups to productivity lies in the unit interval,  $\rho \in (0, 1)$ , the welfare gain is smaller under ACDR non-homothetic preferences than ACR.<sup>2</sup> Since  $\left(1 - \frac{\rho}{\kappa+1}\right) \in \left(\frac{\kappa}{\kappa+1}, 1\right)$ , a more dispersed firm size distribution (smaller  $\kappa$ ) implies a potentially larger deviation.

Meanwhile,  $\bar{\epsilon}^c$  represents the average pass-through (of cost changes to prices). When pass-through is high and approaches unity as in the case of ACR, the welfare implications are similar between BES and ACR. In general, however, it is likely that the welfare gains from trade liberalization are lower under BES non-homothetic preferences, given incomplete pass-through in the BES environment. The less elastic the demand (and the lower the pass-through) is, the larger the downward revision.

### Caveats

In sum, the power law in aggregate bilateral trade flows turns out to be a robust theoretical regularity, as it holds under various perturbations to the supply-side and demand-side structures. The critical condition to guarantee this generality appears to be a power law in firm size, as shown by Melitz and Redding (2015), Chaney (2018), and Arkolakis et al. (2019). However, note that even if this class of models implies a power-law phenomenon in the aggregate trade flows (i.e., constant trade elasticity), they do not necessarily imply the same welfare effects when trade cost changes, the nature of which is sensitive to the preference specifications. Thus, the empirical and theoretical validity of the gravity equation is not a sufficient condition for welfare equivalence across trade models.

This section ends with another caveat, noting that the literature surveyed above has universally assumed iceberg trade cost following Samuelson (1954). The assumption of iceberg trade cost implies that trade cost is proportional to the quantity of production. Thus, with Pareto distribution of firm size/productivity, the presence of trade cost does not affect the relative market share of big versus small firms. This assumption also implies that the same factor intensity is used in marketing/shipping and in production. Thus, it neutralizes the potential effects of marketing/shipping technology on factor prices (Matsuyama, 2007).

The work by Hummels and Skiba (2004) and Irarrazabal, Moxnes, and Opromolla (2015), however, suggests that the reality is likely a combination of additive (per unit shipped) and multiplicative (iceberg) trade costs. The theoretical exploration of Sørensen (2014) and quantitative assessment of Irarrazabal et al. (2015) imply that reduction in additive (per unit) trade cost will have overall larger welfare impacts than reduction in iceberg trade cost of equal yield or equal impact on trade openness.

It seems plausible that with the introduction of additive (per unit) trade cost, the firm-level sales will not preserve its power law universally. Instead, the deviation from the power law would be larger for lower-priced (more efficient) firms since the per-unit trade cost is

a larger share of the consumer price for these firms. This pattern of deviation goes in the opposite direction from that of Arkolakis (2010) suggested by (7) and (8). One possible reconciliation is to introduce quality choice by firms such that more efficient firms also produce higher-quality and higher-priced varieties. This would plausibly help preserve the power law of firm sales for larger firms and the gravity equation at the aggregate for long distances as in Arkolakis (2010) and Chaney (2018). These speculations are subject to validations by future work.

## Micro-Foundation for Power Law in Firm Size

The above survey demonstrates the instrumental role of power law in firm size in generating the power-law phenomena in trade flows (gravity equation). It is thus important to understand the theoretical underpinnings of power laws in firm size. This section reviews the recent literature on this topic. Readers are referred to Gabaix (2009) for an extensive survey of the earlier literature. This article focuses on the connections of recent explanations for the power law in firm size with geography and trade, whenever appropriate.

### Firm Dynamics

The classic explanation for the power law in firm size, or sometimes more specifically, Zipf's law (a power law with a tail index of 1), is developed from the perspective of firm dynamics. It has a long tradition dating back to Gibrat (1931), who asserted that the rate of growth of firm size is independent of size. That is, the growth rate of firms of different size, as a random variable, follows the same distribution. Such a mathematical statement, dubbed "Gibrat's law," a law concerning growth of firm sizes, can be translated into geometric Brownian motion:

$$dx_t = \alpha x_t dt + \sigma x_t dz_t,$$

(12)

where  $x_t$  is the firm size,  $z_t$  is a Brownian motion without drift, and  $\alpha$  and  $\sigma$  are constants. Denote a Brownian motion with drift as  $d\tilde{x}_t = \alpha dt + \sigma dz_t$ . Then, the change of the Brownian motion is normally distributed by applying the central limit theorem to the random-walk representation with infinitesimal steps  $dz_t$ . Because  $d \ln x_t = dx_t/x_t$ , changes in  $\ln x_t$  over a finite time interval  $T$  follow a normal distribution. In other words, changes in  $x_t$  in a finite time interval  $T$  follow a log-normal distribution with mean  $(\alpha - \frac{1}{2}\sigma^2) T$  and variance  $\sigma^2 T$ . An immediate implication is that the limiting distribution of  $x_t$  when  $T$  goes to infinity does not exist. Nonetheless, if there is a reflection barrier for  $x_t$  as a lower bound, then the limiting distribution for  $x_t$  exists and is given by a power law.<sup>3</sup> For a concise exposition of the above, see chapter 3 of Dixit and Pindyck (1994). Because Brownian motion is essentially the continuous version of a random walk with infinitesimal steps, there are also discrete processes (e.g., the Kesten process) that generate power laws in a similar fashion.

## Geography, Trade, and Power-Law Phenomena

---

It is well known that the power law also holds for city size distribution with a tail index near one.<sup>4</sup> Gabaix (1999) identifies the condition under which Zipf's law for city size emerges, and he shows this with both a Brownian motion and a Kesten process and embeds these into a simple urban model. For the firm size distribution, Luttmer (2007) was the first to derive Zipf's law in a general equilibrium model of firm dynamics, in contrast to the previous literature that focused on probability processes that led to a power (or Zipf's) law. See, for example, Simon and Bonini (1958), Steindl (1965), and Ijiri and Simon (1964).

In Luttmer (2007), there are shocks to both demand and productivity, and if both shocks are assumed to follow geometric Brownian motion, then firm size also follows geometric Brownian motion. Here, the shock to demand can be interpreted as changes in "quality." His economic mechanism features firm entry, exit by selection, and firm growth. Incumbent firms need to pay a fixed cost to keep operating, and when some incumbent firms' quality augmented productivity falls too low, they exit. Entrant firms can imitate the incumbents by randomly sampling an incumbent and obtaining the scaled-down productivity of this incumbent (hence it is an imperfect imitation). As both types of firms grow at a similar rate, there exists a balanced growth path in which both types of firms exist. There is a "return process" in the model that plays the role of a reflection barrier. The "return process" exists because firms exit when their quality augmented productivity is below some barrier and enter at a point above this barrier. The stationary firm size distribution on the balanced growth path is a gamma distribution with a power law right tail, where the tail index depends on the parameters of the model. He shows that when the entry costs are high or that imitation is difficult, the resulting tail index is close to one (his data indicate a tail index of 1.06 for the U.S. firm-size distribution), and this seems to be consistent with the fact that U.S. entrant sizes are typically small.

The intuition for his result is as follows. Note first that when a power law approaches Zipf's law, the mean firm size grows without bound. Then, because firm profitability is tied to size, the fact that entrants attempt to imitate a randomly sampled incumbent ties the expected gains from entry to the average size of incumbents. In equilibrium, entry cost must be high so that it can compensate the high expected gains that reflect the large firm size of top firms (fat tails!). All of these are relative to the continuation cost for the incumbent firms (fixed cost). In short, this paper specifies clearly in economics what it takes in a model of firm dynamics to entail a distribution with very large mean that is consistent with a power-law generating process (i.e., geometric Brownian motion). Later, Luttmer (2012) considered different imitation mechanisms and shows that Zipf's law still holds in a mechanism where entrants can make only small improvements over the technologies used by the least productive incumbents. This echoes Luttmer's (2007) condition that entry or imitation is difficult, but under this alternative mechanism both entry and exit rates are high, as observed in the data. For other economic models of firm dynamics based on the random growth mechanism, see also Rossi-Hansberg and Wright (2007) and Acemoglu and Cao (2015).

## Geography, Trade, and Power-Law Phenomena

---

The literature that explains the power or Zipf's law using firm dynamics is appealing in its economic mechanism and its generality in the sense that it can potentially be extended to the power-law phenomena in other objects or fields. Nevertheless, this approach relies strongly on Gibrat's law, whose validity is not without a question (see, e.g., Stanley et al., 1996; Rossi-Hansberg & Wright, 2007).<sup>5</sup> From the perspective of this article, if power law in firm size leads to a gravity equation in international and regional trade flows, one might expect it to be related to geography and trade as well. Few studies in the firm-dynamics literature incorporate trade and geography. In addition, one might also expect the power law in firm size to be related to that in city size, as cities are made of firms and also firms' major markets. Finally, one might also expect the power law in firm size to be related to that in income, which is also a documented empirical regularity. The following sub-sections review recent developments in the literature that incorporate these different angles without resorting to Gibrat's law.

### Static Explanations: Networks, Firm Hierarchy, and Innovation

#### Trade Networks

The degree distribution of a network is known to exhibit a power law. If links among nodes are formed randomly with each pair of nodes having the same probability of linking, such a *random graph* does not entail a power law. A "scale-free network" is one type of graph in which the probabilities of forming links are proportional to the degrees (sizes) of the nodes. This process of link formation is also known as *preferential attachment*. See, for example, Barabási and Albert (1999). In some sense, this is similar to random growth because under preferential attachment, the bigger nodes have similar growth rates in degrees to the smaller nodes as the process goes on. In the economics literature, Chaney (2014) is a prominent example that applied preferential attachment to explain the power law in firm size and explained how this is linked to geography and exports.

Specifically, using French firm-level data, Chaney finds two stylized facts to motivate his theory. First, if a firm exports to a larger number of countries, then it is more likely to enter still more countries in the future. Second, where a firm's exports today affect the locations to which the firm will export goods in the future, since the firm will tend to export to places near its current export destination. Here, "near" could mean either geographic proximity or larger trade linkage. These facts prompted him to develop a model for trade networks.

The model Chaney provides features a discrete set of locations where distances between locations can be defined. There is a search friction so that a firm needs to search for its customers (could be other firms) via either a local search or a remote search. A search is local when the firm searches from its own location, whereas a search is remote when it starts from the firm's existing customers' locations. While the local search is similar to a random attachment that forms a random graph, the remote search is similar to the preferential attachment that is the basis of a scale-free network. As the model features both types of searches, he shows that asymptotically the number of consumers (which is the measure of firm size in that paper) exhibits a power-law tail. Note, however, that the num-

ber of consumers is essentially the number of contacts/links, and does not map directly into sales or employment, so there is a lack of intensive margin in this model. While the empirical firm-size distribution is based on either sales or employment, whether or not considerations of the intensive margin would distort the result is unresolved in Chaney (2014).

As discussed in the subsection *Network, Geography, and Gravity*, Chaney (2018) showed how the gravity equation can arise from a combination of three conditions: (1) a power law in firm size, (2) the average squared distance of exports following an increasing power function in firm size, and (3) a restriction on the relative magnitude of the two power parameters. Chaney (2018) micro-founds the first two conditions using a modified model of Chaney (2014), which does away with the local search, and provides aggregate properties of trade (in addition to firm-level ones). He shows that if Zipf's law in firm size holds, then the distance elasticity of the gravity equation will be one ( $\zeta \approx 1$ ). Also note that the basis for both the power law in firm size and the gravity equation in this framework is the preferential attachment in the network formation, and geography provides the link between the two. The model is appealing in that the second condition is independent of particular geography, and hence the results are robust to technological progress in transportation or changes in national borders.

### Firm Hierarchy

Following the seminal “problem-solving” model of Garicano (2000) to explain the hierarchy within a firm, related theoretical and empirical studies flourished. See, for example, Antràs, Garicano, and Rossi-Hansberg (2006), Garicano and Rossi-Hansberg (2006), Bloom et al. (2014).<sup>6</sup> The model features a “bottom” layer of production workers who own certain basic skills/knowledge and engage in day-to-day production work. However, problems that are beyond the knowledge of production workers arise from time to time, and the workers pass these problems “up” to a layer of managers, whose time is used for problem solving. If there are problems more difficult to solve than what this layer of managers can handle, these problems are then passed to the next higher layer of managers. This model generates a pyramidal firm structure in which the higher the layer, the more capable the managers are in problem solving. The pyramidal structure stems from the fact that more difficult problems do not occur as often, and hence the number of managers in higher layers is lower. If skill is defined as the difficulty of the problem, and if skill is costly to obtain, then the pyramidal structure will be even more pronounced. Naturally, wages for higher-echelon layers of managers are higher. One naturally wonders whether this kind of pyramidal structure would in any way become a fractal structure so that power law in wages/income (also a well-known empirical regularity) can be explained.<sup>7</sup> Nevertheless, such a fractal structure approach to explain power laws using a pyramidal firm hierarchy has not appeared yet.

Instead, Geerolf (2017) presents a modified version of the Garicano model with manager-worker matching and explains both the power laws in firm size and wage/income using a “power law change of variable close to the origin” technique that has been used in the physics literature (e.g., Jan et al., 1999; Sornette, 2002; Sornette, 2006; Newman, 2005).<sup>8</sup>

## Geography, Trade, and Power-Law Phenomena

Formally, suppose that two variables  $s$  and  $t$  have a reciprocal relationship,  $s = t^{-\alpha}$  ( $\alpha > 0$ ), and  $s$  has a positive probability density around the origin,  $\lim_{s \rightarrow 0} f(s) = K > 0$ . Then, the probability density of  $t$  exhibits a power law (in its right tail) as

$$g(t) = f(s(t)) \alpha t^{-\alpha-1}$$
$$\frac{g(t)}{t^{-\alpha-1}} \approx \alpha K \text{ asymptotically.}$$

That is, a power law (in the right tail) for  $t$  only requires that  $\lim_{s \rightarrow 0} f(s) = K > 0$ , but otherwise  $f(s)$  can take on any functional form.<sup>9</sup>

The setup of Geerolf (2017) assumes that the skills of agents are distributed on a continuum  $[1-\Delta, 1]$ , with  $\Delta \in [0, 1)$ . Start with a two-layer firm hierarchy. There exists a cutoff  $z \in (1-\Delta, 1)$  so that a type of manager  $y > z$  is matched with a type of worker  $x \leq z$ . The difficulty of a problem is distributed *uniformly* in  $[0, 1]$ . So, a worker of type  $x$  can solve  $x$  fraction of the problems arising from the production process on his/her own, and passes the  $1-x$  fraction of the problems to his/her manager whose type is denoted as  $y$ . To help the workers with the  $1-x$  fraction of problems, the manager needs to spend  $h$  units of time for each problem. With the labor endowment of the manager normalized to one, the number of workers that a manager of type  $y$  can supervise (i.e., the span of control) is then  $n(y) = \frac{1}{h(1-x)}$ . In this worker-manager matching model, a manager represents a firm, and because a firm with a manager of type  $y$  can solve  $y$  fraction of the problems, its output is also  $y$  per production worker. Thus, the total output of this firm is  $\frac{y}{h(1-x)}$ . This model then features an assortative matching so that a type 1 manager is matched with a type  $z$  worker, and a type  $z$  manager is matched with a type  $1-\Delta$  worker.

As firm size is measured by the span of control, he shows that when  $\Delta \rightarrow 0$ , a power law in firm size emerges with a tail index of 2. For a quick intuition, first note that when  $\Delta \rightarrow 0$ , the support for skills  $[1-\Delta, 1]$  becomes infinitesimal, which means that the managers are supervising very capable workers. But it does not require much time to supervise these workers, and hence the span of control becomes large. Formally, the  $1-x$  term in the span of control  $n(y) = \frac{1}{h(1-x)}$  tends to zero and  $n(y)$  becomes arbitrarily large.

Since the density of  $1-x$  is uniform, the “power law change of variables close to the origin” can be applied, and a power law emerges. In fact, the uniform assumption is stronger than necessary. As mentioned, all that is asked is that the density of  $1-x$  near zero be bounded away from zero. Geerolf then generalizes his model to a setup with multiple layers  $L$ , where the span of control of the top managers becomes a multiplication of the span of controls of the lower layers. This implies a fatter tail than the two-layer case with a tail index of  $1 + \frac{1}{L-1}$ . Thus, when  $L \rightarrow \infty$  to, Zipf’s law is obtained.

From the optimal choice of the matching process,  $w'(y) = n(y)$ , where  $w(\cdot)$  denotes the payoff function. This is because the difference in the payoff/income  $dw$  across different managers equals the additional output  $dy$  per worker multiplied by the number of work-

ers  $n(y)$ . In the many-layer case, the income of a top-level manager  $x_L \in [z_L, 1]$  is given by

$$w(x_L) = w(z_L) + \int_{z_L}^{x_L} n(y) dy.$$

Since  $n(y)$  approximately follows Zipf's law when  $L$  is large, the distribution of the top income becomes a power law with a tail index of 2.

### Innovation

Firm size is closely linked to productivity. For example, prominent firm-heterogeneity models such as Melitz (2003) and Eaton and Kortum (2002) assume Pareto and Fréchet productivity distributions, respectively, and both models entail a power-law tail in firm size because of these assumptions. An obvious question, then, is why should the underlying heterogeneity among firms be Pareto or Fréchet distributed? (or more specifically, have a power-law tail?). Chen, Hsu, and Peng (2018) provided an explanation for both the power laws in productivity and firm size in a rather standard general-equilibrium model of trade. Their model is the same as that of Melitz (2003), except that an innovation stage is added so that productivity is endogenously determined by the firm given its innate capability. The capability of a firm is inversely modeled as its probability of failure in conducting R&D experiments. Since higher productivity induces lower prices and larger sales, a more capable firm devotes more resources to conduct R&D experiments to obtain higher productivity. This article's key finding is that power laws in both productivity and firm size emerge under general underlying heterogeneity of firms. The mathematical mechanism is again the "power law change of variable close to the origin." That is, it only asks that there be a sufficient mass of top firms in the sense that the distribution of failure probability in R&D experiments needs to have a positive and finite density around zero. Moreover, the power laws hold in a general open-economy environment where almost all parameters are allowed to be country specific.

Besides the underlying firm heterogeneity, this article also generalizes the preference and technological constraints compared with standard trade models. The power laws for both productivity and firm size survive when the demand and innovation cost functions are both regularly varying. For example, this includes CES and many non-CES and non-homothetic preferences on the demand side, and general polynomial functions for the innovation cost.

### Geography and Power Laws

While Chaney (2018) provides a unified framework to think about geography, trade, and the power law phenomena, the role of geography is limited to providing a link between the gravity equation and the power law in firm size. As power laws emerge from scale-free structures, such as scale-free growth processes (Kesten processes or Brownian motion), scale-free networks, or more generally fractal structures, one wonders whether *geography itself* can provide a fractal structure that forms the basis of a power law. In other

## Geography, Trade, and Power-Law Phenomena

words, Chaney's mechanism for generating the power laws is still based on preferential attachment in network formation, rather than any geographic/spatial fractal structure. In particular, could the power law in cities (a salient phenomenon in geography) be linked with the power law in firm size in some way?

Hsu (2012) provided a theory that generates the power laws in both city size and firm size. His model is a modern formalization of the original central place theory of Christaller (1933) via a firm entry model with a continuum of goods. Central place theory describes a hierarchy of cities and towns that emerges from homogeneous plains of farmers, who are the base consumers for the goods and services that firms produce (while farmers focus on agriculture). The two main properties of this theory are the *hierarchy and central place properties*. The hierarchy property states that if a good of certain degree of scale economies is produced at a location, then all goods with lower degrees of scale economies are also produced in the same location. The central place property states that in a hierarchy of cities, a next-layer city is located in the middle of the two neighboring larger-sized cities.

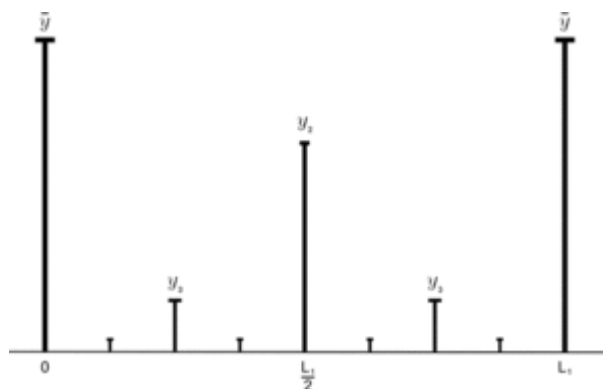


Figure 1. A central place hierarchy.

When both properties hold, such a hierarchy of cities is called a *central place hierarchy*. See Figure 1 for a depiction of this hierarchy. Here, the vertical axis is the commodity space, and each good is indexed by  $y \in [0, \bar{y}]$ , where  $y$  represents the degree of scale economies of the good (in his model,  $y$  is the fixed cost of production). Layer-1 cities are those producing all goods  $[0, \bar{y}]$ ; layer-2 cities are those producing all goods up to  $y_2$ ; layer- $i$  cities are those producing all goods up to  $y_i$ . Thus, a layer- $i$  city can be defined by the top good  $y_i$  it produces, and the hierarchy property implies that a layer- $(i + 1)$  city must be smaller than a layer- $i$  city.

Central place theory was first developed as an abstraction of the pattern that Christaller conceived as the hierarchy of cities and towns on the plain of southern Germany. In the modern-day economy, the so-called farmers can be broadly interpreted as those who are tied to extraction of natural resources and hence do not move around. Also, the validity of central place ideas does not depend on whether these “primary” industries represent a



## Geography, Trade, and Power-Law Phenomena

---

large or small share of the economy. The theory works as long as there exist immobile people who are spread over the entire geographic space.

Even though central place theory has been a key building block of economic geography, it has been praised for its deep economic insight but criticized for its lack of microfoundation (see, e.g., Fujita, Krugman, & Venables, 1999B). Attempts to provide microfoundation include Eaton and Lipsey (1982); Quinzii and Thisse (1990); Fujita, Krugman, and Mori (1999A); Tabuchi and Thisse (2006); Tabuchi and Thisse (2011); Hsu (2012); Hsu, Holmes, and Morgan (2014); and de Palma et al. (2019). In particular, Hsu (2012) derived both hierarchy and central place properties in the same framework. His unique contribution is to provide an explanation for both power laws in city size and firm size by central place theory.<sup>10</sup> The mechanism of the paper is explained in two steps as follows.

First, he explains how a central place hierarchy emerges. The model invokes a firm-entry mechanism that is based on spatial price discrimination (Lederer & Hurter, 1986). In such a mechanism, firms deliver goods to different locations, and in each location, firms engage in Bertrand competition. Thus, the resulting price is the second-lowest “delivered marginal cost” (unit cost of production and unit transport cost that increases with distance). Such a competition mechanism implies that firms that enter and survive are equal distances apart. Importantly, since goods differ in their degree of scale economies, the firms that produce goods with higher degrees will be further apart in the geographic space because they need larger market areas to survive. Such firms with larger market areas are also the larger firms. Thus, the differences in the degrees of scale economies are translated into a hierarchy of firms. With a small positive externality among firms (such as shared consumers in the production location if workers are hired in that location), the hierarchy property emerges because firms producing different goods tend to be located in the same place, and those firms producing goods with higher degree of scale economies appear in fewer locations. Because of the fact that competing firms are equal distances apart, a central place property emerges when the hierarchy property is already in place.

Second, a central place hierarchy as seen in Figure 1 is already similar to a fractal structure except that the density of goods in terms of the degree of scale economies is left open. Hsu (2012) showed that when the density function is regularly varying at the origin (i.e., only for goods with small-scale economies), then the power laws for both city and firm sizes emerge. A brief account for the mechanism is as follows. The size of a city is proportional to its total production, which is positively related to the range of goods it produces and the market areas of the goods in different subranges. Naturally, a layer- $i$ 's city size is proportional to the sum of the production across all subranges  $(y_{k+1}, y_k)$  for all  $k \geq i$ . If the density function is regularly varying around the origin, then city size  $Y_i$  becomes a geometric series that is proportional to  $(\delta/2)^i$  for some constant  $\delta > 0$ . In such a city hierarchy with  $R_i = 2^i - 1$ , ranks and sizes of cities and firms are approximately log-linearly related (i.e., the power laws). Hsu shows that the condition on the above-mentioned density is weak since it includes many well-known and widely used distributions

(See his Table C1 in Hsu, 2012). In other words, with this mild condition, a central place hierarchy becomes a fractal structure, and the two power laws follow.<sup>11</sup>

Even though the paper engages in more discussion about cities, it is important to note that in this model, the hierarchy of cities and the hierarchy of firms are exactly the same thing except they are viewed from different angles.<sup>12</sup> In particular, firm size and geography are linked by the fact that larger firms serve larger market areas. This also echoes the empirical and theoretical findings in Chaney (2018).

### Network and City Size Distribution

So far, our discussions surround the power law in firm size. Nevertheless, given the previous discussion on the link between geography, city size, and firm size, it is also important to note a recent theory connecting geography and city size distribution by Berliant and Watanabe (2018). Their paper proposes a theory in which scale-free transportation networks *à la* Barabási and Albert (1999) give rise to power laws in city size, and they show that the estimated model generates city sizes that are sufficiently close to the actual city size distribution.

The model is simple and can be briefly described here. The number of cities in the economy is fixed at  $J$ , with the city size (population/labor) of city  $i$  denoted as  $s_i$ . Each city produces a distinct good, and individual utility function is a Cobb-Douglas one with each of the  $J$  goods having the same expenditure share ( $1/J$ ). The production technology is constant returns to scale, and one unit of labor produces one unit of commodity. The market for each good is competitive. As firm size is not well defined in this environment, this model is silent on firm-size distribution. The critical assumption is that for a given network structure, shipping a commodity  $j$  from city  $j$  to city  $i$  requires an iceberg transport cost such that the prices paid by consumers in  $i$  is given by

$$p_i^j = \tau^{l_i^j} p_j^j,$$

(13)

where  $p_j^j$  is the price paid by consumers at city  $j$  for good  $j$ ,  $\tau \geq 1$  is the iceberg cost factor by going one step from one node to a neighboring node, and  $l_i^j$  is the geodesic length between city  $j$  and  $i$ , which is the number of steps along the shortest path between these two cities.

Define a city's accessibility by  $a_i \equiv -\sum_k l_i^k / J$ , i.e., the negative of the average geodesic length of city  $i$ . Berliant and Watanabe show that city size  $s_i$  is proportional to  $\tau^{a_i}$ . Furthermore, they show that accessibility  $a_i$  is a linear function in the logarithm of city  $i$ 's degree  $k_i$ . These imply a log-linear relationship between city size  $s_i$  and its degree  $k_i$ , which is, in turn, log-linear in the rank of city  $i$  because the degree distribution follows a power law under any scale-free network. Thus, city size distribution also exhibits a power law. Interestingly, when transport cost  $t$  increases, the tail index of the city size distribution is

reduced, implying a fatter tail and a more dispersed/skewed distribution.<sup>13</sup> That is, geography plays a definitive role in explaining city size distribution—larger geographic barriers lead to concentration of economic activities in a few cities.

## Concluding Remarks

This article reviews several power-law phenomena related to trade and geography. It first reviews the theoretical literature on the gravity equation and concludes that the power law in firm size is a sufficient condition for the gravity equation to hold in a large set of trade models. These include all the ACR-class models with CES preferences as well as some trade models with non-CES homothetic, directly additive non-homothetic, and indirectly additive non-homothetic preferences where the distribution of firm productivity follows a power law. The power law in aggregate trade is also shown by Arkolakis (2010) to be a good first-order approximation, even if the firm-level trade elasticity is heterogeneous due to endogenous choice of market access. Finally, Chaney (2018) identifies the model-free sufficient conditions for generating the gravity equation and demonstrates that these conditions are empirically plausible. Again, the power law in firm size is identified as a key condition. In contrast to this seemingly universal applicability of the gravity equation, the literature has arrived at the conclusion that welfare evaluation of gains from trade is not insensitive to the underlying model structures. In particular, non-homothetic preferences and hence variable markups will alter the simple ACR welfare formula. Thus, the “social physics” does not extend to the normative domain.

We also review several important theories of firm-size distribution. The popular explanation based on firm dynamics is discussed in detail, while static explanations based on networks (Chaney, 2014, 2018), hierarchical structure of firms (Geerolf, 2017), and innovation (Chen et al., 2018) are also reviewed. Whereas Chaney’s theory generates both power laws in trade (gravity equation) and firm size, Geerolf (2017) explains both power laws in firm size and income, and the model of Chen et al. (2018) generates both power laws in productivity and firm size. Moreover, the central place theory in its modern form (Hsu, 2012) explains both power laws in city size and firm size by a geographic fractal structure. These various theories advance understandings of the potential mechanisms behind the various power-law phenomena. The common theme across these theories is to deal with different aspects of agent heterogeneity in a neat way so that certain scale-free properties can emerge without imposing strict functional-form restrictions on the agent heterogeneity.

Recall that these power-law phenomena are important because: (1) they imply that top firms have significant influence on the macroeconomic performance (consistent with the granular economies phenomenon); and (2) related to the first point, the power-law coefficients are often tightly connected with welfare evaluation (as suggested by ACR and AC-DR). Further, the validity of these power laws forms a strong justification for making power-function assumptions in economic models, which make complex and large-scale quantitative analysis possible.<sup>14</sup> The recent development of quantitative analysis in both interna-

tional and regional economics manifests this point (see, e.g., Costinot & Rodríguez-Clare, 2015; Redding & Rossi-Hansberg, 2017, for surveys of these developments in the two respective fields). With the aid of power-function assumptions, these quantitative analyses can often accommodate arbitrary numbers of countries, industries, goods, firms and workers, etc., and this high-dimensionality nature of quantitative analysis leads economists much closer to effective and meaningful policy evaluation than ever.

### Further Reading

Allen, T., Arkolakis, C., & Takahashi, Y. (forthcoming). Universal gravity. *Journal of Political Economy*.

Anderson, J. E., & Yotov, Y. V. (2010). The changing incidence of geography. *American Economic Review*, 100, 2157–2186.

Bernard, A. B., & Moxnes, A. (2018). Networks and trade. *Annual Review of Economics*, 10, 65–85.

Bernard, A. B., Moxnes, A., & Saito, Y. U. (2019). Production networks, geography, and firm performance. *Journal of Political Economy*, 127, 639–688.

Cabral, L., & Mata, J. (2003). On the evolution of the firm size distribution: Facts and theory. *American Economic Review*, 93, 1075–1090.

Chaney, T. (2016). Networks in international trade. In Y. Bramoullé, A. Galeotti, & B. Rogers (Eds.), *The Oxford Handbook of the Economics of Networks*. Oxford: Oxford University Press.

Chang, P.-L., & Lee, M.-J. (2011). The WTO trade effect. *Journal of International Economics*, 85, 53–71.

di Giovanni, J., Levchenko, A. A., & Rancière, R. (2011). Power laws in firm size and openness to trade: Measurement and implications. *Journal of International Economics*, 85, 42–52.

Durantón, G. (2007). Urban evolutions: The fast, the slow, and the still. *American Economic Review*, 97, 197–221.

Fally, T. (2015). Structural gravity and fixed effects. *Journal of International Economics*, 97, 76–85.

Gabaix, X., & Ioannides, Y. M. (2004). The evolution of city size distributions. In J. V. Henderson & J.-F. Thisse (Eds.), *Handbook of Regional and Urban Economics* (Vol. 4, pp. 2341–2378). North Holland: Elsevier.

Head, K., Mayer, T., & Thoenig, M. (2014). Welfare and trade without Pareto. *American Economic Review Papers and Proceedings*, 104, 310–316.

Helpman, E., Melitz, M. J., & Yeaple, S. R. (2004). Export versus FDI with heterogeneous firms. *American Economic Review*, 94, 300–316.

Nigai, S. (2017). A tale of two tails: Productivity distribution and the gains from trade. *Journal of International Economics*, 104, 44–62.

Rossi-Hansberg, E., & Wright, M. L. (2007). Urban structure and growth. *Review of Economic Studies*, 74, 597–624.

Simonovska, I., & Waugh, M. E. (2014). The elasticity of trade: Estimates and evidence. *Journal of International Economics*, 92, 34–50.

Su, H.-L. (2019). On the city size distribution: A finite mixture interpretation. National Taiwan University Working Paper.

Tintelnot, F., Kikkawa, A. K., Mogstad, M., & Dhyne, E. (2019). Trade and domestic production networks. NBER Working Paper no. 25120.

## References

Acemoglu, D., & Cao, D. (2015). Innovation by entrants and incumbents. *Journal of Economic Theory*, 157, 255–294.

Anderson, J. E., & van Wincoop, E. (2003). Gravity with gravitas: A solution to the border puzzle. *American Economic Review*, 93, 170–192.

Antràs, P., Garicano, L., & Rossi-Hansberg, E. (2006). Offshoring in a knowledge economy. *Quarterly Journal of Economics*, 121, 31–77.

Arkolakis, C. (2010). Market penetration costs and the new consumers margin in international trade. *Journal of Political Economy*, 118, 1151–1199.

Arkolakis, C., Costinot, A., & Rodríguez-Clare, A. (2012). New trade models, same old gains? *American Economic Review*, 102, 94–130.

Arkolakis, C., Costinot, A., Donaldson, D., & Rodríguez-Clare, A. (2019). The elusive pro-competitive effects of trade. *Review of Economic Studies*, 86(1), 46–80.

Axtell, R. L. (2001). Zipf distribution of US firm sizes. *Science*, 293, 1818–1820.

Barabási, A.-L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286, 509–512.

Beckmann, M. J. (1958). City hierarchies and the distribution of city size. *Economic Development and Cultural Change*, 6, 243–248.

Berliant, M., & Watanabe, A. H. (2018). A scale-free transportation network explains the city-size distribution. *Quantitative Economics*, 9, 1419–1451.

## Geography, Trade, and Power-Law Phenomena

---

Berliant, M., & Watanabe, H. (2015). Explaining the size distribution of cities: Extreme economies. *Quantitative Economics*, 6, 153–187.

Bernard, A. B., Eaton, J., Jensen, J. B., & Kortum, S. (2003). Plants and productivity in international trade. *American Economic Review*, 93, 1268–1290.

Bertoletti, P., Etro, F., & Simonovska, I. (2018). International trade with indirect additivity. *American Economic Journal: Microeconomics*, 10, 1–57.

Bloom, N., Garicano, L., Sadun, R., & Van Reenen, J. (2014). The distinct effects of information technology and communication technology on firm organization. *Management Science*, 60, 2859–2885.

Chaney, T. (2008). Distorted gravity: The intensive and extensive margins of international trade. *American Economic Review*, 98, 1707–1721.

Chaney, T. (2014). The network structure of international trade. *American Economic Review*, 104, 3600–3634.

Chaney, T. (2018). The gravity equation in international trade: An explanation. *Journal of Political Economy*, 126, 150–177.

Chang, P.-L., & Lee, M.-J. (2011). The WTO trade effect. *Journal of International Economics*, 85, 53–71.

Chen, Y.-F., Hsu, W.-T., & Peng, S.-K. (2018). Innovation, firm size distribution, and gains from trade. Singapore Management University Working Paper no. 17-2018.

Christaller, W. (1933). *Central Places in Southern Germany*, translated by C. W. Baskin (1966), Englewood Cliffs, NJ: Prentice-Hall.

Costinot, A., & Rodríguez-Clare, A. (2015). Trade theory with numbers: Quantifying the consequences of globalization. In E. Helpman, K. Rogoff, & G. Gopinath (Eds.), *Handbook of International Economics* (Vol. 4, pp. 197–261). North-Holland: Elsevier.

de Palma, A., Papageorgiou, Y. Y., Thisse, J.-F., & Ushchev, P. (2019). About the origin of cities. *Journal of Urban Economics*, 111, 1–13.

Desmet, K., & Rappaport, J. (2017). The settlement of the United States, 1800–2000: The long transition towards Gibrat's law. *Journal of Urban Economics*, 98, 50–68.

Dixit, A. K., & Pindyck, R. S. (1994). *Investment under uncertainty*. Princeton, NJ: Princeton University Press.

Durantón, G., Morrow, P. M., & Turner, M. A. (2014). Roads and trade: Evidence from the US. *Review of Economic Studies*, 81, 681–724.

Eaton, B. C., & Lipsey, R. G. (1982). An economic theory of central places. *Economic Journal*, 92, 56–72.

## Geography, Trade, and Power-Law Phenomena

---

Eaton, J., & Kortum, S. (2002). Technology, geography, and trade. *Econometrica*, 70, 1741–1779.

Eaton, J., Kortum, S., & Kramarz, F. (2011). An anatomy of international trade: Evidence from French firms. *Econometrica*, 79, 1453–1498.

Edmond, C., Midrigan, V., & Xu, D. Y. (2015). Competition, markups, and the gains from international trade. *American Economic Review*, 105, 3183–3221.

Eeckhout, J. (2004). Gibrat's law for (all) cities. *American Economic Review*, 94, 1429–1451.

Feenstra, R. C., & Weinstein, D. (2017). Globalization, competition, and U.S. welfare. *Journal of Political Economy*, 125, 1041–1074.

Fujita, M., Krugman, P., & Mori, T. (1999a). On the evolution of hierarchical urban systems. *European Economic Review*, 43, 209–251.

Fujita, M., Krugman, P., & Venables, A. J. (1999b). *The spatial economy: Cities, regions, and international trade*. Cambridge, MA: MIT Press.

Gabaix, X. (1999). Zipf's law for cities: An explanation. *Quarterly Journal of Economics*, 114, 739–767.

Gabaix, X. (2009). Power laws in economics and finance. *Annual Review of Economics*, 1, 255–294.

Gabaix, X. (2011). The granular origins of aggregate fluctuations. *Econometrica*, 79, 733–772.

Garicano, L. (2000). Hierarchies and the organization of knowledge in production. *Journal of Political Economy*, 108, 874–904.

Garicano, L., & Rossi-Hansberg, E. (2006). Organization and inequality in a knowledge economy. *Quarterly Journal of Economics*, 121, 1383–1435.

Garicano, L., & Rossi-Hansberg, E. (2015). Knowledge-based hierarchies: Using organizations to understand the economy. *Annual Review of Economics*, 7, 1–30.

Geerolf, F. (2017). A theory of Pareto distributions. UCLA Working Paper.

Gibrat, R. (1931). *Les inégalités économiques*. Paris, France: Recueil Sirey.

Head, K., & Mayer, T. (2015). Gravity equations: Workhorse, toolkit, and cookbook. In E. Helpman, K. Rogoff, & G. Gopinath (Eds.), *Handbook of International Economics* (Vol. 4, chap. 3, pp. 131–196). North-Holland: Elsevier.

Hsu, W.-T. (2012). Central place theory and city size distribution. *Economic Journal*, 122, 903–932.

## Geography, Trade, and Power-Law Phenomena

---

Hsu, W.-T., Holmes, T. J., & Morgan, F. (2014). Optimal city hierarchy: A dynamic programming approach to central place theory. *Journal of Economic Theory*, *154*, 245–273.

Hummels, D., & A. Skiba. (2004). Shipping the good apples out? An empirical confirmation of the Alchian-Allen conjecture. *Journal of Political Economy*, *112*, 1384–1402.

Ijiri, Y., & Simon, H. A. (1964). Business firm growth and size. *American Economic Review*, *54*, 77–89.

Irarrazabal, A., Moxnes, A., & Opromolla, L. D. (2015). The tip of the iceberg: A quantitative framework for estimating trade costs. *Review of Economics and Statistics*, *97*, 777–792.

Jan, N., Moseley, L., Ray, T., & Stauffer, D. (1999). Is the fossil record indicative of a critical system? *Advances in Complex Systems*, *2*, 137–141.

Krugman, P. (1979). Increasing returns, monopolistic competition, and international trade. *Journal of International Economics*, *9*, 469–479.

Krugman, P. (1980). Scale economies, product differentiation, and the pattern of trade. *American Economic Review*, *70*, 950–959.

Krugman, P. (1997). *Development, geography, and economic theory*. Cambridge, MA: MIT Press.

Lederer, P. J., & Hurter, A. P., Jr. (1986). Competition of firms: Discriminatory pricing and location. *Econometrica*, *54*, 623–640.

Luttmer, E. G. (2012). Technology diffusion and growth. *Journal of Economic Theory*, *147*, 602–622.

Luttmer, E. G. J. (2007). Selection, growth, and the size distribution of firms. *Quarterly Journal of Economics*, *122*, 1103–1144.

Matsuyama, K. (2007). Beyond icebergs: Towards a theory of biased globalization. *Review of Economic Studies*, *74*, 237–253.

McCallum, J. (1995). National borders matter: Canada-U.S. regional trade patterns. *American Economic Review*, *85*, 615–623.

Melitz, M. J. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica*, *71*, 1695–1725.

Melitz, M. J., & Ottaviano, G. I. P. (2008). Market size, trade, and productivity. *Review of Economic Studies*, *75*, 295–316.

Melitz, M. J., & Redding, S. J. (2015). New trade models, new welfare implications. *American Economic Review*, *105*, 1105–1146.



## Geography, Trade, and Power-Law Phenomena

---

Monte, F., Redding, S., & Rossi-Hansberg, E. (2018). Commuting, migration and local employment elasticities. *American Economic Review*, *108*(12), 3855–3890.

Newman, M. E. (2005). Power laws, Pareto distributions and Zipf's law. *Contemporary Physics*, *46*, 323–351.

Quinzii, M., & Thisse, J.-F. (1990). On the optimality of central places. *Econometrica*, *58*, 1101–1119.

Redding, S. J., & Rossi-Hansberg, E. (2017). Quantitative spatial economics. *Annual Review of Economics*, *9*, 21–58.

Redding, S. J., & Sturm, D. M. (2008). The costs of remoteness: Evidence from German division and reunification. *American Economic Review*, *98*, 1766–1797.

Rossi-Hansberg, E., & Wright, M. L. (2007). Establishment size dynamics in the aggregate economy. *American Economic Review*, *97*, 1639–1666.

Rozenfeld, H. D., Rybski, D., Gabaix, X., & Makse, H. A. (2011). The area and population of cities: New insights from a different perspective on cities. *American Economic Review*, *101*, 2205–2225.

Samuelson, P. A. (1954). The transfer problem and transport costs, II: Analysis of effects of trade impediments. *Economic Journal*, *64*, 264–289.

Simon, H. A., & Bonini, C. P. (1958). The size distribution of business firms. *American Economic Review*, *48*, 607–617.

Sørensen, A. (2014). Additive versus multiplicative trade costs and the gains from trade liberalizations. *Canadian Journal of Economics*, *47*, 1032–1046.

Sornette, D. (2002). Predictability of catastrophic events: Material rupture, earthquakes, turbulence, financial crashes, and human birth. *Proceedings of the National Academy of Sciences*, *99*, 2522–2529.

Sornette, D. (2006). *Critical phenomena in natural sciences: Chaos, fractals, self-organization and disorder: Concepts and tools*. New York, NY: Springer Science and Business Media.

Stanley, M. H., Amaral, L. A., Buldyrev, S. V., Havlin, S., Leschhorn, H., Maass, P., . . .

Stanley, H. E. (1996). Scaling behaviour in the growth of companies. *Nature*, *379*, 804–806.

Steindl, J. (1965). *Random Processes and the Growth of Firms: A Study of the Pareto Law*. London: Griffin.

Tabuchi, T., & Thisse, J.-F. (2006). Regional specialization, urban hierarchy, and commuting costs. *International Economic Review*, *47*, 1295–1317.

## Geography, Trade, and Power-Law Phenomena

---

Tabuchi, T., & Thisse, J.-F. (2011). A new economic geography model of central places. *Journal of Urban Economics*, 69, 240–252.

Tinbergen, J. (1962). *Shaping the world economy: Suggestions for an international economic policy*. New York, NY: Twentieth Century Fund.

### Notes:

(1.) In the case of variable trade elasticity, the local welfare change formula of ACR may still be used:

$$d \ln W(\tau) = -\frac{1}{\zeta(\tau)} d \ln \lambda(\tau).$$

For example, the local formula holds in Edmond, Midrigan, and Xu (2015) and Feenstra and Weinstein (2017). Conceptually, one needs to obtain the elasticity at different levels of trade cost and integrate the right-hand side of the above equation over the relevant interval of trade cost, to derive the discrete change in welfare.

(2.) Nevertheless, in the ACDR class, trade flows and trade elasticity remain sufficient statistics for welfare gains from trade, as in ACR.

(3.) A reflection barrier  $\underline{x}$  is a lower bound such that  $x_t$  will remain where it is at time  $t$  if it encounters a negative shock that would move it below  $\underline{x}$  without any interference.

(4.) Eeckhout (2004) argues that the log-normal distribution is a better description for the “entire” distribution of human settlements (not just cities). There is an ensuing debate about this. As Eeckhout (2004) uses the U.S. census places as cities (human settlements), it is unclear why census places are good proxies for cities. Rozenfeld et al. (2011) instead use an algorithm to identify population clusters in the United States with population data of fine resolution (census tracts) and argue that Zipf’s law still better describes the distribution of the population cluster of more than 10,000 people. Regardless of the debate, the consensus of the literature is that the right tail of the city-size distribution is well approximated by a power law, whereas the entire distribution might be something else.

(5.) There is a similar doubt about the validity of Gibrat’s law for city growth. See, for example, Redding and Sturm (2008) and Desmet and Rappaport (2017). In particular, Redding and Sturm (2008) showed that a city’s growth rate depends on its “market potential,” which in turn depends on nearby cities’ sizes and growth. Hence, the growth rates of cities are not independently and identically distributed.

(6.) Also see the survey article Garicano and Rossi-Hansberg (2015).

(7.) A fractal structure is one in which the shape of the smaller parts of the structure resembles that of the bigger one or the entire structure.

(8.) The name of the technique is given by Sornette (2006, Section 14.2.1).

## Geography, Trade, and Power-Law Phenomena

---

(9.) As is, the power-law change of variables technique is simply a mathematical one that allows explanation for the power laws (at the right tail) to be explained by the properties of some distributions near zero. Economics come in as to which mappings for such change of variables are plausible. As will be seen in Geerolf (2017), the right tail of firm-size distribution is mapped to the neighborhood where the fraction of problems that workers cannot solve becomes very small (origin). In Chen et al. (2018), it is mapped to the neighborhood where the failure probability of firms conducting R&D experiments becomes very small.

(10.) Beckmann (1958) showed how power law in city size emerges from a fractal structure of population and market area. However, his structure is assumed rather than derived. He did not explain the power law in firm size either.

(11.) One drawback of this weak condition on density is that it approximates the two power laws at the left tail instead of the right tail. Nevertheless, the theory also provides a broader condition for the power laws. Namely, Proposition 2 states that if the sizes of the subranges  $\Delta_i \equiv y_k - y_{i+1}$  form approximately a geometric sequence, then the power laws emerge for the entire domain.

(12.) Note that whereas Hsu (2012) provided a theory explaining power laws as an equilibrium phenomena, Hsu et al. (2014) showed how central place hierarchy and the power law in city size emerge as a socially optimal outcome. Nevertheless, since there is no clear definition of firms in a social planner's problem, this study is silent on firm-size distribution.

(13.) This comparative statics in transport cost  $\tau$  is not specific to scale-free networks. It holds for all network structure. Also see Berliant and Watanabe (2015) for another formulation to explain skewed distribution for city size.

(14.) Although Fréchet distribution is not a power function, it has a power law tail.

### **Pao-Li Chang**

School of Economics, Singapore Management University

### **Wen-Tai Hsu**

School of Economics, Singapore Management University