Central place theory and city size distribution

Wen-Tai HSU
Singapore Management University, WENTAIHSU@smu.edu.sg

Published in The Economic Journal, 2011 June, Volume 122, Issue 563, Pages 903-932
http://doi.org/10.1111/j.1468-0297.2012.02518.x

CENTRAL PLACE THEORY AND CITY SIZE DISTRIBUTION*

Wen-Tai Hsu

This article proposes a theory of city size distribution via a hierarchy approach rather than the popular random growth process. It does so by formalising central place theory using an equilibrium entry model and specifying the conditions under which city size distribution follows a power law. The force driving the city size differences in this model is the heterogeneity in economies of scale across goods. The city size distribution under a central place hierarchy exhibits a power law if the distribution of scale economies is regularly varying, which is a general class that encompasses many well-known, commonly used distributions.

> The difference between a high-order metropolis like Chicago and a lower-order town like Peoria or Burlington was not merely Chicago's much larger population. Chicago's high rank meant that its market attracted customers for many more goods and services from a much wider region ... Just as one can rank human settlements according to the number of people who live in them, so can one rank all economic goods according to the number of people and concentrations of wealth needed to create a market for them. The hierarchy of urban settlements is also a hierarchy of markets.
>
> (Cronon, 1991, pp. 279–80).

City size distribution is highly skewed and heavy-tailed: there are many more small cities than there are large cities, and the distribution can be approximated by the power law, a.k.a., the Pareto distribution, with a tail index of around one. If the tail index is exactly one, then the distribution is called *Zipf's law* (following Zipf, 1949) or the *rank-size rule*, because the product of the rank and size of a city is a constant.

Observation of the power law becomes clearest when city ranks are plotted against city sizes on a log-log scale, i.e. a Zipf plot. Figure 1 presents a Zipf plot using US 2000 census data for all Metropolitan Statistical Areas (MSAs). The ordinary least square (OLS) regression of the log of rank on the log of size for all 362 MSAs entails an R^2 of 0.9857 and a slope, of which the absolute value is the approximate tail index, of −0.9491. The fact that a straight line fits fairly well leads one to think that the size distribution of cities may be generated by the power-law distribution defined by the tail

* Corresponding author: Wen-Tai Hsu, Department of Economics, National University of Singapore, Faculty of Arts & Social Sciences, AS2 Level 6, 1 Arts Link, Singapore 117570. Email: wentai@nus.edu.sg.

This article was previously circulated under the title 'Central Place Theory and Zipf's Law'. I am grateful to Tom Holmes, Sam Kortum and Erzo G. J. Luttmer for their advice and continuous support. For their helpful comments, I thank Marcus Berliant, V.V. Chari, Jiahua Che, Gilles Duranton, Jan Eeckhout, Masa Fujita, Xavier Gabaix, Luis Garicano, Yannis Ioannides, Patrick Kehoe, James Mirrlees, Tomoya Mori, Roger Myerson, Derek Neal, Shi Qi, Esteban Rossi-Hansberg, Maryam Saeedi, Robert Shimer, Tony E. Smith, Takatoshi Tabuchi, Matthew Turner, Harald Uhlig, Ping Wang, William Wheaton, two anonymous referees, and the seminar participants at University of Minnesota, Federal Reserve Bank of Minneapolis, University of Chicago, University of Toronto, Kyoto University, National Graduate Institute for Policy Studies, City University of Hong Kong, Chinese University of Hong Kong, Fudan University, the 2008 Annual Meeting of the American Real Estate and Urban Economics Association, the EITSS conference of Nagoya University, and the 2009 Meeting of the Urban Economics Association. All errors are mine.

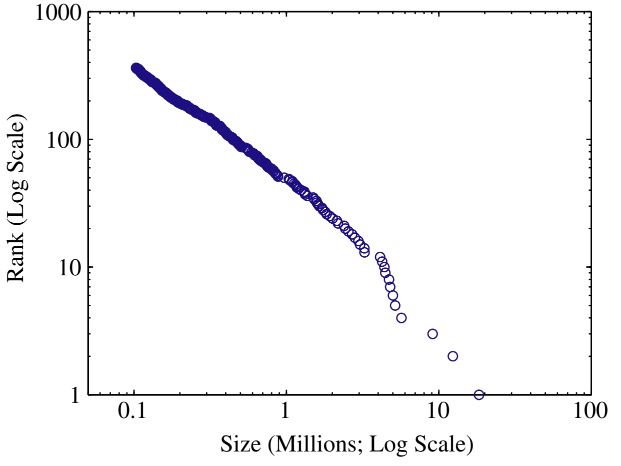

Fig. 1. *Zipf Plot for All 362 MSAs*

Source. 2000 US census.

probability, $P(S > s) = a/s^{\zeta}$, $s \geq \underline{s}$, for positive constants a and $\underline{s}$ and tail index ζ. Unlike most empirical work on city size distribution, which relies on city definitions that are based on administrative units, Rozenfeld *et al.* (2011) adopt an innovative 'from the bottom up' approach to construct cities, i.e. population clusters, by applying a computer algorithm on high-resolution data from the UK and US. They find that the power law holds fairly well for cities as small as 5,000 inhabitants in the UK and 12,000 inhabitants in the US.[1]

The striking regularity of city size distribution has attracted considerable interest in providing explanations for it. The popular approach has been to derive the city size distribution from the steady state of a random growth process. In a random growth process, large cities arise due to their long histories of favourable productivity shocks; however, as the probabilities of such events are low, there are not many large cities, i.e. a skewed distribution. For why this steady-state distribution should be Zipf's, see Gabaix (1999); for why it should be log-normal, see Eeckhout (2004).[2] See Duranton (2006, 2007) and Rossi-Hansberg and Wright (2007) for full-fledged general equilibrium models that match the empirical city size distribution reasonably well.[3] However, the locations of cities in the geographic space typically play no role in such models, and the size differences are usually driven by luck/shocks rather than by different functions of differently sized cities.

[1] An alternative parametrisation for the city size distribution is the log-normal. Using *census places* data in the US, Eeckhout (2004) finds that the entire distribution ('entire' meaning all human settlements) fits the log-normal well while the power law remains a good approximation to the upper tail.

[2] A log-normal distribution of city sizes entails a slight concave Zipf's plot. Some deviations from the benchmark model in this article can also generate slightly concave Zipf's plots. However, this article focuses on the power law, and hence these details are omitted.

[3] For surveys of this literature, see Gabaix and Ioannides (2004) and Gabaix (2009). The body of literature explaining city size distribution using static models or a non-random growth process is relatively small; see Eaton and Eckstein (1997); Brakman *et al.* (1999), and Krugman (1996).

This article explains city size distribution using a model in which cities of different sizes serve different functions in the economy. It builds on the insights of *central place theory,* which was first developed by German geographers Christaller (1933) and Lösch (1940). The idea underlying this theory is that goods differ in their degrees of scale economies. Goods with substantial scale economies, e.g. stock exchanges or symphony orchestras, will be found in only a few locations, whereas those with small scale economies, e.g. petrol stations or convenience stores, will be found in many. Moreover, large cities tend to have a wide range of goods, whereas small cities provide only those with a low degree of scale economies. In Christaller's (1933) scheme, the *hierarchy property*[4] holds if larger cities provide all of the goods that smaller cities also provide. We can readily see that this setup implies a skewed city size distribution.

In this article, a city system consists of multiple layers of cities; cities of the same layer have the same functions, i.e. they host the same set of industries. This model can be viewed as a multiple-goods extension of that proposed by Lederer and Hurter (1986), in which firms play a two-stage game: they compete on price in each local market after making their location choices. The differentiation among cities is driven by the heterogeneity of the scale economies among goods, which can be modelled by the heterogeneity in the fixed cost of production, demand or transportation cost. For clean exposition, the analysis is carried out with heterogeneity in the fixed cost only. However, it will become clear why the case with heterogeneity in demand or transportation cost is isomorphic to the case of fixed cost (see Section 1.3).

Two important results of this model are that a class of equilibria consistent with the hierarchy property exists, and there is always only one next-layer city between two neighbouring larger cities. The latter result, which I call the *central place property,* implies that the ratio of the market areas of one layer to the next is two, which is analogous to the $K = 3$ *market principle* put forward by Christaller (1933). (In the extension to the plane, if there is always only one next-layer city located in the equilateral triangle area among three neighbouring larger cities, then the ratio of market areas is three.) In this article, a *central place hierarchy* is a city system in which both the hierarchy and central place properties hold.[5]

This article makes four contributions. First, it provides a new and parsimonious model as a microfoundation for central place theory. Although this theory has a long history in economic geography, its modelling has been mechanical. (See Fujita *et al.* (1999*b*), for a critique of the lack of a microfoundation in the central place theory literature.) A few attempts to formalise central place theory have been made in the economics literature (Eaton and Lipsey, 1982, Quinzii and Thisse, 1990, Fujita *et al.*, 1999*a*, Tabuchi and Thisse, 2011). Fujita *et al.* (1999*a*) generate a hierarchical city system through an evolutionary approach that extends the core-periphery model of Fujita and Krugman (1995). Given the existence of the first and largest city, they

[4] This property is often referred to in the literature as the *hierarchy principle.*

[5] Following Henderson's (1974) type-of-cities theory, there is an extensive literature on city hierarchy that addresses how cities specialise in different industries and how industrial fundamentals affect city size. Central place theory prescribes a plausible pattern of diversity and a specific pattern of specialisation, but it does not allow sufficient room to explain more flexible patterns of specialisation across cities. In this view, central place theory and type-of-cities theory are complements rather than substitutes in modelling city hierarchy. See the conclusion for further discussion of this issue.

successfully model the hierarchy property. Compared with this article, however, theirs does not feature the central place property and it has no particular implications for city size distribution.[6]

Second, this article demonstrates that the power law for cities arises from *regularly varying* distributions of scale economies, a rather general class that includes many well-known, commonly used distributions. This is shown in two steps. I first specify the condition under which the central place hierarchy becomes a *fractal structure*, in which the smaller parts have shapes that are similar to the larger parts, which leads to the power law (Proposition 2). Then, I show that regularly varying distributions of scale economies ensure that this condition holds (Proposition 3). Although Beckmann (1958) may have been the first to point out the link between central place theory and the power law, he provides no economic model for his hierarchical structure, and his hierarchy lacks the dimension of industries, which is crucial in this article. Also note that the regularly varying distributions (distributions with a regularly varying left tail) in this article are of a different class from the *regular* distributions defined in Gabaix and Landier (2008), and neither class is a subset of the other. It is worth emphasising, however, that both articles employ the general properties of extreme value theory to make predictions independent of the details of the distributions.

Third, the model is consistent with the number-average-size (NAS) rule, a recently discovered empirical regularity regarding industrial locations that was first documented by Mori *et al.* (2008) using Japanese data. The NAS rule states that the relationship between the number and average size of the cities in which an industry is located is linear on a log-log scale with a negative slope. To illustrate, suppose that Major League Baseball is an industry. As there are 26 cities (MSAs) that host at least one team, the number of concern is 26, and the average size of concern is simply the average population of these 26 cities. This regularity provides a view of how city sizes and industrial activities are related, and it is important to note the result in Mori *et al.* (2008) that if the hierarchy property holds, then the NAS rule is essentially equivalent to the power law. I also examine US industrial-location data using the approach of Mori *et al.* (2008) and find that the NAS rule holds very well for three- and four-digit industries as defined by the North American Industry Classification System (NAICS). The details of this empirical evidence are presented in Appendix A.

Fourth, this article derives the power law for firms, the evidence for which is provided by Axtell (2001) and Luttmer (2007). Zipf's law seems to hold better for firm size than for city size because the slope is very close to one. Note that a central place hierarchy also leads to a hierarchy of firms, as those firms that are located exclusively in large cities serve larger market areas (i.e. more consumers) and thus must be larger. This article provides a different angle from that in the literature on the power law for firms,

[6] Eaton and Lipsey (1982) provide a spatial competition model in which there are two goods and multiple firms for each good. Based on multipurpose shopping, they show how firms producing two different goods may co-locate. In this sense, they thus model a primitive hierarchy property. Quinzii and Thisse (1990) provide a rationale for the hierarchy property via a social planner's problem. However, neither of these articles obtains the central place property. Tabuchi and Thisse (2011) model both the hierarchy and central place properties. Although their hierarchical structure is the most similar to that in this article, their model is very different, and they do not analyse city size distribution.

which also draws predominantly upon random growth processes (see, for example, Simon and Bonini, 1958; Luttmer, 2007).

The modelling in this article proceeds in two steps. In the first step, a basic model is developed to deliver the fundamental hierarchical structure. A closed-form solution is readily obtained, which provides a tractable framework for investigation of the conditions for the power laws and the NAS rule. However, the equilibria giving rise to central place hierarchies constitute only a subset of the continuum of equilibria. In the second step, an agglomeration force is created by incorporating workers' demand, and the central place hierarchy as an equilibrium becomes locally unique.[7] A numerical exercise shows that this generalisation does not change the power-law feature of the city size distribution, although the distribution becomes more skewed.

This article models central place hierarchies as equilibrium results. See Hsu and Holmes (2009) for a social planner's problem that generates these hierarchies via a dynamic programming approach to central place theory. The rest of the article is organised as follows. Section 1 lays out the basic model and derives central place hierarchies as equilibrium results. Section 2 specifies the conditions under which the power laws and the NAS rule hold. Section 3 generalises the model by adding workers' demand and analyses its effects on city size distribution. Section 4 concludes the article.

1. Central Place Theory: The Basic Model

1.1. *Model and One-good Equilibrium*

The geographic space is the real line, and the location is indexed as $x \in \mathbb{R}$. There is a [0,1] continuum of consumption goods and two types of agents: farmers and firms. The farmers are immobile and are uniformly distributed on the real line with a density of one. Each farmer demands one unit of each good in [0,1] inelastically.

For any good to be produced at a given location, a fixed cost is required to set up production. Denote the fixed cost of production for a good as y, and denote the (cumulative) distribution function of y as $F : [\underline{y}, \bar{y}] \subset \mathbb{R}_+ \rightarrow [0, 1]$. Producing one unit of each good requires constant marginal cost c. The transportation cost is t per unit per mile travelled. For each good, there is an infinite pool of potential firms. The firms and farmers play the following two-stage game (Lederer and Hurter, 1986).

(*i*) *Entry and location stage* The potential firms simultaneously decide whether to enter. Upon entering, each entrant chooses a location and pays the fixed cost to set up at that location. Assume the tie-breaking rule: if a potential firm sees a zero-profit opportunity, then it enters.

(*ii*) *Price competition stage* The firms deliver goods to the farmers. Given its own and other firms' locations, each firm sets a delivered price schedule over the real line. For each good, each location on the real line is a market in which the firms

[7] Note that the central place hierarchy in this article is also locally stable. It is interesting to note the similarity that although there are also multiple equilibria in the models of Fujita *et al.* (1999*a*) and Tabuchi and Thisse (2011), the hierarchy equilibria in these models are also locally stable. Also see Berliant (2006) for a discussion of multiplicity in these models.

engage in Bertrand competition. Each farmer decides the specific firm from which to buy each good.

Consider the subgame perfect equilibrium (SPNE) of any particular good y. Consider two neighbouring firms at a distance of L. Denote the firm on the left-hand side as A and that on the right-hand side as B. The marginal costs of delivering the good to a consumer who is x distance from A are thus:

$$MC_A = c + tx,$$
$$MC_B = c + t(L - x).$$

Bertrand competition at each x results in the firm with the lower marginal cost grabbing the market and charging the price of its opponent's marginal cost. Without loss of generality, let A be located at 0. Thus, the equilibrium prices on $[0, L]$ can be written as

$$p(x) = \begin{cases} c + t(L - x) & x \in [0, \frac{L}{2}], \\ c + tx & x \in [\frac{L}{2}, L]. \end{cases}$$

The gross profit for firm A from the market area on its right-hand side and that for B from that on its left-hand side are both $tL^2/4$. Figure 2 illustrates the marginal cost of both firms and the equilibrium price, as well as the gross profits from the market area between A and B.

Consider any entrant's strategy at the first stage. Let this entrant be named C. If C were to enter into a market area between A and B, then it is straightforward to show that C's profit-maximising location would be exactly in the middle of the two, given A and B's locations. Any deviation from the middle will strictly decrease C's profit, and C will enter if and only if this maximal profit is non-negative. Therefore, firms must be an equal distance apart, and the gross profit of any firm with a market area of L is $tL^2/2$. Note that there is no room for arbitrage, and firms can thus exercise price discrimination effectively. To see this, refer to Figure 2. For any consumer located at

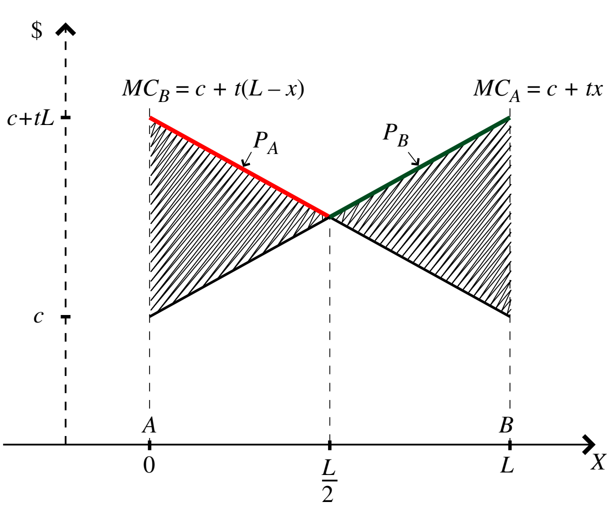

Fig. 2. *Second-stage Competition: Prices and Gross Profits*

$x \in [0, L/2]$, the marginal cost of selling a product purchased at x to the consumers on their left is the same as $MC_B = P_A$, and that of selling it to those on their right must be larger than the equilibrium price. Thus, it is impossible for this consumer to make a profit by reselling the product. The foregoing derivation of an SPNE for an arbitrary good leads to Lemma 1.

LEMMA 1. *Fix the level of fixed cost y and define $\underline{L}(y)$ as the solution to the zero-profit condition $t\,[\underline{L}\,(y)]^2/2 = y$. Thus, $\underline{L}(y) = \sqrt{2y/t}$. There is a continuum of equilibria in which one firm is located at every point in $\{x + nL\}_{n=-\infty}^{\infty}$, where $L \in [\underline{L}(y), 2\underline{L}(y))$ and $x \in [0, \underline{L}(y))$.*

The continuum of equilibria exists because any distance L in the interval $[\underline{L}(y), 2\underline{L}(y))$ is an equilibrium distance; $L \geq \underline{L}(y)$ implies that all firms earn a non-negative profit (no exit), whereas $L/2 < \underline{L}(y)$ implies that any new entrant between any two existing firms must earn a negative profit (no entry).

1.2. *Hierarchy Equilibrium*

An equilibrium is a collection of locations of firms, delivered price schedules, and farmers' consumption choices such that the allocation for each good y is an SPNE. In this article, I consider a particular type of equilibrium, that is, one in which the hierarchy property holds.

DEFINITION 1. *A hierarchy equilibrium is an equilibrium in which, at any production location, the set of goods produced must take the form $[\underline{y}, y]$ for some level of fixed cost y.*

In a hierarchy equilibrium, a decreasing sequence $\bar{y} = y_1 > y_2 > \cdots > y_I \geq \underline{y}$, exists for some $I \in \mathbb{N} \cup \{\infty\}$, denoting the cut-offs. A hierarchy equilibrium is said to satisfy the *central place property* if the market area of the firms producing $(y_{i+1}, y_i]$ is half that of the firms producing $(y_i, y_{i-1}]$.

DEFINITION 2. *A hierarchy equilibrium that satisfies the central place property is called a central place hierarchy. Due to the hierarchy property, any production location produces goods in the range of $[\underline{y}, y_i]$ for some y_i, and it is called a layer-i city.*

In fact, it turns out that any hierarchy equilibrium is a central place hierarchy, and it is characterised as follows. Fix an $x \in \mathbb{R}$ and set the grid for $(y_{i+1}, y_i]$ as $\{x + nL_i\}_{n=-\infty}^{\infty}$, where $L_1 = m\underline{L}[\bar{y}]$, $m \in [1, 2)$, $L_i = L_1/2^{i-1}$, and the cut-off y_i is given by the zero-profit condition

$$y_i = \frac{tL_i^2}{2} \quad \forall\, 2 \leq i \leq I, \tag{1}$$

where the number of layers I will be specified later. Without loss of generality, let $x = 0$. Then, the location configuration so constructed is precisely that given in Figure 3, except that only four layers are depicted in the Figure. By construction, both the hierarchy and central place property are satisfied.

The fact that the foregoing construction is an equilibrium can be seen in the following recursive view. The good $\bar{y}$ is in equilibrium because $m \in [1, 2)$ (Lemma 1). For all $y \in (y_2, \bar{y}]$, firms earn a positive profit, as they share the same market area as

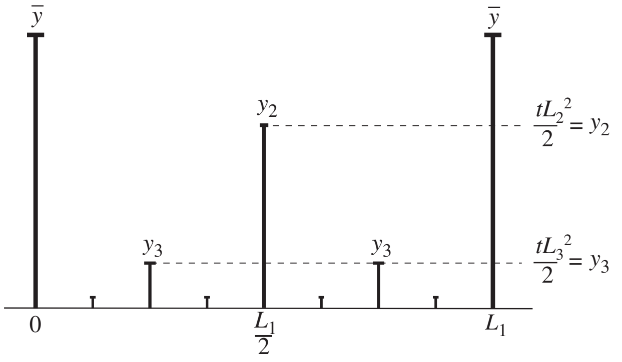

Fig. 3. *A Central Place Hierarchy*

Notes. The layer-i cities produce goods in $[\underline{y}, y_i]$. The cut-offs y_i are determined by the zero-profit conditions. The market areas for goods $(y_{i+1}, y_i]$ are half of that for $(y_i, y_{i-1}]$.

type-$\bar{y}$ firms but they have smaller fixed costs. As y_2 is determined by the zero-profit condition, no additional entrants of $y \in (y_2, \bar{y}]$ will enter between two existing firms, because any such entrant would have the same market area as a type-y_2 firm but with a larger fixed cost. This proof applies recursively; because type-y_2 firms earn zero profits, all firms with $y \in (y_3, y_2]$ earn positive profits and type-y_3 firms earn zero profits etc. As the gross profits $tL_i^2/2$ shrink over the layers toward 0, the number of layers I is determined by the last $y_i = tL_i^2/2$ such that $y_i \geq \underline{y}$. Thus, from $tL_I^2/2 \geq \underline{y}$ and $tL_{I+1}^2/2 < \underline{y}$, we get

$$I = \left\lfloor \frac{2\ln(m) + \ln(\bar{y}/\underline{y})}{2\ln(2)} + 1 \right\rfloor, \tag{2}$$

where $\lfloor x \rfloor$ denotes the largest integer that is smaller than or equal to a real number x. Note that (2) already implies that $I = \infty$ if $\underline{y} = 0$.

To see that the central place hierarchy is the unique type of hierarchy equilibrium, first note that any candidate for a hierarchy equilibrium must have $n_i \in \mathbb{N}$ layer-i cities that are evenly spaced between the two neighbouring larger cities. This spacing occurs because, in equilibrium, all firms producing the same good must be evenly spaced (Lemma 1). The hierarchy equilibrium constructed above is simply a special case where $n_i = 1$ for all i. Now, suppose that all goods in $(y_i, \bar{y}]$ are in equilibrium but $n_i \geq 2$ and type-y_i firms earn zero profits. However, this implies that a new entrant with a y slightly larger than y_i entering the middle point between two existing type-y firms can earn a positive profit, as its market area is larger than any type-y_i firms and its fixed cost is only infinitesimally larger. This contradicts the condition that all goods in $(y_i, \bar{y}]$ are in equilibrium. Hence, for any hierarchy equilibrium, $n_i = 1$ for all i. The results thus far are summarised in Proposition 1.

Proposition 1 (Central place hierarchy). *For each $L_1 = m\underline{L}(\bar{y})$, $m \in [1, 2)$, let $L_i = L_1/2^{i-1}$, y_i be given by the zero-profit condition* (1) *and the number of layers I be given by* (2). *Fix an $x \in \mathbb{R}$, and set the grid for $(y_{i+1}, y_i]$ as $\{x + nL_i\}_{n=-\infty}^{\infty}$. Then, for each $m \in [1, 2)$, the location configuration so constructed is the unique hierarchy equilibrium and satisfies the central place property.*

1.3. *Alternative Sources of Heterogeneity*

As noted in the Introduction, the heterogeneity in scale economies across goods can also be modelled in terms of the heterogeneity in demand or transportation cost. Suppose that each individual consumes y units of good y, and that the fixed cost of production is the same across goods and normalised to one. We can use the same notation for the distribution function of $y : F(\cdot)$. It is an easy exercise to derive all of the results in this article step-by-step, except that the order of goods in terms of the size of the market areas is now reversed. To see this, first note that the gross profit of firms producing y is $ytL^2/2$. Then, under a central place hierarchy, the zero-profit condition determining the cut-off y_i is

$$\frac{1}{y_i} = \frac{tL_i^2}{2}. \tag{3}$$

Alternatively, suppose that both the demand and the fixed cost are normalised to one and that y represents the transportation cost with the distribution function $F(\cdot)$. Then, the gross profit of firms producing y is $yL^2/2$. Thus, we get the same zero-profit condition for the cut-off y_i given by (3) with $t = 1$. The rest of the derivation is similar.

1.4. *Discussion*

As in Fujita *et al.* (1999*b*) or similar models, I employ immobile farmers as a primitive assumption and the uniform distribution for simplicity. Such assumptions are usually understood as indicating that there are resources throughout the geographic space that require on-site extraction/utilisation, such as farming, fishing, mining etc. It does not matter whether the farming/fishing/mining sector accounts for a small or large proportion of the population (the fraction of farmers is reflected by the demand from farmers, and this is normalised to one here); as long as there are people throughout the space who are immobile and do not provide all goods to themselves, dispersed towns and cities will emerge as 'central places' for these immobile people and smaller towns.

2. Power Laws and the NAS Rule

2.1. *Power Law for Cities*

In a central place hierarchy, all firms in the range $(y_{k+1}, y_k]$ produce for a market of size L_k. Thus, the output of the firms in this range is $\hat{Y}_k \equiv L_k[F(y_k) - F(y_{k+1})]$. Define the size of a layer-i city by the total units produced in that city (as a measure of the level of economic activity):

$$Y_i \equiv \sum_{k=i}^{I} \hat{Y}_k = \sum_{k=i}^{I} L_k[F(y_k) - F(y_{k+1})].$$

Figure 4 illustrates the definitions of Y_i. The area shaded with lines and the area shaded with dots represent the total quantity produced in a layer-1 and a layer-2 city respectively.

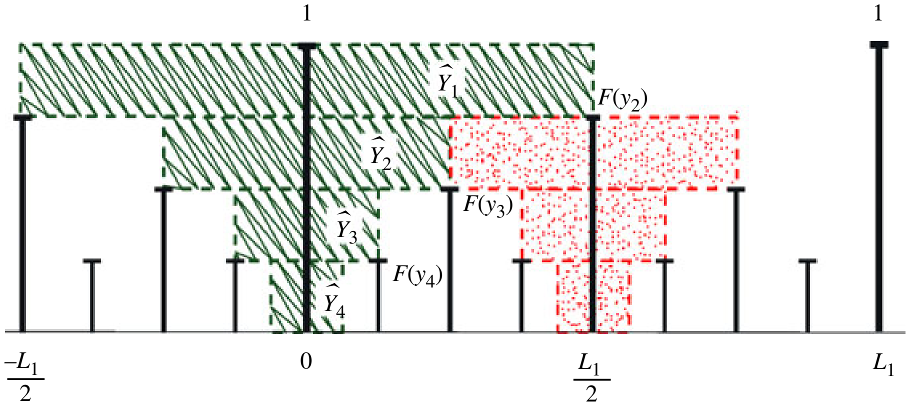

Fig. 4. *City Size*

Notes. The area shaded with lines and the area shaded with dots denote the size of a layer-1 and a layer-2 city respectively. Both shaded areas are composed of rectangles, each of which represents the total production of the respective range of goods.

For every layer-1 city, there is one layer-2 city and 2^{i-2} layer-i cities. Thus, the total number of cities up to layer-i is

$$R_i = 1 + 1 + \sum_{k=3}^{i} 2^{k-2} = 2^{i-1}.$$

We are interested in the relationship between size Y_i and R_i, which approximates the ranks of the layer-i cities. Note that because the rank doubles from one layer to the next, Zipf's law is approximated if city size Y_i shrinks by approximately half from one layer to the next (the rank-size rule!). Similarly, if city size Y_i shrinks by an approximately constant fraction from one layer to the next, then the power law is approximated.

It is worth pointing out that the above account of the ranks of cities involves looking at a region/city system of length L_1 that is a clone of many others in the geographic space. An alternative view which better reflects the hierarchial relations across cities is proposed as follows. If one city B is located within the market area of a larger city A, then it is within A's *economic hinterland*. As this relation is recursive in the spatial structure, it is best to define the economic hinterland of a city as a subset of the geographic space in which all of the goods that are produced are also consumed within the same subset. That is, even when a smaller city is not in the market area of A, if it is in the market area of B and B is in the market area of A, then the smaller city is also in A's economic hinterland. Figure 5 shows an example of the economic hinterland of a layer-1 and a layer-2 city. Note that the layer-1 cities' economic hinterlands partially overlap, whereas the regions/city systems in the first view do not overlap.

Within any layer-1 city's economic hinterland, the number of cities up to the ith layer is given by $R_i = \sum_{k=1}^{i} 2^{k-1} = 2^i - 1$. For sufficiently large is, $R_i/R_{i+1} \approx 1/2$. For simplicity, the following analysis is carried out using the first view of non-overlapping regions but all the results also hold asymptotically in this hinterland view.

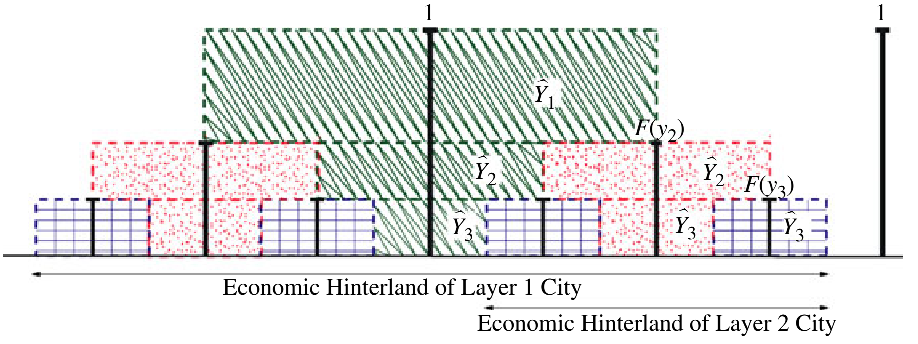

Fig. 5. *Alternative View of City Hierarchy*

Notes. Only three layers of cities are shown here. The economic hinterlands of a layer-1 and a layer-2 city are indicated. The shaded areas are the city sizes.

2.1.1. *A simple condition for the power law*

There is, indeed, a simple but powerful condition that directly links central place hierarchies and the power law, regardless of the underlying economics behind that hierarchy. Given a central place hierarchy, the location patterns of cities of different layers are fixed and different underlying economics matters only in relation to how the fractions of goods ($z_i = F(y_i)$) in the different layers are determined. The following Proposition specifies the condition for the fractions of goods that renders the central place hierarchy a fractal structure.

Proposition 2 (Bounds on fraction ratios). *Suppose that there are infinitely many layers in a central place hierarchy. Let z_i denote the fraction of goods produced in a layer-i city, and let $\Delta_k = z_k - z_{k+1}$. Suppose that there is a $\delta > 0$ and a $\rho > 1$, such that for all $i \in \mathbb{N}$,*

$$\frac{\delta}{\rho} \leq \frac{\Delta_{i+1}}{\Delta_i} \leq \rho\delta.$$

Then,

$$\frac{1}{2}(\rho^{-1} - 1)\delta \leq \frac{Y_{i+1}}{Y_i} - \frac{\delta}{2} \leq \frac{1}{2}(\rho - 1)\delta. \tag{4}$$

Proof. Observe $Y_i \propto \sum_{k=i}^{\infty} \Delta_k / 2^k$. For weights

$$w_{k,i} \equiv \frac{\frac{\Delta_{i+k}}{2^{i+k}}}{\sum_{k=0}^{\infty} \frac{\Delta_{i+k}}{2^{i+k}}},$$

we have

$$\frac{Y_{i+1}}{Y_i} - \frac{\delta}{2} = \frac{1}{2}\sum_{k=0}^{\infty} w_{k,i}\left(\frac{\Delta_{i+k+1}}{\Delta_{i+k}} - \delta\right).$$

Therefore,

$$\frac{1}{2}(\rho^{-1}-1)\delta \leq \frac{1}{2}\sum_{k=0}^{\infty} w_{k,i}\left(\frac{\Delta_{i+k+1}}{\Delta_{i+k}}-\delta\right) \leq \frac{1}{2}(\rho-1)\delta.$$

Observe that δ is approximately the ratio of the increments (Δ_k). The approximate slope of the Zipf plot is

$$\frac{\ln(R_{i+1}/R_i)}{\ln(Y_{i+1}/Y_i)} \approx \frac{\ln(2)}{\ln(\delta/2)} = -\frac{\ln(2)}{\ln(2)-\ln(\delta)}.$$

The closer the bound ρ on the ratios of the increments between two adjacent layers is to one, the better the approximation to the power law. In other words, Zipf's law requires only that the increments of the fraction of goods of two adjacent layers do not vary too much $(\delta = 1)$, whereas the power law relaxes the ratio of increments between two layers from one.

2.1.2. *Regularly varying distributions*

The next question, naturally, is how the behaviour of $F(\cdot)$ translates into a power law for cities. A few basic concepts of regular variation are needed.[8]

DEFINITION 3. *A measurable, positive function g is said to be regularly varying at zero (at infinity) if, for any $u > 0$, and for some $\alpha \in \mathbb{R}$,*

$$\lim_{y\downarrow 0\ (\to\infty)} \frac{g(uy)}{g(y)} = u^{\alpha}.$$

If $\alpha = 0$, then g is said to be slowly varying. A function g is regularly varying with index α if and only if there exists a slowly varying function $\ell(y)$ such that

$$g(y) = y^{\alpha}\ell(y).$$

In what follows, $g \in RV_\alpha$ denotes that g is regularly varying at zero with index α.[9] Suppose that $\underline{y} = 0$, and hence there are infinitely many layers. Recall from Proposition 1 that $y_{k+1} = y_k/4$ for all $k \geq 2$. Observe that the ratio between the increments between two layers can be written as

$$\delta_k \equiv \frac{\Delta_{k+1}}{\Delta_k} = \frac{F(y_{k+1}) - F(y_{k+2})}{F(y_k) - F(y_{k+1})} = \frac{1 - F(y_{k+1}/4)/F(y_{k+1})}{F(4y_{k+1})/F(y_{k+1}) - 1}.$$

According to Definition 3, if $F \in RV_\alpha$, then in a small enough neighbourhood of 0, there are infinitely many k values such that

$$\delta_k \approx \frac{1-(1/4)^{\alpha}}{4^{\alpha}-1} = \left(\frac{1}{4}\right)^{\alpha}. \tag{5}$$

By Proposition 2, the power law is approximated with a tail index close to $1/(1 + 2\alpha)$.

A distribution function F on $(0, \bar{y}]$ can be regularly varying only with a non-negative index α because an $F \in RV_\alpha$ on $(0, \bar{y}]$ with $\alpha < 0$ must be decreasing in a small neigh-

[8] See Bingham *et al.* (1987) for a textbook/encyclopedia treatment of regular variation.

[9] We are not concerned with the regularly varying function at infinity because the analysis here focuses on the source of heterogeneity being the fixed cost. See more discussion at the end of this subsection.

bourhood of 0, which violates the requirement of a distribution function.[10] However, a distribution function can be defined via a transformation of a non-increasing function $G \in RV_\alpha$ with $\alpha < 0$:

$$F(y) \equiv \frac{G(\underline{y}) - G(y)}{G(\underline{y}) - G(\bar{y})}, \tag{6}$$

where the domain of F is $[\underline{y}, \bar{y}]$ for some $\underline{y} > 0$. For a $\underline{y}$ that is close enough to zero, such an F behaves like a regularly varying function with a negative index α. This is because, for a $\underline{y}$ close enough to 0, there exists a sufficiently small neighbourhood of $\underline{y}$ such that, for all y_{k+1} in that neighbourhood:

$$\delta_k = \frac{F(y_{k+1}) - F(y_{k+2})}{F(y_k) - F(y_{k+1})} = \frac{G(y_{k+2}) - G(y_{k+1})}{G(y_{k+1}) - G(y_k)} = \frac{1 - G(y_{k+1}/4)/G(y_{k+1})}{G(4y_{k+1})/G(y_{k+1}) - 1} \approx \frac{1 - (1/4)^\alpha}{4^\alpha - 1} = \left(\frac{1}{4}\right)^\alpha.$$

In any case, when the index α associated with the distribution function is positive (negative), then the slope of the Zipf plot is smaller (greater) than one. The following proposition summarises the foregoing discussion and provides statements based on the density functions.[11]

Proposition 3 (Regularly varying distributions). *Let $\delta = (1/4)^\alpha$ and fix any $\rho > 1$. Then, for a sufficiently small $\underline{y} \geq 0$, there exists an integer $K > 0$ such that condition* (4) *holds for all layers $I \geq i \geq K$ (with the possibility that $I = \infty$), if one of the following conditions is met:*

(a) the distribution function of fixed cost $F \in RV_\alpha$ with $\alpha \in [0, \infty)$;

(b) $G \in RV_\alpha$ with $\alpha \in (-1/2, 0)$ such that F is defined by (6);

(c) the density function of fixed cost $f \in RV_{\alpha-1}$, for $\alpha \in (-1/2, \infty)$.[12]

In all cases, the approximate slope of the Zipf plot, i.e. the plot of log of rank on log of size, is $-1/(1 + 2\alpha)$.

Proof. See Appendix B for the proof of the density part.

The class of distributions that satisfies the conditions in Proposition 3 is rather general, as it includes several well-known, commonly used distributions, such as the Pareto, Weibull, F, Beta (which subsumes the uniform) and Gamma, which subsumes the chi-square, exponential and Erlang. See Appendix C for more examples.[13]

[10] See Lemma 1 in Bingham and Teugels (1975). Note that most expositions in the literature are on functions that are regularly varying at the right-tail. All the results used in this article are from the literature, e.g. Bingham and Teugels (1975), Bingham *et al.* (1987), hold for functions that are regularly varying at the left tail. One can check the results by noting that when a function $g(y)$ is regularly varying at infinity with index α, then $h(y) \equiv g(1/y)$ is regularly varying at zero with index $-\alpha$.

[11] Note that $F \in RV_\alpha$ does not imply the density function $f \in RV_{\alpha-1}$ unless the density in the neighbourhood of 0 is monotone. See Bingham *et al.* (1987), Theorem 1.7.2*b*. For most parametric distributions, a regularly varying $F(\cdot)$ implies a regularly varying density, and *vice versa*. However, statements regarding density are still useful, as there are no closed-form expressions for some distribution functions.

[12] In (*b*), it must be the case that $\underline{y} > 0$ and hence $I < \infty$.

[13] Moreover, the convolution of regularly varying distribution functions is also regularly varying (Jessen and Mikosch, 2006), as is the convolution for densities (Bingham *et al.*, 2006). If a sequence of random variables on $(0, \infty)$ is *i.i.d.* and regularly varying, then all of the polynomials statistics are also regularly varying (Jessen and Mikosch, 2006).

In the case in which the heterogeneity of scale economies is modelled by the demand or transportation cost, it is the right tail of $F(\cdot)$ that matters, as most cities host industries with small scale economies (large demand or a large transportation cost). Hence, the power law arises when $F(\cdot)$ has a regularly varying right tail.

2.2. *The Power Law for Firms and the NAS Rule*

2.2.1. *The power law for firms*

Not only is the central place hierarchy a hierarchy of cities, it is also a hierarchy of firms. Any firm producing $y \in (y_{i+1}, y_i]$ has the size $L_i = L_1/2^{i-1} \equiv s_i$. Call this size type of firms class-i. Note that the measure of class-i firms, m_i, is $2^{i-1}[F(y_i) - F(y_{i+1})] = 2^{i-1}(z_i - z_{i+1}) = 2^{i-1}\Delta_i$. Consider the accumulative measure of firms of classes up to i as the rank of class-i firms:

$$M_i = \sum_{k=1}^{i} m_k.$$

Indeed, the conditions required to ensure that a firm size distribution follows the power law are the same as those for a city size distribution, although the implied slopes are different.

Proposition 4 (Bounds proposition on firm size distribution). *Suppose that there are infinitely many layers in a central place hierarchy, and there is a $\delta > 0$ and $\rho > 1$ such that, for all $i \in \mathbb{N}$, $\frac{\delta}{\rho} \leq \Delta_{i+1}/\Delta_i \leq \rho\delta$. Then,*

$$\frac{(2\delta/\rho)^{i+1}-1}{(2\delta/\rho)^{i}-1} \leq \frac{M_{i+1}}{M_i} \leq \frac{(2\rho\delta)^{i+1}-1}{(2\rho\delta)^{i}-1}.$$

Proof. See Appendix B.

Proposition 4 implies that

$$\lim_{\rho\to 1} \frac{M_{i+1}}{M_i} = \frac{(2\delta)^{i+1}-1}{(2\delta)^{i}-1},$$

which means that M_{i+1}/M_i is approximately 2δ, and the slope of the Zipf plot for firms is $-[1 + \ln(\delta)/\ln(2)]$ for large enough i's and for a ρ sufficiently close to one. If the distribution of fixed cost $F \in RV_\alpha$, then the slope of the Zipf plot is approximately $-(1 - 2\alpha)$. When α is close to 0, both the city and firm size distributions are close to Zipf's law.

2.2.2. *The NAS rule*

Recall that the number of cities in layer i is 2^{i-2} for $i \geq 2$. Also note that the number of cities producing $(y_{i+1}, y_i]$ is equal to the number of cities up to layer-i (R_i). Thus, the average size of cities producing $y \in (y_{i+1}, y_i]$ is

$$\mathrm{AS}_i = \frac{Y_1 + \sum_{k=2}^{i} 2^{k-2} Y_k}{R_i}. \tag{7}$$

Now, i is not only an index for the layers of cities, but is also an index for the industry groups $((y_{i+1}, y_i])$ located exclusively in cities that are no smaller than layer-i cities. The following proposition presents the implications of the power law on the slopes of the NAS plots, that is, the plots of the log of average size against the log of numbers. This result is in line with that in Mori *et al.* (2008). In the following proposition, the results on how $\ln \mathrm{AS}_i$ changes with $\ln R_i$ for integer i are obtained from analysing how the derivative of $\ln \mathrm{AS}_i$ with respect to $\ln R_i$ changes with i, as if i can potentially be a non-integer number.

PROPOSITION 5 (The NAS rule). *Suppose that $\underline{y} = 0$; hence, there exist infinitely many layers of cities. Suppose that the power law for cities holds such that $Y_{i+1}/Y_i = 2^{-(1+2\alpha)}$, with $\alpha \in (-1/2, 1/2)$.*[14] *If $\alpha \in [0, 1/2)$, then $\mathrm{d}\ln(\mathrm{AS}_i)/\mathrm{d}\ln(R_i)$ is strictly decreasing in i, and*

$$0 > \frac{\mathrm{d}\ln(\mathrm{AS}_i)}{\mathrm{d}\ln(R_i)} > -1, \tag{8}$$

$$\lim_{i\to\infty} \frac{\mathrm{d}\ln(\mathrm{AS}_i)}{\mathrm{d}\ln(R_i)} = -1. \tag{9}$$

If $\alpha \in (-1/2,\ 0)$, then

$$\lim_{i\to\infty} \frac{\mathrm{d}\ln(\mathrm{AS}_i)}{\mathrm{d}\ln(R_i)} = -(1+2\alpha). \tag{10}$$

Proof. See Appendix B.

Using Japanese data, Mori *et al.* (2008) and Mori and Smith (2011) find that the NAS plot is quite linear with a slope of around -0.7. As reported in Appendix A, the NAS plot using US data exhibits similar features with a slope of around -0.75. Note that the distinction between the two cases in Proposition 5 is whether $\alpha < 0$. In both cases, even though they are different in their implications on Zipf's plot in terms of whether the corresponding slope is greater than one, the implied NAS slope is <1, which is consistent with the empirical findings. To see this, observe that if $\alpha \in [0,\ 1/2)$, the local slope (in absolute value) is < 1 and converging to one. Thus, if finite but numerous layers are observed in the data, then the NAS plot should be quite linear with a slope of <1. When $\alpha \in (-1/2,\ 0)$, then the slope converges to $1 + 2\alpha < 1$, which again implies a quite linear NAS plot with a slope of < 1.

[14] Note that $\alpha < 1/2$ is not a bad restriction for city size distribution, as we rarely observe tail indices of $<1/2$ in city size data; see Soo (2005). In this model, a tail index greater than 1/2 implies $\alpha < 1/2$.

3. The General Model

In the basic model, given any hierarchy equilibrium, another equilibrium can be obtained by deviating the locations of the firms producing any particular good by moving their locations by the same distance and in the same direction. This is because such deviation changes nothing in terms of pricing and profits. In this Section, a more general model is presented by incorporating workers' demand to create an agglomeration force that refines the hierarchy equilibrium. A numerical exercise is carried out to examine how this incorporation affects the city size distribution.

3.1. *Model and Equilibrium*

Although the current model seems to be a partial equilibrium one, it actually has the following general equilibrium interpretation. Assume that farmers are immobile and that each farmer, using a unit of land and their unit labour endowment, produces a units of the agricultural good, which is set as the *numeraire* as well as the unit of utils. All costs and prices in the previous sections are hence denominated in the agricultural good. Following the literature and for tractability, the transportation cost of the agricultural good/*numeraire* is set to zero. Consuming one or more than one unit of any $y \in [\underline{y}, \bar{y}]$ entails $\bar{p}$ units of utils, and the utility of a farmer at location x is hence

$$u(x) = \int_{\underline{y}}^{\bar{y}} \max\{\bar{p} - p(x, y), 0\}\ \mathrm{d}F(y) + a.$$

Then, simply allow sufficiently large a and $\bar{p}$ so that each farmer purchases one unit of each y in the the unit measure of manufactures.

In addition to the farmers and firms in the basic model, there is a third type of agent: workers, whose labour is the variable input to produce manufactures. A unit of manufactures requires $\phi \in (0, 1)$ units of labour. The marginal cost of production is then $c(x) = \phi w(x)$, where $w(x)$ denotes the wages at location x. Each worker has one unit of labour endowment and the same preference as the farmers. The difference between farmers and workers is that the latter have endogenous entry and location decisions like firms, and they have the backyard technology needed to home-supply themselves with any manufactures with unit cost r. If a potential worker does not enter, then he or she enjoys a reservation utility of $\underline{u}$.

Again, $\bar{p}$ is sufficiently large that each worker, like each farmer, consumes one unit of each y. Denote the equilibrium prices charged to workers as $p_w(x, y)$. (The prices that workers face are not necessarily the same as those that farmers do because of the existence of workers' reservation price r.) A worker at x has a utility of

$$u(x) = \int_{\underline{y}}^{\bar{y}} (\bar{p} - \min\{p_w(x, y), r\})\mathrm{d}F(y) + w(x). \tag{11}$$

To describe the market clearing conditions, let $j(x, y, x')$ be the indicator function of the purchase decision of a farmer at x' about the type-y good produced by a firm at x, $\imath(x, y)$ be the indicator function of whether a firm producing a type-y good sets up at x, $N(x)$ be the total measure of workers at x, and $Y(x)$ be the total units of all types of goods produced at x. The market clearing for each local labour market and the markets for goods require

$$N(x) = \phi Y(x), \tag{12}$$

$$Y(x) = \int_0^1 \imath(x,y)\left[\sum_{\{x':j(x,y,x')=1\}} N(x') + \int_{-\infty}^{\infty} j(x,y,x')\mathrm{d}x'\right]\mathrm{d}F(y). \tag{13}$$

Definition 4. *An equilibrium is a collection* $\{\imath(\cdot,\cdot), N(\cdot), p_w(\cdot,\cdot), p(\cdot,\cdot), w(\cdot), j(\cdot,\cdot,\cdot), Y(\cdot)\}$ *ensuring that for each type y, the allocation constitutes an SPNE and* (12) *and* (13) *hold. A hierarchy equilibrium is defined in the same way as in the basic model.*

3.2. *Local Uniqueness of the Hierarchy Equilibrium*

If the reservation price r is less than the marginal cost $c(x)$ for all x, then all workers will home-supply themselves with all manufactures. In this case, the equilibrium entry for workers implies that

$$w(x) = \underline{u} - \bar{p} + r \equiv w.$$

Hence, $c(x) = c \equiv \phi(\underline{u} - \bar{p} + r)$. Thus, if $r < c$, or equivalently, $r \leq \phi(\underline{u} - \bar{p})/(1-\phi)$, then all firms derive their profits from farmers only and the model is equivalent to the basic one. If $r = c$, then firms can charge r to attract workers to buy their goods. This is the borderline case that the workers' demand increases the production of all firms, but the firms still do not derive any profits from the workers. Hence, a hierarchy equilibrium in the basic model remains a hierarchy equilibrium in the general model, but the city size distribution changes. In fact, this distribution becomes more skewed, because large cities sell more to workers and hence become even larger. See the next subsection for further details.

If $r > c$, i.e. $r > \phi(\underline{u} - \bar{p})/(1-\phi)$, then firms derive profits from the demand of workers. Suppose that $\underline{y} > 0$, and therefore the number of layers is finite. Let L_I denote the smallest market area among all firms. (In a central place hierarchy, L_I is the distance between any layer-I city and either one of its two neighbouring cities.) If r is not very large such that

$$r < c + tL_I, \tag{14}$$

then r is less than the after-delivery marginal cost of any good supplied by any non-local firm. As r is the second lowest marginal cost in any production location, every worker will buy goods from local firms at a price of r and home-produce the rest. Hence, the cost of living for any worker is, again, r, and the marginal cost is given again by $c = \phi w = \phi(\underline{u} + r - \bar{p})$. When $r > c + tL_I$, firms also sell to workers in other cities, and the marginal costs across locations start to vary. Although it is more realistic to have varying marginal costs, they also render the equilibrium locations intractable. Thus, from this point on, the focus is on the case of $r < c + tL_I$. This condition will hold if t or $\underline{y}$ is sufficiently large.

Any worker who resides at location x without any production receives zero wages, and he or she enjoys a utility smaller than $\bar{p} - c$. Hence, to ensure that no worker enters such a location, we also need the constraint that

$$\underline{u} \geq \bar{p} - c. \tag{15}$$

Under $r > c$, (14) and (15), (13) can be rewritten as

$$Y(x) = \int_0^1 \imath(x, y)\left[N(x) + \int_{-\infty}^{\infty} j(x, y, x')\mathrm{d}x'\right]\mathrm{d}F(y),$$

which entails (18) in the following proposition, which considers a central place hierarchy. Similarly, the profit of a type-y firm with $y \in (y_{i+1}, y_i]$ is

$$\pi(x, y) = \frac{tL_i^2}{2} + (r - c)N(x) - y,$$

which gives rise to the zero-profit condition (17). Further details of the characterisation of a central place hierarchy are given below.

Proposition 6 (Central place hierarchy). *Suppose that the distribution function F is continuous, $r > c$, and* (14) *and* (15) *hold. Then, there exist central place hierarchies characterised by each layer-i city's population N_i, total production Y_i, cut-off good y_i, the largest market area L_i, and the number of layers I, such that the following hold.*

(*i*) *L_1 satisfies*

$$\frac{tL_1^2}{2} + (r - c)N_1 - \bar{y} \geq 0, \frac{tL_2^2}{2} + (r - c)N_2 - \bar{y} < 0. \tag{16}$$

(*ii*) *For $i \geq 2$,*

$$y_i = \frac{tL_i^2}{2} + (r - c)N_i, L_i = \frac{L_1}{2^{i-1}}. \tag{17}$$

(*iii*) *For all $\{i \in 1, 2, \ldots, I\}$,*

$$Y_i = F(y_i)N_i + \sum_{k=i}^{I}[F(y_k) - F(y_{k+1})]L_k, \tag{18}$$

$$N_i = \phi Y_i, \tag{19}$$

where I is the largest integer satisfying $tL_I^2/2 + (r - c)N_I \geq \underline{y}$.

Proof. See Appendix B.

Recall that when $r \leq c$, an equilibrium can be created from any hierarchy equilibrium by sliding all firms' locations for any particular good by the same distance and in the same direction. Call such a deviation a *parallel deviation*. Therefore, any deviation from a central place hierarchy that involves parallel deviations for a subset of goods (possibly with different distances for different goods) remains an equilibrium under

$r \leq c$. Define a Δ-deviation as a deviation that involves a subset of goods, each of which has a parallel deviation with the deviating distance less than or equal to $\Delta > 0$. Define the Δ-neighbourhood of a central place hierarchy as the set of location configurations that involve Δ-deviations from the hierarchy. See Figure 6 for a depiction of the Δ-deviation of a particular good. When $r > c$, Proposition 7 shows that any central place hierarchy is a *locally unique* equilibrium by showing that any central place hierarchy is the only equilibrium in its Δ-neighbourhood when the deviating subset and Δ are both sufficiently small.

PROPOSITION 7 (Local uniqueness). *Suppose that the conditions of Proposition 6 hold. Consider the Δ-deviation of a sufficiently small subset of goods from a central place hierarchy. If $\Delta \in (0, (r - c)/t)$, then the new location configuration does not constitute an equilibrium.*

Proof. See Appendix B.

The intuition behind Proposition 7 is as follows. Consider first the Δ-deviation of a single good y. If Δ is sufficiently small, then the type-y firms still sell to workers in the closest city and earn profits from them. For any particular type-y firm, given other type-y firms' locations, unilaterally moving closer to the closest city by a distance of δ decreases its profit from farmers but increases its profit from workers due to the savings in transportation costs. Such a move with a small enough δ must increase the firm's overall profit because the increase in the profit from workers is linear in the population of the closest city, whereas the decrease in the profit from farmers is linear in δ. Thus, at $\delta = 0$, the marginal profit in moving closer the closest city is positive and the location configuration with this Δ-deviation does not entail a Nash equilibrium in the location stage for type-y firms. This argument easily generalises to the case of a Δ-deviation that involves a sufficiently small subset (possibly with positive measure) of goods, because all that is needed is to ensure that the 'new cities,' if any, formed by the deviating workers

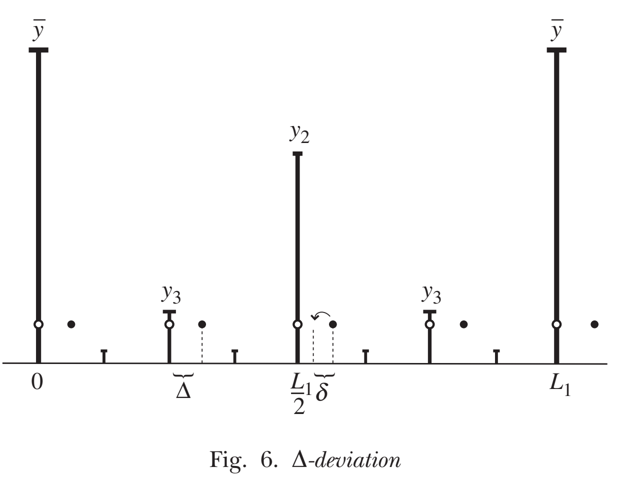

Fig. 6. Δ-*deviation*
Notes. For one particular good, the locations of all firms deviate from the central place hierarchy by a distance of Δ. Given all other firms' locations, one firm contemplates whether to move closer to the closest city by a distance of δ.

are not too large compared with the existing cities in the central place hierarchy. Also note that a central place hierarchy is *locally stable* in the sense that the deviating firms in any off-equilibrium location configuration in a small enough neighbourhood of the hierarchy have incentives to move towards, if not directly jumping back to, their prescribed locations in the hierarchy.

3.3. *City Size Distribution*

Combining (18) and (19) gives

$$N_i = \begin{cases} \phi \sum_{k=i}^{I} L_i[F(y_i) - F(y_{i+1})] & r < c, \\ \dfrac{1}{\phi^{-1} - F(y_i)} \sum_{k=i}^{I} L_i[F(y_i) - F(y_{i+1})] & r \geq c. \end{cases} \tag{A}$$

Let us call the term $Y_i = \sum_{k=i}^{I} L_i[F(y_i) - F(y_{i+1})]$ the base demand, which is actually the definition of city size in the basic model. The city population is just a multiple ϕ of the base demand in the $r < c$ case, and all of the results with regard to the power laws and the NAS rule from the basic model hold. It is difficult to obtain analytical results when $r > c$, as there are no closed-form solutions to cut-offs y_i. However, we can obtain the general effect of adding workers' demand by observing the difference between the $r \geq c$ and $r < c$ cases in (A).

It follows from $\phi \in (0,1)$ that $[\phi^{-1} - F(y_i)]^{-1} > \phi$. Hence, an increase in the base demand increases the city population more in the $r \geq c$ case than in the $r < c$ case, and the larger the ϕ, the larger the difference between $[\phi^{-1} - F(y_i)]^{-1}$ and ϕ. These effects are stronger for larger cities (small i's), as $F(y_i)$ is larger. As $\lim_{\phi \downarrow 0} \phi / [\phi^{-1} - F(y_i)]^{-1} = 1$, the $r < c$ case can be approximated with a small ϕ. Hence, adding workers' demand (i.e. changing from the $r < c$ case to the $r \geq c$ case) or increasing the degree to which workers are needed (i.e. increasing ϕ) actually results in a more skewed city size distribution. Numerical comparative statics of ϕ are carried out to illustrate this point. The algorithm used to compute the equilibrium size distributions is drawn from Proposition 6 and described in Appendix D.

For simplicity, assume that $\underline{u} = \bar{p}$, and thus for any $\phi < 1$, $r - c = (1 - \phi)r > 0$. The F employed is defined by $G(y) = [\ln(1 + \beta y)]^{\alpha}$ via (6). Figure 7 shows the Zipf plots for five values of $\phi = \{0.01, 0.21, 0.41, 0.61, 0.81\}$ and the corresponding estimates of the tail indices.[15] For all ϕ values, there are 10 layers.

Figure 7 verifies that when ϕ increases, larger cities see greater increases in size, which results in a size distribution with a lower tail index, i.e. greater skewness. Most importantly, the addition of workers to the model has little effect on the shape of the Zipf plots, as they are quite linear (the R^2 values are all in the range of (0.989, 0.994)). A plausible explanation for why the power law still holds approximately is as follows.[16] Divide (18) by Y_i, and we get

$$1 = \frac{F(y_i)N_i}{Y_i} + \frac{\sum_{k=i}^{I} L_i[F(y_i) - F(y_{i+1})]}{Y_i}.$$

[15] The parameters are $\underline{u} = \bar{p} = 2$, $\bar{y} = 1$, $\underline{y} = 10^{-6}$, $t = 10^{11}$, $r = 0.3$, and $\alpha = -0.1$

[16] I thank an anonymous referee for pointing this out.

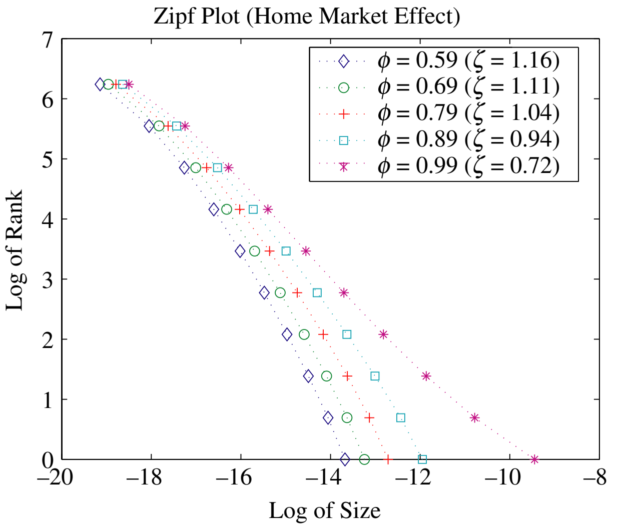

Fig. 7. *Comparative Statics of ϕ on City Size Distribution*
Notes. The values of ϕ and the estimated tail indices (denoted as ζ) are shown.

As $F(y_i)N_i/Y_i = \phi F(y_i) \to 0$ for large is, $\sum_{k=i}^{I} L_i[F(y_i) - F(y_{i+1})]/Y_i$ approaches one for large i's. Most cities are of large i's, and hence if the y_i's do not change drastically, then Y_i, and hence N_i, should exhibit the power law according the analysis in Section 2.

4. Conclusion

This article provides a parsimonious model that formalises central place theory. The differences in city size are driven by the heterogeneity of scale economies, and the central place hierarchy automatically establishes the skewness of the city size distribution. The power law for cities and firms and the NAS rule arise when the distribution of scale economies is regularly varying. In fact, this is the condition for ensuring that a central place hierarchy is a fractal structure. The hierarchy approach to the fractal structure in this article differs from the generic explanations of *self-organised criticality* (Bak *et al.*, 1987) and *scale-free networks* (Barabási and Albert, 1999). As these approaches can explain the power law distributions that are observed for numerous size distributions in the physical, biological and social sciences, it would be interesting to probe the hierarchy approach's potential to explain phenomena other than cities and firms.

The hierarchy property posits that larger cities are more diverse not only because they have more industries than smaller cities but also because they specialise in industries with more scale economies. This is arguably a reasonable view, especially when we look at industries in broader classifications. Mori *et al.* (2008) and Mori and Smith (2011) find that the hierarchy property holds well for 3-digit JSIC but dissipates for 4-digit JSIC (the finest Japanese industrial classification). In Appendix A, in which I examine the NAS rule using US data, I find that this rule holds well for 3 and 4-digit NAICS, but dissipates for 5 and 6-digit NAICS. These findings hint that, at the finest levels, specialisation matters to the extent that heavy industries can be located in small towns due to numerous factors outside central place theory. The development of a comprehensive theory that produces more realistic patterns of diversity and specialisation is a desirable direction for future research.

Interestingly, the basic model presented here can be directly applied to the economic activities within a metropolitan area, as the population distribution of the suburbs of such an area can be viewed as more or less uniformly distributed, especially in US cities that are characterised by a large degree of urban sprawl. Thus, the employment/retail centres in a metropolitan area are analogues to the cities in a wide geographic space. Suppose that the commuting cost within a metropolitan area is zero. Then, the size of an employment/retail centre can simply be defined as the employment/total quantity sold at that centre. Retail activities provide a particularly vivid image of central place patterns (Berry, 1967). It would thus be interesting to examine whether there are corresponding power laws and an NAS rule for the employment/retail centres within metropolitan areas.

Appendix A. Empirical Evidence on the Number-Average Size Rule: the US Case

The empirical literature on central place theory within the field of economic geography consists primarily of case studies. The earliest such research focuses on the spatial patterns of human settlements, with the most notable example, of course, being Christaller's (1933) study of the spatial patterns of cities and towns in southern Germany, which seems to fit the assumption of a homogeneous farming plain. Berry and Garrison (1958*a, b*) provide further case studies of the sizes and functions of central places in Snohomish County, Washington.

Few empirical studies examined the main propositions/assumptions of central place theory using large-scale data, such as national data, before that carried out by Mori *et al.* (2008). These authors document empirical regularities in both the hierarchy property and the NAS rule using Japanese urban-industrial data. To check the hierarchy property, they define the diversity of a city by the number of industries present within it. Using the three-digit Japanese classification of industries (JSIC), they find that in 71% of the industry-city pairs a given industry is present in all cities with diversity greater than that of the given city. Moreover, they show that this finding is extremely unlikely to be a random event.

Recall that the NAS rule states that the number and average size of the cities in which an industry is located follow a log-linear relationship. In Mori *et al.* (2008), the regressions of the log of the average size on the log of the number for the 3-digit JSICs of 1980/1981 and 1999/2000 entail R^2s that are both over 0.99, with the slopes being $-0.7204/-0.7124$.

Using County Business Pattern Data, I find that the NAS rule also holds for the US. Figure A1 shows the NAS plot (log of average size on the vertical axis and log of the number on the horizontal axis) for the 77 three-digit North American Industry Classification System (NAICS) (1997 version). Similar to Mori *et al.* (2008), I exclude the Agriculture, Forestry, Fishing and Hunting (11) and Mining (21) sectors, as they are not considered to have central place functions. Public Administration (92) is also excluded, because its locations are not determined solely by economic incentives and thus could be arbitrary. The 'presence' of an industry in a city is defined as any positive employment in a MSA or combined metropolitan statistical area (CMSA).

The OLS regression gives

$$\log(\text{average size}) = \underset{(0.0141)}{17.789} - \underset{(0.0026)}{0.7477} \log(\text{number}),\ R^2 = 0.9991,$$

where the numbers in parentheses are the standard errors.

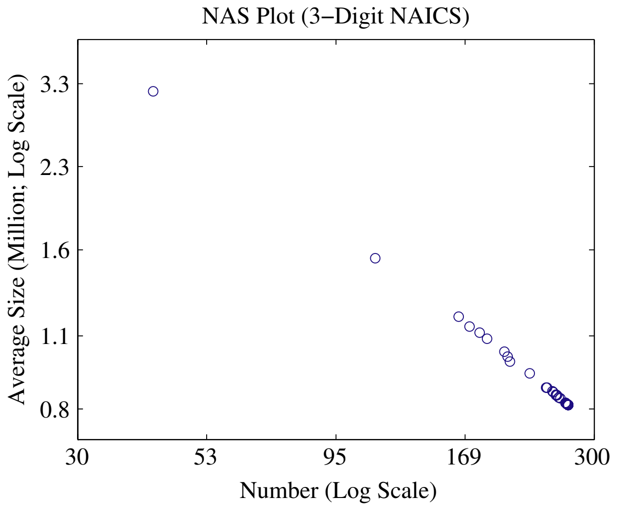

Fig. A1. *Number-average-size Plot for Three-digit NAICS*
Source. 2000 County Business Pattern Data, US Census Bureau.

A straight line fits almost perfectly![17]

It is natural to wonder how the NAS plot and the regression result would change if the NAICS digit or the threshold of 'presence' were changed.[18] Figure A2 shows four variations: 3-digit NAICS and presence defined by the employment of at least 50 people (left-top), 3-digit NAICS and a threshold of employment of at least 100 people (right-top), 4-digit NAICS and any positive employment (left-bottom) and 4-digit NAICS and a threshold of at least 50 people (right-bottom).[19] The (R^2, slope) pairs are (0.997, −0.737), (0.994, −0.721), (0.988, −0.694) and (0.944, −0.622) respectively. Overall, the log-linear relationship is robust across these variations in the definitions of industry and presence.

Appendix B. Proofs

Proof of Proposition 3. For the density part of (a), recall that $f \in RV_{\alpha-1}$ if and only if $f(y) = y^{\alpha-1}\ell(y)$ for some slowly varying $\ell(y)$. Observe that

$$\frac{F(\theta y) - F(y)}{\ell(y)} = \int_y^{\theta y} \frac{\ell(x)}{\ell(y)} \frac{\mathrm{d}x}{x^{1-\alpha}} = y^\alpha \int_1^\theta \frac{\ell(yu)}{\ell(y)} \frac{\mathrm{d}u}{u^{1-\alpha}}.$$

[17] Although the plot shown in Figure A1 consists of 77 three-digit NAICS, 51 of them are located in all of the MSAs; thus, they have the same number (267) and average size (813,544). In light of the limited number of informative sample points in 3-digit NAICS, it is important to note the robustness of the NAS rule in 4-digit NAICS (shown in Figure A2).

[18] The definition of the presence of an industry in a city is a natural challenge to the validity of the NAS rule. To address this issue further, Mori and Smith (2011) have developed a more stringent definition based on the identification of industry clusters. That is, if an industry cluster coincides with the location of a city, then that industry is present in the city.

[19] To determine which MSA/CMSA is excluded by the threshold, I use CBP business employment data by industry and by county. For establishments with more than 1,000 employees, I use cell counts by size class, industry and county taken from Holmes and Stevens (2004). As these authors show, the employment size classes in the CBP can be deemed to constitute a fine grid, and the true mean (using national data) in each size class should be a close enough approximation.

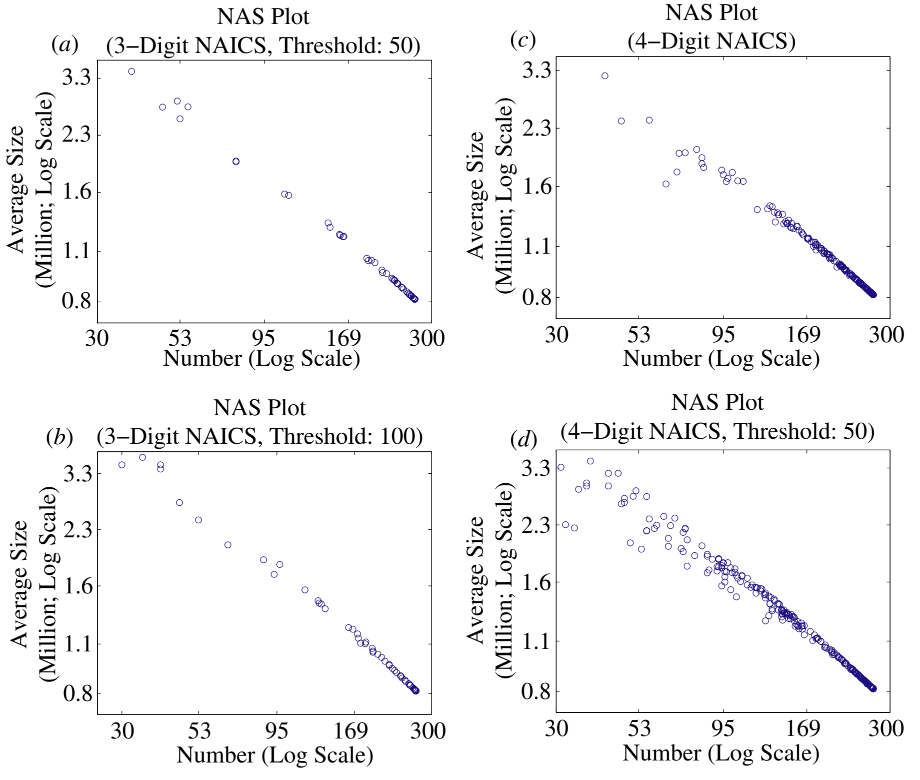

Fig. A2. *NAS Plots for Different Industry Definitions and Thresholds of Employment*
Source. 2000 County Business Pattern Data, Census Bureau.

Applying the Uniform Convergence Theorem for slowly varying functions,[20]

$$\lim_{y\downarrow 0}\frac{F(\theta y)-F(y)}{y^{\alpha}\ell(y)}=\lim_{y\downarrow 0}\int_{1}^{\theta}\frac{\ell(yu)}{\ell(y)}\frac{\mathrm{d}u}{u^{1-\alpha}}=\int_{1}^{\theta}\frac{\mathrm{d}u}{u^{1-\alpha}}=\frac{\theta^{\alpha}-1}{\alpha}.$$

Note that this limit is equal to $\ln(\theta)$ when $\alpha = 0$. For any positive θ,

$$\lim_{y\downarrow 0}\frac{F(y)-F(\theta y)}{F(y/\theta)-F(y)}=\lim_{y\downarrow 0}\frac{-\dfrac{[F(\theta y)-F(y)]}{y^{\alpha}\ell(y)}}{\dfrac{[F(y/\theta)-F(y)]}{y^{\alpha}\ell(y)}}=\theta^{\alpha}.$$

Setting $\theta = 1/4$, it can be seen that (4) holds for $\delta = (1/4)^{\alpha}$ for infinitely many k values such that y_k is in a sufficiently small neighbourhood of 0.

For the density part of (*b*), the proof is the same as that for case (*a*) and $[F(y) - F(\theta y)]/[F(y/\theta) - F(y)]$ has the same limit in θ^{α}. The only difference is that there are now only finite y_{k+1} values in a right neighbourhood of $\underline{y}$ that are close to this limit.

[20] Bingham *et al.* (1987), Theorem 1.2.1.

Proof of Proposition 4. Observe that

$$\frac{M_{i+1}}{M_i} = \frac{\sum_{k=1}^{i+1} 2^k \Delta_k}{\sum_{k=1}^{i} 2^k \Delta_k} = 1 + \frac{2^{i+1}}{\sum_{k=1}^{i} 2^k \left(\prod_{s=k}^{i} \Delta_s / \Delta_{s+1} \right)}.$$

Hence,

$$1 + \frac{2^{i+1}}{\left(\frac{\rho}{\delta}\right)^{i+1} \sum_{k=1}^{i} (2\delta/\rho)^k} \leq \frac{M_{i+1}}{M_i} \leq 1 + \frac{2^{i+1}}{(1/\rho\delta)^{i+1} \sum_{k=1}^{i} (2\rho\delta)^k}.$$

$$\Longleftrightarrow \frac{(2\delta/\rho)^{i+1} - 1}{(2\delta/\rho)^{i} - 1} \leq \frac{M_{i+1}}{M_i} \leq \frac{(2\rho\delta)^{i+1} - 1}{(2\rho\delta)^{i} - 1}.$$

Proof of Proposition 5. First, consider $\alpha = 0$, i.e. the exact Zipf's law case. Then, (7) can be rewritten as

$$\ln(AS_i) = \ln\left(\frac{Y_1}{2}\right) + \ln(i+1) - \ln(R_i).$$

Let $v = \ln(AS_i)$, and $s = \ln(R_i) = (i-1)\ln(2)$. Thus, the foregoing equation can be rewritten as $v = \ln(Y_1/2) - s + \ln[s/\ln(2) + 2]$, and

$$\frac{\mathrm{d}\ln(AS_i)}{\mathrm{d}\ln(R_i)} = \frac{\mathrm{d}v}{\mathrm{d}s} = -1 + \frac{1}{s + 2\ln(2)}, \tag{B.1}$$

which implies that (8) and (9) are true, as $s > 0$, and $s \to \infty$ as $i \to \infty$. For $\alpha > 0$, let $\beta = 2^{2\alpha}$. Then, (7) can be rewritten as

$$\ln(AS_i) = \ln(Y_1) + \ln\left[1 + \frac{\beta^{2-i} - \beta}{2(1-\beta)}\right] - \ln(R_i).$$

It is easily verified that

$$\frac{\mathrm{d}\ln(AS_i)}{\mathrm{d}\ln(R_i)} = \frac{\mathrm{d}v}{\mathrm{d}s} = -1 + \frac{\beta^{1-[s/\ln(2)]}}{3\beta - 2 - \beta^{1-[s/\ln(2)]}} \frac{\ln(\beta)}{\ln(2)}, \tag{B.2}$$

which immediately implies that (9) is true. This $\mathrm{d}\ln(AS_i)/\mathrm{d}\ln(R_i) > -1$ follows from the fact that $\beta > 1$ and $s \geq 0$. To show that $\mathrm{d}\ln(AS_i)/\mathrm{d}\ln(R_i) < 0$, we need to show that the second term on the right-hand side in (B.2) is <1, which would be true if

$$\frac{\beta^{1-[s/\ln(2)]}}{3\beta - 2 - \beta^{1-[s/\ln(2)]}} < \frac{1}{2\alpha}.$$

Observe that

$$\frac{\beta^{1-[s/\ln(2)]}}{3\beta - 2 - \beta^{1-[s/\ln(2)]}} \leq \frac{\beta}{2(\beta-1)} = \frac{2^{2\alpha-1}}{2^{2\alpha} - 1} < \frac{1}{2\alpha}.$$

The last inequality follows from the fact that $\alpha \in (0, 1/2)$. That $\mathrm{d}\ln(AS_i)/\mathrm{d}\ln(R_i)$ is strictly decreasing in i follows directly from (B.1) and (B.2).

Now, consider $\alpha \in (-1/2, 0)$. Observe that

$$\lim_{s\to\infty} \frac{\beta^{1-[s/\ln(2)]}}{3\beta - 2 - \beta^{1-[s/\ln(2)]}} = -1.$$

Hence, (B.2) gives (10).

Proof of Proposition 6. Using Points 1, 2 and 4, the argument for each good being in an SPNE is the same as Proposition 1 in the $r \leq c$ case. The additional profits from sales to workers do not alter the fact that firms would prefer to be located in the middle of their neighbours. The rest of the proof shows that there is a solution that satisfies Points 1 to 4. Observing (17)–(19), $\{Y_i, N_i, L_i\}_{i=1}^{I}$ are obviously decreasing sequences. Combining (17)–(19),

$$y_i = \frac{tL_i^2}{2} + \frac{\phi(r-c)}{1-\phi F(y_i)} \sum_{k=i}^{I} [F(y_k) - F(y_{k+1})] L_k. \tag{B.3}$$

Looking at (B.3) recursively back from the layer-I cities,

$$y_I = \frac{tL_I^2}{2} + \frac{\phi(r-c)}{1-\phi F(y_I)} F(y_I) L_I. \tag{B.4}$$

Given L_1 and I, note that $F(0) = 0$ and $F(\infty) = 1$ and define

$$\Gamma_I(y; L_1) = \frac{tL_I^2}{2} + \frac{\phi(r-c)}{1-\phi F(y)} F(y) L_I - y.$$

Hence, $\Gamma_I(0) > 0$ and $\Gamma_I(\infty) < 0$. Because $F(\cdot)$ is left continuous, a solution to $\Gamma_I(y) = 0$ exists. Denote this solution as y_I^*. Obtain such a solution for each I, and find the I that ensures that $y_{I+1}^* < \underline{y} \leq y_I^*$. Then, redefine $y_{I+1}^* = \underline{y}$.

Define similar mappings of $\Gamma_i(y)$ recursively using (B.3), given $y_{i+1}^*, y_{i+2}^*, \ldots, y_I^*$. Solving $\Gamma_i(y) = 0$ given $y_{i+1}^*, y_{i+2}^*, \ldots, y_I^*$ entails a correspondence of vectors dependent on L_1, $\mathbf{y}(L_1) = \{y_i^*(L_1)\}_{i=1}^{I}$ with $y_i^*(L_1) > y_{i+1}^*(L_1)$ for all i values. $\mathbf{y}(L_1)$ is strictly increasing in L_1. Equation (16) can be rewritten as

$$y_1^*(L_1) - \bar{y} \geq 0, y_2^*(L_1) - \bar{y} < 0. \tag{B.5}$$

Indeed, because $F(\cdot)$ is continuous, $\Psi(L_1) \equiv y_1^*(L_1) - \bar{y}$ has a closed graph, $\Psi(0) < 0$, and $\Psi(\infty) > 0$. Thus, there exists an L_1^* such that $\Psi(L_1^*) = \epsilon' \geq 0$, and ϵ' can be arbitrarily small. That is, there exists an L_1^* such that Point 1 (or, equivalently, (B.5)) holds.

Proof of Proposition 7. I first prove the Proposition for the Δ-deviation of an arbitrary good y, and then show how it generalises to a sufficiently small subset of goods.

Consider a Δ-deviation of an arbitrary good y with Δ small enough such that $c + t\Delta < r$. For any type-y firm, the two closest cities must be a layer-I city and a layer-i one for some $i < I$. The condition (14) and the fact that $c + t\Delta < r$ imply that $\Delta < L_I$. Hence, the distances of any type-y firm from the two closest cities must be Δ and $L_I - \Delta$. Any type-y firm must sell to the city that is Δ distance from it. It may also sell to the city that is $L_I - \Delta$ distance from it, if L_I is sufficiently small.

Consider $y \in (y_{k+1}, y_k]$ with $k < I$. Then, every type-y firm's closest non-layer-I city must be a layer-i city with $i \leq k$, and the distance between each type-y firm and the layer-i city is Δ. Given the locations of other firms, consider the profit that a type-y firm can earn by unilaterally moving $\delta \in [0, (r-c)/t)$ closer to its nearest layer-i city (see Figure 6). Denote this profit by $\hat{\pi}_{k,i}(\delta)$. In the case of a type-y firm selling to two cities,

$$\hat{\pi}_{k,i}(\delta) = \frac{t}{2}\left(L_k^2 - \delta^2\right) + \{r - [c + t(\Delta - \delta)]\} N_i + \{r - [c + t(L_I - \Delta + \delta)]\} N_I - y. \tag{B.6}$$

The first-order derivative is

$$\frac{\mathrm{d}\hat{\pi}_{k,i}(\delta)}{\mathrm{d}\delta} = t[(N_i - N_I) - \delta]. \tag{B.7}$$

Thus, given other firms' locations, $\mathrm{d}\hat{\pi}_{k,i}(0)/\mathrm{d}\delta = t(N_i - N_I) > 0$ implies that the optimal δ must be positive and the type-y good is obviously not in equilibrium. In the case of a type-y firm selling only to one city, we simply remove the third term on the right-hand side of (B.6) and have $\mathrm{d}\hat{\pi}_{k,i}(0)/\mathrm{d}\delta = tN_i > 0$.

For $y \in [\underline{y}, y_I]$, any type-$y$ firm sells only to its nearest city, unless $\Delta = L_I/2$. This is because, when $\Delta \neq L_I/2$, no two type-y firms that are closest to any given city are of equal distance to that city and hence the workers in that city only buy y from the firm that is closer. Employing similar arguments, when $\Delta \neq L_I/2$, each type-y firm has an incentive to move toward its nearest city, given other firms' locations. For the case of $\Delta = L_I/2$, assume that the sales to the workers in any city are evenly split between the two nearest firms. Then, given other firms' locations, each type-y firm has an incentive to move towards the larger of the two nearest cities, because it can thus grab the entire market of y in that city.

The above result trivially generalises to a Δ-deviation of a measure-zero subset. Now, when a Δ-deviation involves a positive measure of goods, a deviating type-y firm may sell to some deviating workers, in addition to the farmers and those city workers who stay put. Some of these deviating workers may form 'small cities' if the deviating firms line up at the same locations, some of them may spread out continuously on a line segment and some may be sporadically located at unconnected different points. The sales to the third group can be ignored and the sales to the second group of workers are similar to those to farmers. Without loss of generality, suppose the deviation involving y is to the right. For an arbitrary deviating type-y firm, denote $\tilde{N}^{\ell}$ ($\tilde{N}^{r}$) as the total mass of the above-mentioned first group of workers (possibly from different locations) to the left (right) of it that it sells to. To see why the deviating subset needs to be sufficiently small, take the case of a type-y firm selling to two cities and $y \in (y_{k+1}, y_k]$ with $k < I$. In this case, (B.7) evaluated at 0 becomes

$$\frac{\mathrm{d}\hat{\pi}_{k,i}(0)}{\mathrm{d}\delta} = t(N_i - N_I + \tilde{N}^{\ell} - \tilde{N}^{r}).$$

The second group of workers does not factor in this formula because their role is similar to farmers and is nullified by taking δ to 0. As $N_i - N_I + \tilde{N}^{\ell} > 0$, as long as $\tilde{N}^{r}$ is small enough, $\mathrm{d}\hat{\pi}_{k,i}(\delta)/\mathrm{d}\delta > 0$ and thus the type-y good is not in equilibrium. Of course, the total mass of deviating workers $\tilde{N}^{\ell} + \tilde{N}^{r}$ is limited by the size of the subset involved in the deviation. This argument trivially applies to all the other cases.

Appendix C. Examples of Regularly Varying Distribution Functions

Table C1 lists distributions that are regularly varying at the left tail. If the domain of a distribution listed is unbounded from above, then it is understood that some proper truncation of the right-tail is needed (due to the existence of the upper bound of fixed cost $\bar{y}$). For clarity, the functions shown are proportional to the distribution and density functions, hence ignoring the adjustment of the parameters due to normalisation and/or truncation.

Appendix D. Algorithm Used to Compute the Equilibrium Size Distributions in Section 3

Recall that the number of layers must be limited to ensure that they are finite in this case. First, make a grid of L_1, running from the top down. Given that each L_1 and $L_I = L_1/2^{I-1}$, I can solve y_I by (B.4) for each $I \in \mathbb{N}$. The larger the I, the smaller the y_I. The $y_I(L_1)$ and $I(L_1)$ are those that satisfy $y_I \geq \underline{y}$ and $y_{I+1} < \underline{y}$. After $I(L_1)$ and $y_I(L_1)$ are pinned down, solve y_i recursively using (B.3) and backwardly from y_{I-1} to y_1. Thus, a sequence $\{y_i(L_1)\}_{i=1}^{I}$ is obtained. Find the L_1 such that

Table C1

*Examples of Regularly Varying Distribution**

Name	$F(y)\sim$	$f(y)\sim$	Index
Beta		$y^{\alpha-1}(1-y)^{\beta-1}$	$\alpha > 0$
Gamma		$y^{\alpha-1}e^{-y/\beta}$	$\alpha > 0$
F		$y^{\alpha-1}/\left(1+\frac{2\alpha y}{\beta}\right)^{(2\alpha+\beta)/2}$	$\alpha > 0$
Weibull	$1-e^{-\left(\frac{y}{\lambda}\right)^{\alpha}}$	$\frac{\alpha}{\lambda}\left(\frac{y}{\lambda}\right)^{\alpha-1}e^{-\left(\frac{y}{\lambda}\right)^{\alpha}}$	$\alpha > 0$
Kumaraswamy	$1-(1-y^{\alpha})^{\beta}$	$\alpha\beta y^{\alpha-1}(1-y^{\alpha})^{\beta-1}$	$\alpha > 0$
Inv. of Pareto	βy^{α}	$\beta\alpha y^{\alpha-1}$	$\alpha > 0$
Log-Logistic	$y^{\alpha}/(\beta^{\alpha}+y^{\alpha})$	$(\alpha/\beta)*(y/\beta)^{\alpha-1}/[1+(y/\beta)^{\alpha}]^{2}$	$\alpha > 0$
U-quadratic	$\frac{4}{\bar{y}^3}\left[\left(y-\frac{\bar{y}}{2}\right)^3+\left(\frac{\bar{y}}{2}\right)^3\right]$	$\frac{12}{\bar{y}^3}\left(y-\frac{\bar{y}}{2}\right)^2$	1
Rayleigh	$1-e^{-\frac{y^2}{2\sigma^2}}$	$\frac{y}{\sigma^2}e^{-y^2/2\sigma^2}$	1
Generalised Pareto	$1-(1-\beta y/\sigma)^{1/\beta}$	$\frac{1}{\sigma}\left(1-\frac{\beta y}{\sigma}\right)^{(1-\beta)/\beta}$	1
	$y^{\alpha}/[1-\ln(y)]$	$\frac{y^{\alpha-1}[1+\alpha-\alpha\ln(y)]}{[1-\ln(y)]^2}$	$\alpha \in \mathbb{R}$
	$y^{\alpha}e^{1-\sqrt{1-\ln(y)}}$	$[1+2\alpha\sqrt{1-\ln(y)}]y^{\alpha-1}e^{1-\sqrt{1-\ln(y)}}/2\sqrt{1-\ln(y)}$	$\alpha \in \mathbb{R}$
	G is Log-Logistic ($\alpha < 0$)		0
Pareto	$G(y)=\beta y^{\alpha}$		$\alpha < 0$
	$G(y)=[\ln(1+\beta y)]^{\alpha}$		$\alpha < 0$
	G is Kumaraswamy ($\alpha < 0$)		$\alpha\beta < 0$
	$G(y)=y^{\alpha}e^{-y/\beta}$		$\alpha < 0$
		Beta density with $\alpha < 0$	$\alpha < 0$
		Gamma density with $\alpha < 0$	$\alpha < 0$
		F density with $\alpha < 0$	$\alpha < 0$

*The second and third columns show the distribution and density functions respectively. The fourth column indicates the index with which the distribution is regularly varying. For those with an index < 0, it is understood that $y > 0$, and the G function by which the distribution function F is defined via (6) is listed. The parameters of the distributions are all positive unless otherwise noted.

$y_1(L_1) \geq \bar{y}$ and $y_2(L_1) < \bar{y}$. Redefine $y_1 = \bar{y}$. Finally, given L_1 and y_i, $\{Y_i, N_i, L_i\}_{i=1}^{I}$ are calculated according to Proposition 6.

National University of Singapore

Submitted: 6 January 2011
Accepted: 12 November 2011

References

Axtell, R.L. (2001). 'Zipf distribution of US firm sizes', *Science*, vol. 293, pp. 1818–20.

Bak, P., Tang, C. and Wiesenfeld, K. (1987). 'Self-organised criticality: an explanation of $1/f$ noise', *Physical Review Letters*, vol. 59(4), pp. 381–4.

Barabási A.-L. and Albert, R. (1999). 'Emergence of scaling in random networks', *Science*, vol. 286, pp. 509–12.
Beckmann, M.J. (1958). 'City hierarchies and the distribution of city size', *Economic Development and Cultural Change*, vol. 6(3), pp. 243–8.
Berliant, M. (2006). 'Well isn't that spatial?! Handbook of regional and urban economics: a view from Economic Geography', book review, *Journal of Economic Geography*, vol. 6, pp. 107–10.
Berry, B.J.L. (1967). *Geography of Market Centers and Retail Distribution*, Englewood Cliffs, NJ: Prentice-Hall.
Berry, B.J.L. and Garrison, W.L. (1958*a*). 'The functional base of the central place hierarchy', *Economic Geography*, vol. 34(2), pp. 145–54.
Berry, B.J.L. and Garrison, W.L. (1958*b*). 'A note on central place theory and the range of a good', *Economic Geography*, vol. 34(4), pp. 304–11.
Bingham, N.H., Goldie, C.M. and Omey, E. (2006). 'Regularly varying probability densities', *Publications De L'Institut Mathématique,* Nouvelle série, tome. 80(94), pp. 47–57.
Bingham, N.H., Goldie, C.M. and Teugels, J.L. (1987). *Regular Variation*, Cambridge, UK: Cambridge University Press.
Bingham, N.H. and Teugels, J.L. (1975). 'Duality for regularly varying functions', *Quarterly Journal of Mathematics*, vol. 26(1), pp. 333–53.
Brakman, S., Garretsen, H., Van Marrewijk, C. and van den Berg, M. (1999). 'The return of Zipf: a further understanding of the rank-size distribution', *Journal of Regional Science*, vol. 39, pp. 183–213.
Christaller, W. (1933). *Central Places in Southern Germany*, translated by C. W. Baskin (1966), Englewood Cliffs, NJ: Prentice-Hall.
Cronon, W. (1991). *Nature's Metropolis: Chicago and the Great West*, New York, NY: Norton.
Duranton, G. (2006). 'Some foundations for Zipf's law: Product proliferation and local spillovers', *Regional Science and Urban Economics*, vol. 36(4), pp. 543–63.
Duranton, G. (2007). 'Urban evolutions: the fast, the slow, and the still', *American Economic Review*, vol. 97(1), pp. 197–221.
Eaton, J. and Eckstein, Z. (1997). 'Cities and growth: theory and evidence from France and Japan', *Regional Science and Urban Economics*, vol. 27, pp. 443–74.
Eaton, B.C. and Lipsey, R.G. (1982). 'An economic theory of central places', Economic Journal, vol. 92(365), pp. 56–72.
Eeckhout, J. (2004). 'Gibrat's law for (all) cities', *American Economic Review*, vol. 94(5), pp. 1429–51.
Fujita, M. and Krugman, P. (1995). 'When is the economy monocentric? von Thünen and Chamberline unified', *Regional Science and Urban Economics*, vol. 25, pp. 505–28.
Fujita, M., Krugman, P. and Mori, T. (1999*a*). 'On the evolution of hierarchical urban systems', *European Economic Review*, vol. 43, pp. 209–51.
Fujita, M., Krugman, P. and Venables, A.J. (1999*b*). *The Spatial Economy: Cities, Regions, and International Trade*, Cambridge, MA: The MIT Press.
Gabaix, X. (1999) 'Zipf's law for cities: an explanation', *Quarterly Journal of Economics*, vol. 114(3), pp. 739–67
Gabaix, X. (2009). 'Power laws in economics and finance', *Annual Review of Economics*, vol. 1, pp. 255–93.
Gabaix, X. and Ioannides, Y.M. (2004). 'The evolution of city size distributions', in (J.V. Henderson and J.-F. Thisse, eds.), *Handbook of Regional and Urban Economics.* Edition 1, vol. 4, Chapter 53. pp. 2341–78, Amsterdam: Elsevier.
Gabaix, X. and Landier, A. (2008). 'Why has CEO pay increased so much?', *Quarterly Journal of Economics*, vol. 123(1), pp. 49–100.
Henderson, J.V. (1974). 'The types and size of cities', *American Economic Review*, vol. 64, pp. 640–56.
Holmes, T.J. and Stevens, J.J. (2004). 'Spatial Distribution of Economic Activities in North America', in (J.V. Henderson and J.-F. Thisse, eds.), *Handbook of Regional and Urban Economics*, edition 1, vol. 4, ch. 63. pp. 2797–843, Amsterdam: Elsevier.
Hsu, W.-T. and Holmes, T.J. (2009). 'Optimal city hierarchy: A dynamic programming approach to central place theory', Working Paper, National University of Singapore.
Jessen, A.H. and Mikosch, T. (2006).'Regularly varying functions', *Publications De L'Institut Mathématique,* Nouvelle série, tome. vol. 80(94), pp. 171–92.
Krugman, P. (1996). 'Confronting the mystery of urban hierarchy', *Journal of the Japanese and the International Economies*, vol. 10, pp. 399–418.
Lederer, P.J. and Hurter, A.P. Jr. (1986). 'Competition of firms: Discriminatory pricing and location', *Econometrica*, vol. 54(3), pp. 623–40.
Lösch, A. (1940). *The Economics of Location*, translated by W. Woglom (1956), New Haven, CT: Yale University Press.
Luttmer, E.G.J. (2007). 'Selection, growth, and the size distribution of firms', *Quarterly Journal of Economics*, vol. 122(3), pp. 1103–44.
Mori, T., Nishikimi, K. and Smith, T.E. (2008). 'The number-average size rule: a new empirical relationship between industrial location and city size', *Journal of Regional Science*, vol. 48, pp. 165–211.
Mori, T. and Smith, T.E. (2011). 'An industrial agglomeration approach to central place and city size regularities', *Journal of Regional Science*, vol. 51(4), pp. 694–731.

Quinzii, M. and Thisse, J.-F. (1990). 'On the optimality of central places', *Econometrica*, vol. 58(5), pp. 1101–19.
Rossi-Hansberg, E. and Wright, M.L.J. (2007). 'Urban structure and growth', *Review of Economic Studies*, vol. 74(2), pp. 597–624.
Rozenfeld, H., Rybski, D., Gabaix, X. and Makse, H. (2011). 'The area and population of cities: new insights from a different perspective on cities', *American Economic Review*, vol. 101(5), pp. 2205–25.
Simon, H.A. and Bonini, C.P. (1958). 'The size distribution of business firms', *American Economic Review*, vol. 48(4), pp. 607–17.
Soo, K.T. (2005). 'Zipf's law for cities: a cross country investigation', *Regional Science and Urban Economics*, vol. 35, pp. 239–63.
Tabuchi T. and Thisse, J.-F. (2011). 'A new economic geography model of central places', *Journal of Urban Economics*, vol. 69, pp. 240–52.
Zipf, G.K. (1949). *Human Behavior and the Principle of Least Effort*, Cambridge, MA: Addison-Wesley.