3-2013

Optimal City Hierarchy: A Dynamic Programming Approach to Central Place Theory

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

Thomas J. HOLMES
University of Minnesota - Twin Cities

Frank MORGAN
Williams College

Citation

Optimal City Hierarchy: A Dynamic Programming Approach to Central Place Theory

Wen-Tai Hsu[*]   Thomas J. Holmes[†]   Frank Morgan[‡§]

July 26, 2012

Abstract

Central place theory is a key building block of economic geography and an empirically plausible description of city systems. This paper provides a rationale for central place theory via a dynamic programming formulation of the social planner's problem of city hierarchy. We show that there must be one and only one immediate smaller city between two neighboring larger-sized cities in any optimal solution. If the fixed cost of setting up a city is a power function, then the immediate smaller city will be located in the middle, confirming the locational pattern suggested by Christaller (1933). Moreover, the optimal city hierarchy can be decentralized. We also show that the solution can be approximated by iterating the mapping defined by the dynamic programming problem. The main characterization results apply to a general hierarchical problem with recursive divisions.

JEL: R12; R13

Keywords: central place theory, city hierarchy, dynamic programming, principle of optimality, fixed point

1 Introduction

Central place theory describes how a city hierarchy is formed from a featureless plain of farmers as consumers. It is a key building block of economic geography (King, 1984) and dates back at least to Christaller (1933). Many have argued for its empirical plausibility as a description of city hierarchy (Fujita, Krugman, and Venables, 1999; Berliant, 2008; Mori, Nishikimi, and Smith, 2008; Mori and Smith, 2011). Although original central place theory is not a rigorous economic theory based on incentives and equilibrium, many economists have found its insights appealing, and a few attempts have been made to formalize it, including those by Eaton and Lipsey (1982), Quinzii and Thisse (1990), Fujita, Krugman, and Mori (1999), Tabuchi and Thisse (2011), and Hsu (2012).[1]

The basic idea behind this theory is that goods differ in their degree of scale economies relative to market size. Goods for which this ratio is large, e.g., stock exchanges or symphony orchestras, will be found in only a few places, whereas goods for which the ratio is small, e.g., gas stations or convenience stores, will be found in many places. Moreover, large cities tend to have a wide range of goods, whereas small cities provide only goods with low scale economies. Naturally, small cities are in the market areas of large cities for those goods that they themselves do not provide. In Christaller's scheme, the *hierarchy property*[2] holds if larger cities provide all of the goods that smaller cities also provide and more.

In this paper, a *city system* is composed of multiple layers of cities, and cities of the same layer have the same functions, i.e., they host the same set of industries. The driving force behind the differentiation of cities is the heterogeneity of scale economies among goods, which is modeled by heterogeneity in the setup costs of production. In addition to the hierarchy property, another defining feature of city hierarchy in central place theory, that called the *central place property*, is that there is only one next-layer city between neighboring larger cities and it is halfway in between. Christaller (1933) calls this the K = 3 *market principle*.[3] The city hierarchy described by central place theory (hereafter *central place hierarchy*) is a city system in which both the hierarchy and central place properties hold. Figure 1 provides an illustration of such a city hierarchy in a one-dimensional geographic space.[4]

---

[1] Besides central place theory, another important theory of city hierarchy is Henderson's (1974) type-of-cities theory, which emphasizes cities' roles in industrial specialization. Also see the extension to city growth by Rossi-Hansberg and Wright (2007).

[2] This is often called the *hierarchy principle* in the literature.

[3] On the plane, if there is always only one next-layer city located at the centroid of the equilateral triangle area in between three neighboring larger cities, then the ratio of the market areas is 3.

[4] The vertical axis shows the range of goods produced and goods are indexed by some measure of the
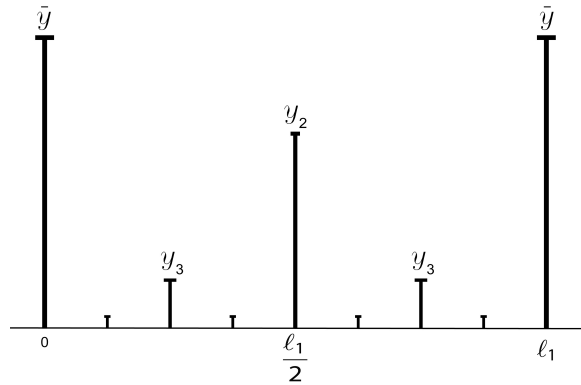
Figure 1: Central Place Hierarchy on the Line.

This paper takes aim at providing a rationale for central place theory via a social planner's problem. An innovative feature of this paper is that the social planner's problem is formulated as a dynamic programming problem in a geographic space (instead of in time). In this paper, we ask what optimal city hierarchy would arise from a uniformly populated space via the tradeoff between transport costs and the setup costs of production (and hence, the setup costs of cities). Lucas and Rossi-Hansberg (2002) pioneer the application of dynamic programming to a spatial problem.[5] Whereas they study city residents' choice of work and residence locations within a city, we apply the technique to a spatial problem of city locations. Earlier literature in regional science also addresses central place theory using a recursive structure, e.g., Beckmann (1958, 1970), but does not make use of dynamic programming.

Quinzii and Thisse (1990) also ask how a central place hierarchy might emerge from a socially optimal solution, and while they provide conditions under which the hierarchy property emerges in the optimal solution, their optimal solution does not feature the central place property. In contrast, this paper takes as given the hierarchy property and asks instead whether the spacing in central place theory is optimal. In addition to Quinzii and Thisse (1990), the other above-mentioned attempts at modeling central place theory, with the exception of Hsu (2012), have mostly ignored the locational issue, i.e., focused on the hierarchy property. Thus, this paper complements the literature by squarely confronting the locational issue. As we will clarify, the hierarchical location choice problem is more

degree of scale economies, e.g., fixed cost of production, $y \in [0, \bar{y}]$, for some $\bar{y} > 0$. The hierarchy property implies that each city provides goods in $[0, y]$ for some y. Hence, a layer-i city provides goods in $[0, y_i]$, and obviously, $y_1 = \bar{y}$.

[5]Broadly speaking, spatial problems, like time problems, are closely associated with recursivity and often have to make use of some functional approaches. For example, see Mirrlees (1972), who probes the properties of an optimal population density function in a monocentric city.

complex than one would have imagined or hoped, even with the assumption of uniform distribution.

In this model there are a continuum of goods that varies in setup cost of production and a continuum of locations containing individuals who need to consume all the different goods. This underlying modeling structure is the same as that used in Hsu (2012). In both papers, the central issue is what structure of cities will emerge in this economy, but the formulations of the problem are very different. Hsu (2012) focuses on the market equilibrium outcome. This paper instead solves the social planner's problem. The social planner's problem is more technically challenging than the market equilibrium outcome, requiring us to approach the problem in a different way. In particular, to verify equilibrium in Hsu (2012), it was sufficient to examine local conditions in which marginal firms had zero profit. Here, we need to tackle the global optimality of the planner's solution, and this leads to the development of our dynamic programming approach, an approach that may have additional applications, as we discuss below. In addition to the difference in formulations, the two papers focus on different questions. Here, the main focus is on whether or not the central place property is satified. In Hsu (2012), the central place property is a simple result, and his focus is whether equilibrium city size distribution follows the power law. We further elaborate on the connections between these papers in Section 4.

Our main results are as follows. First, we show that under rather weak conditions on the structure of setup costs, there will always be one and only one immediate sub-city, i.e., the largest among all cities in between two neighboring larger-sized cities. The intuition for why there is at least one city is straightforward, and we provide it here. We assume that the setup cost goes to zero as city size goes to zero. Thus, between any pair of cities it is always worthwhile to place another city, perhaps one that is very small, to save on the transport costs generated by consumers buying low setup cost goods. The intuition of why there is only one intermediate sub-city is more complicated, and we defer this to later.

Second, based on the first result, we formulate a sequence problem and the corresponding dynamic programming problem while providing characterization for both problems. To find an optimal hierarchy, the social planner's problem can be formulated as looking for a sequence of the locations and sizes of immediate sub-cities to minimize the per capita cost. When the size and location of an immediate sub-city are chosen, the location divides the area bounded by the two neighboring larger-sized cities into two areas, each of which is a new area in which a new sub-city is to be determined. Thus, the sequence problem form of the social planner's problem involves an infinite bifurcation of

areas. Such recursivity naturally allows a dynamic programming formulation. We show that the two problems are equivalent (the *principle of optimality*). More importantly, for the mapping defined by the dynamic programming problem, we show that there exists a unique fixed point and it equals the minimized cost function of the sequence problem, and that the fixed point can be approximated by iterations of the mapping, even though the mapping is not necessarily a contraction as there is no discount factor. Our numerical examples show that the iterations converge to the solution at a rather fast rate.

We find that all of these characterization results apply to a general problem of recursive divisions and are hence potentially useful in various hierarchical problems. For example, there is a large literature on firm hierarchy, e.g., Qian (1994), Garicano (2000), and Garicano and Rossi-Hansberg (2006), and the techniques developed in this paper can potentially contribute to this literature. Other possible venues of application include outsourcing in trade and the structure of fiscal decentralization. We present the results on the general problem first and then show how these results apply to the city hierarchy problem.

Third, we find an interesting case in which we can find the unique fixed point analytically. We show that when the setup cost function is a power, the central place property holds, i.e., the optimal location of any immediate sub-city is exactly in the middle between two neighboring larger-sized cities. This functional form is of particular interest because Hsu (2012) showed that under this condition, the resulting equilibrium size distribution of cities follows a power law, a well-known empirical regularity.[6]

Fourth, we determine when the social planner's solution coincides with the market equilibrium outcome in Hsu (2012). In particular, we show that the optimal solution can be achieved through the equilibrium outcome; that is, the optimal solution can be decentralized. However, other suboptimal equilibria also exist.

The rest of this paper is organized as follows. Section 2 introduces the environment, defines the social planner's problem, and derives two key lemmas that simplify the problem. Section 3 formulates both the sequence and dynamic programming problems and provides characterization results for both the city hierarchy problem and a more general problem. It also shows that the central place property holds under the power law distrib-

---

[6]Deviations from the power law may be found when smaller cities or towns are included (Eeckhout 2004), but the power law remains a good approximation at least for the right tail (Eeckhout 2004, Rozenfeld, Rybski, Gabaix, and Makse 2011). The model in this paper is also consistent with such deviations. See further discussion in Section 3.4. Unlike other theories of urban systems and city size distribution, Hsu's explanation of city size distribution is based on what cities do differently and how things occur geographically, rather than on a random growth process of cities. For explanations along this line, see Simon (1955), Gabaix (1999), Eeckhout (2004), Duranton (2006, 2007), Rossi-Hansberg and Wright (2007), and Córdoba (2008).

ution of setup costs. Section 4 compares the optimal allocation with Hsu's (2012) equilib-
rium. This section also briefly discusses the problem in two-dimensional space. Section
5 gives our conclusions. Several proofs are relegated to a separate appendix which is
available online on the authors' websites.

2 Model and Immediate Sub-cities

2.1 Environment

The geographic space is the real line $\mathbb{R}$ on which an infinite mass of consumers is uni-
formly distributed with a density of one. There is a continuum of commodities labeled
$y \in [0, z_1]$, where z_1 is exogenously given. Each consumer demands one unit of each
$y \in [0, z_1]$. To produce any good y, a setup cost $\phi(y)$ is required. This setup cost includes
the setup/fixed cost of producing y and any potential city-wide (fixed) external cost of
producing y. The marginal cost is a constant γ. To transport any good requires a cost of
t per unit of distance. The goods are ranked in terms of their setup costs, and we assume
that no two goods have the same setup cost. Hence, ϕ is strictly increasing. We also as-
sume that ϕ is continuous with $\phi(0) = 0$. In addition, we assume the *hierarchy property*: at
any location, if a good $z \leq z_1$ is produced, then all $y \in [0, z]$ are also produced.

We interpret production locations as cities. Without the hierarchy property, the opti-
mal distance between two production locations can actually be solved good by good, and
industries would not have to co-agglomerate at an optimal solution. Obviously, there are
benefits to industries co-agglomerating, but we make those benefits implicit by assum-
ing the hierarchy property so as to focus on the hierarchical location choice problem. To
provide microfoundations for the hierarchy property per se is a rather challenging and
worthwhile research agenda. For such an effort, see Eaton and Lipsey (1982), Quinzii and
Thisse (1990), Fujita, Krugman, and Mori (1999), Tabuchi and Thisse (2006, 2011), and
Hsu (2012).

We assume a uniform distribution of consumers for tractability. However, one can also
think of them as farmers who would locate themselves uniformly if agricultural produc-
tivity, or other instances of the on-site extraction of natural resources, were uniform and
if the farming technology were Leontief in land and labor. The original development of
central place theory per Christaller was in fact an attempt to explain the "industrial activ-
ities" that serve the farmers (on the farming plains of southern Germany). Nonetheless,
it is important to note that it does not actually matter whether agricultural employment
is large or small; as long as there are immobile consumers spreading across the entire

geographic space, there exists the need for cities and towns to spread out to serve these consumers.

2.2 The problem

We label a location that produces all goods up to z as a z-city. Denote the cost of setting up a z-city as $\Phi(z) \equiv \int_0^z \phi(y)\, dy$. According to the hierarchy property, z also refers to a city's size. The social planner's objective is to decide the locations and sizes of cities to minimize the per capita cost of production to serve every consumer a unit of each good in $[0, z_1]$.

A z_1-city serves the entire possible range of goods. Since all goods must be consumed, including z_1 with the highest setup costs, z_1-cities must exist in any social planning solution. The first question then facing the social planner is how far apart z_1-cities should be spaced on the real line. We denote the distance between two z_1-cities as ℓ_1 and call all of the cities in between two neighboring z_1-cities a *city hierarchy*, which for now includes the possibility of no smaller cities. Without loss of generality, let the area/interval between a particular pair of *neighboring* z_1-cities be $[0, \ell_1]$, i.e., the two cities are located at 0 and ℓ_1, respectively. Because consumers must be served by the nearest production locations, no consumers on $[0, \ell_1]$ will be served by the cities outside $[0, \ell_1]$. Thus, to search for a solution, the social planner looks for an ℓ_1 and a city hierarchy on $(0, \ell_1)$ without information on the city hierarchies outside this interval. However, if there is an optimal ℓ_1 and an optimal city hierarchy on $(0, \ell_1)$, this optimal city hierarchy can be duplicated on $(k\ell_1, (k+1)\ell_1)$, $k \in \mathbb{Z}$ with two neighboring intervals sharing a common z_1-city at the border. Hence, it is always optimal to evenly space z_1-cities, although uneven spacing of z_1-cities may also be optimal if there are multiple solutions of ℓ_1. In sum, the social planner's problem involves two stages. In the first stage, the social planner decides ℓ_1. In the second stage, the social planner determines the city hierarchy given ℓ_1. The focus of our analysis is the second stage, which spans Sections 2 and 3. The optimal choice of ℓ_1 is analyzed in Section 3.5.

Given ℓ_1, let the discrete set of cities on $(0, \ell_1)$ be denoted as

$$W \equiv \left\{ \begin{array}{c} (z_i, L_{z_i}, I)\, |z_i \in (0, z_1],\ i = 1, 2, ...I,\ I \in \mathbb{N} \cup \{\infty\},\ z_i > z_{i+1}, \\ L_{z_i} \text{ is the set of locations of } z_i\text{-cities} \end{array} \right\}.$$

That is, z_i is the i-th largest among all cities on $(0, \ell_1)$. For now, there may be multiple z_i cities, and L_{z_i} and $|L_{z_i}|$ denote the set of locations and the number of z_i-cities on $(0, \ell_1)$, respectively. The number I is the number of *layers* of cities, and I can be (countably)

infinite.

The optimization problem, given ℓ_1, is to search for a city hierarchy W that solves

$$C^* (\ell_1, z_1) \equiv \inf_W \frac{1}{\ell_1} \left[\sum_{z_i} |L_{z_i}| \Phi (z_i) + \text{total transport cost} \right], \tag{1}$$

Three points are worth noting. First, we ignore the per capita variable cost because it is always γz_1, regardless of the allocation W. Second, "total transport cost" is calculated as follows. Since each city hierarchy W defines *a partition of market areas on* $(0, \ell_1)$ for each good $y \in [0, z_1]$,[7] the transport cost for each y is thus the sum of transport costs incurred in each market area. We obtain the "total transport cost" by integrating over y. The concept is clear, but the expression is messy and not helpful for the following analysis. Hence, we do not include this notational burden. Third, we write $\inf$ instead of $\min$ in (1) because the existence of a minimizer is not yet proven. An infimum obviously exists because the objective is bounded between zero and the cost of building no city hierarchy on the interval of length ℓ_1.

2.3 Two key lemmas

The following two lemmas provide key characteristics of an optimal hierarchy that enables us to set up the planner's problem as a dynamic programming problem.

Lemma 1. *It is never optimal to have an interval without any city in it.*

Proof. Consider an interval $[0, \ell]$ such that there are no cities in $(0, \ell)$ in between two cities located at 0 and ℓ, respectively. Let z denote the size of the smaller of the two cities at the end points. Now, consider adding a z'-city in the middle in between with $z' \leq z$. Then, the savings in transport cost per good is $2 \int_0^{\ell/2} txdx - 4 \int_0^{\ell/4} txdx = t\ell^2/8$. Accounting for the increase in setup cost, the net saving from having a z'-city is given by $S (z'; \ell) \equiv \int_0^{z'} \left[\frac{t\ell^2}{8} - \phi(y) \right] dy$. Because ϕ is continuous and strictly increasing, and $\phi (0) = 0$, $S (z'; \ell) > 0$ for sufficiently small $z' > 0$, given ℓ. The result follows from the fact that there always exists sufficiently small z' such that adding a z'-city improves the allocation. ∎

Two direct consequences of Lemma 1 are that the number of layers I is countably infinite and there are countably infinitely many cities between any two cities. It is also straightforward that if $\phi (0) > 0$, such a proof breaks down for two cities sufficiently close

---

[7]Here, although there are no markets, we use the *market area* of a production location to refer to the interval in which the consumers are served by the location.
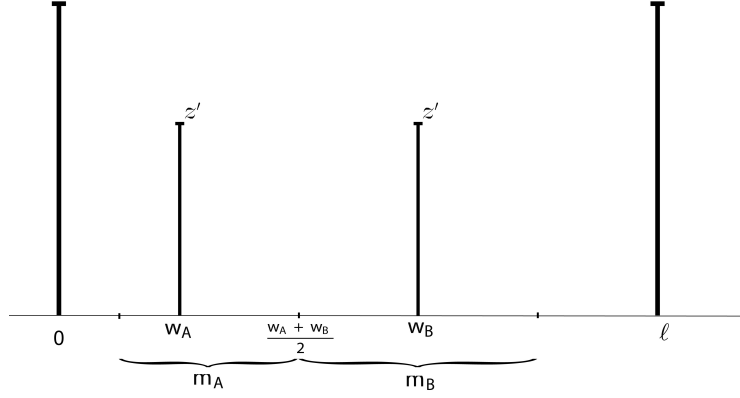
Figure 2: A suboptimal situation: Two cities of the same size z' without a larger city in between them. City k's location and market area are denoted as w_k and m_k, respectively, for $k \in \{A, B\}$.

to each other, i.e., ℓ is sufficiently small. Thus, when $\phi(0) > 0$, the number of cities in an optimal city hierarchy is necessarily finite.

Lemma 2. *It is never optimal to have two cities of the same size $z' < z_1$ without a larger city in between.*

Proof. Suppose that in an optimal allocation there are two cities of the same size z' without a larger city in between them. Without loss of generality, let the two z'-cities be placed in an interval $(0, \ell)$ and two cities whose sizes are larger than z' be placed at the endpoints, i.e., 0 and ℓ. Call the two z'-cities A and B, and denote their locations as w_A and w_B, respectively. See Figure 2 for an illustration. When simultaneously increasing z' at the two z'-cities infinitesimally, there are savings in transport costs because some consumers are closer to z'-cities than to the endpoint cities. Let m_k, $k \in \{A, B\}$ be the interval in which consumers find that city k is the nearest place to buy z', and denote the savings in transport costs due to consumers in m_k switching as $\tilde{s}_k^1$. Note that the optimality of z' requires that $\sum_k \tilde{s}_k^1 = 2\phi(z')$ because the total savings in transport cost from simultaneously increasing z' infinitesimally should equal the total setup cost of z'. Now, denote the savings in transport cost when increasing z' infinitesimally only at city k as s_k^1. Observe that s_k^1 consists of two parts. The first part is the savings experienced by the consumers in m_k, which equal $\tilde{s}_k^1$. The second part exists because a positive measure of consumers outside m_k and in $m_{k'}$, $k' \neq k$, also find city k closer than either of the endpoint cities. Hence, $s_k^1 > \tilde{s}_k^1$, and $\sum_k s_k^1 > \sum_k \tilde{s}_k^1 = 2\phi(z')$. This implies that $s_k^1 > \phi(z')$ for at least one k, which, in turn, implies that the allocation can be improved by increasing z' at this k. The result follows. ∎

8

The key insight is that there is a point in between two z'-cities at which consumers find both of the z'-cities closer than either of the endpoint cities, and when z' increases at either of the z'-cities separately, the consumers around this point benefit either way. Hence, the sum of benefits of increasing the range of goods at each city separately is larger than the benefits of increasing the ranges of goods in both cities simultaneously, the latter of which should equal the total setup cost of z', $2\phi(z')$, if having two z'-cities is optimal. This implies that for *at least one city*, the benefits of increasing the range at this city individually is larger than $\phi(z')$, and we must strictly prefer to increase the range at this city. However, this contradicts the assumption that it is optimal to have two z'-cities.

Note that neither Lemma 1 or 2 depends on the assumption that consumers are uniformly distributed. Now, consider the process of building a city hierarchy in between two z_1-cities. Lemma 1 states that having no cities in between is not optimal. Let z_2 denote the size of the largest cities in between two z_1-cities. Then, Lemma 2 implies that there can be only one z_2-city in an optimal solution. From the perspective of the two z_1-cities, z_2-city is the *immediate sub-city*. Similarly, in between a z_2-city and a z_1-city, there is one and only one immediate sub-city, and this process goes on recursively. This process following the two lemmas entails a simplified problem, which we study in Section 3.

Lemma 2 provides a partial rationale for the central place property, as there is "one" immediate sub-city. If the immediate sub-city is always located in the middle, then the spacing will conform to the central place property. However, this is not necessarily the case, as we subsequently explain.

2.4 An immediate sub-city is not necessarily in the middle

As two cities having the same size z' without a larger city in between, as illustrated in Figure 2, cannot be optimal, suppose that city B has larger size $z'' > z'$ and city A remains at z'. Consider the benefits of moving city B closer to the center. Although this increases savings in transport costs for goods in $(z', z'']$ because they are more centered, it also moves goods $[0, z']$ toward the center and increases the average transport costs for people buying at city B. If there were no hierarchy property, the social planner would split goods and move good $(z', z'']$ to the center and keep $[0, z']$ at A and B. However, the hierarchy property places a constraint. Hence, what is actually optimal depends on the distribution of setup costs. In the following, we provide an example in which the immediate sub-city is not in the middle in the optimal solution.

Let $t = z_1 = \ell_1 = 1$. Consider a discontinuous setup cost requirement function: for an arbitrarily small $e \in (0, 1)$, $\phi(y) = 1/13$ for $y \in [0, e]$ and $\phi(y) = 1$ for $y \in (e, 1]$. It is

readily verified that, in between two z_1-cities, the per capita cost is minimized by evenly placing two immediate sub-cities with $z' = e$.[8] Note that the above lemmas do not have to hold here because ϕ is not smooth and $\phi(0) > 0$. Now, take a smooth approximation of ϕ (with ϕ being continuous, strictly increasing, and $\phi(0) = 0$). The solution must be close to the previous solution, but by Lemma 2, there is just one immediate sub-city. Therefore, the solution must be two sub-cities near $1/3$ and $2/3$ with the range of production of one city being slightly larger than the other. In particular, the immediate sub-city is not half-way in between the two z_1-cities. This example illustrates the possibilities of uneven spacing of cities. The intuition will be clear in our comparison between this example and the central place property result in Proposition 5.

3 Dynamic Programming and the Central Place Property

3.1 The sequence and dynamic programming problems

3.1.1 The sequence problem

Lemmas 1 and 2 indicate that in between two z_1-cities it is optimal to place one and only one immediate sub-city, which is denoted as a z_2-city. The location of the z_2-city divides the interval of length ℓ_1 into two parts. Let $\ell_{2,1}$ and $\ell_{2,2}$ be the distances from the z_2-city to the z_1-city on the left and right side, respectively. When the values of z_2, $\ell_{2,1}$ and $\ell_{2,2}$ are chosen, the recursive nature of the problem becomes apparent because the cost calculations for the goods in $(z_2, z_1]$ become irrelevant to decisions regarding the size and location of the immediate sub-city in each of the two intervals of length $\ell_{2,1}$ and $\ell_{2,2}$. That is, the cost minimization problem given z_2 and $\ell_{2,1}$ and that given z_2 and $\ell_{2,2}$ take the same form as the one given z_1 and ℓ_1.

The two lemmas imply that the city building process, viewed from the top down, involves endless bifurcations. Figure 3 depicts the result from the first three rounds of bifurcations. To write the problem in sequence form, we must develop our notation carefully. As previously mentioned, given ℓ_1 and z_1, the first round of bifurcation involves choosing a z_2-city, the location of which divides the interval of length ℓ_1 into intervals

---

[8]To see this, first note that savings in transport costs per good from having $n \geq 1$ cities is bounded above by the savings when placing these n cities evenly, which equals $\frac{t\ell_1^2}{4} \times \frac{n}{n+1}$ and increases in n. The savings in transport costs of having two evenly-spaced e-cities is then $1/6$ per good and is larger than the setup costs incurred, $2/13$. Obviously, it is not desirable to have more than two e-cities. With $\phi(y)$ being a constant for all $y \in [0, e]$, there are no more cities on $(0, 1)$ besides these e-cities. For $y \in (e, 1]$, since $\phi(y) = 1 > \frac{t\ell_1^2}{4} > \frac{t\ell_1^2}{4} \times \frac{n}{n+1}$, increasing z' from e does not reduce the per capita cost.
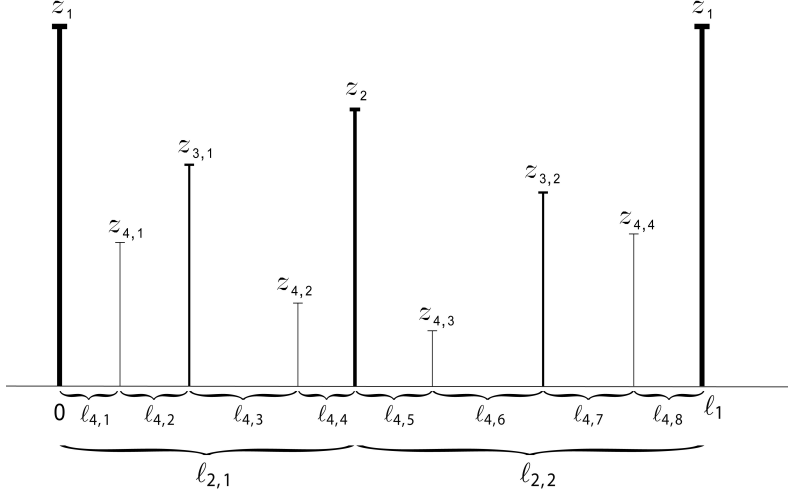
Figure 3: Illustration of Sequence Problem

of length $\ell_{2,1}$ and $\ell_{2,2}$. Then, given z_2, $\ell_{2,1}$, and $\ell_{2,2}$, the second round of bifurcation in-volves choosing a $z_{3,1}$-city and a $z_{3,2}$-city in the intervals of length $\ell_{2,1}$ and $\ell_{2,2}$, respec-tively. The interval of $\ell_{2,1}$ is further divided into $\ell_{3,1}$ and $\ell_{3,2}$, and the interval of $\ell_{2,2}$ is further divided into $\ell_{3,3}$ and $\ell_{3,4}$. In general, the i-th round of bifurcation involves setting up cities of sizes $z_{i+1,1}, z_{i+1,2}, ..., z_{i+1,K_i}$, where $K_i = 2^{i-1}$, which divides intervals of length $\ell_{i,1}, \ell_{i,2}, ..., \ell_{i,K_i}$, respectively. Formally, let $\boldsymbol{z}_1 \equiv \{z_1\}$, and for all $i \in \mathbb{N}$, let $\boldsymbol{\ell}_i \equiv \{\ell_{i,k}\}_{k=1}^{K_i}$ and $\boldsymbol{z}_{i+1} \equiv \{z_{i+1,k}\}_{k=1}^{K_i}$, where $\ell_{1,1} \equiv \ell_1$ and $z_{2,1} \equiv z_2$. We define

$$\Gamma_1\left(\ell_1, z_1\right) \equiv \Gamma\left(\ell_1, z_1\right) \equiv \left\{\left(\boldsymbol{\ell}_2, \boldsymbol{z}_2\right) | z_2 \in [0, z_1], \ell_{2,1}, \ell_{2,2} \in (0, \ell_1) \text{ and } \ell_{2,1} + \ell_{2,2} = \ell_1\right\}. \quad (2)$$

and for $i \geq 2$,

$$\Gamma_i\left(\boldsymbol{\ell}_i, \boldsymbol{z}_i\right) \equiv \left\{ \begin{array}{c} \left(\boldsymbol{\ell}_{i+1}, \boldsymbol{z}_{i+1}\right) | z_{i+1,2k-1}, z_{i+1,2k} \in [0, z_{i,k}] \text{ for all } k = 1, 2, ..., K_{i-1}, \\ \ell_{i+1,2k-1}, \ell_{i+1,2k} \in (0, \ell_{i,k}) \text{ and } \ell_{i+1,2k-1} + \ell_{i+1,2k} = \ell_{i,k}, \text{ for all } k = 1, 2, ..., K_i. \end{array} \right\}$$

Then, define

$$\Pi\left(\ell_1, z_1\right) \equiv \left\{\left(\boldsymbol{\ell}_i, \boldsymbol{z}_i\right)_{i=1}^{\infty} | \left(\boldsymbol{\ell}_{i+1}, \boldsymbol{z}_{i+1}\right) \in \Gamma_i\left(\boldsymbol{\ell}_i, \boldsymbol{z}_i\right), \text{ for all } i = 1, 2, ...\right\}.$$

Any $(\boldsymbol{\ell}, \boldsymbol{z}) \equiv (\boldsymbol{\ell}_i, \boldsymbol{z}_i)_{i=1}^{\infty} \in \Pi\left(\ell_1, z_1\right)$ is called a *feasible* sequence, given (ℓ_1, z_1).

For an immediate sub-city, z'-city, in between two neighboring larger-sized cities, let ℓ be the distance between the two neighboring larger-sized cities, and let the z'-city's distance to one of the two cities be $\alpha\ell$, for $\alpha \in (0, 1)$. The savings in transport costs for

11

each good in $[0, z']$ is

$$s^1(\ell, \alpha) \equiv 2\int_0^{\frac{\ell}{2}} txdx - \left(2\int_0^{\frac{\alpha\ell}{2}} txdx + 2\int_0^{\frac{(1-\alpha)\ell}{2}} txdx\right) = \frac{t\ell^2}{2}\alpha(1-\alpha),\qquad(3)$$

Then, the optimal magnitude of z' is determined by

$$s^1(\ell, \alpha) = \frac{t\ell^2}{2}\alpha(1-\alpha) = \phi(z').\qquad(4)$$

The left-hand side of (4) is the savings in transport costs when increasing z' marginally, whereas the right-hand side is the corresponding setup cost. If z' is low such that $\phi(z') < \frac{t\ell^2}{2}\alpha(1-\alpha)$, it incurs positive net savings (savings in transport costs net of setup costs) by increasing z'. Similarly, when $\phi(z') > \frac{t\ell^2}{2}\alpha(1-\alpha)$, one can improve the allocation by decreasing z'. In sum, Lemmas 1 and 2 and (4) imply that in any optimal city hierarchy the following constraint holds:

$$\left\{\begin{array}{l} z_{i+1,2k-1}, z_{i+1,2k} \in (0, z_{i,k}) \\ \ell_{i+1,2k-1}, \ell_{i+1,2k} \in (0, \ell_{i,k}), \ell_{i+1,2k-1} + \ell_{i+1,2k} = \ell_{i,k} \\ z_{i+1,k} = \phi^{-1}\left(\frac{t}{2}\ell_{i+1,2k-1}\ell_{i+1,2k}\right) \end{array}\right\}.\qquad(5)$$

Equivalently, any optimal city hierarchy is associated with a sequence $\boldsymbol{\alpha} = \{\alpha_{i,k}\}$ such that $\ell_{i+1,2k-1} = \alpha_{i,k}\ell_{i,k}$ (hence $\ell_{i+1,2k} = (1 - \alpha_{i,k})\ell_{i,k}$) and (5) holds.

Note that in defining the choice set of $(\boldsymbol{\ell}, \boldsymbol{z}) \equiv (\boldsymbol{\ell}_i, \boldsymbol{z}_i)_{i=1}^{\infty}$ by Γ_i and Π above, we leave (4) implicit and take the closure of $(0, z_{i,k})$. According to Lemmas 1 and 2, we know that situations in which $z_{i+1,2k-1}$ or $z_{i+1,2k}$ equals 0 or $z_{i,k}$ are never optimal (except possibly for $i = 1$), but we do not lose any generality by including this possibility. When the choice of $z_{i+1,2k-1}$, according to (4) and given $\ell_{i,2k-1}$, is such that $z_{i+1,2k-1} > z_{i,k}$, one can always relabel i, k to ensure that the constraint $z_{i+1,2k-1}, z_{i+1,2k} \in [0, z_{i,k}]$ is obeyed.[9] Thus, the choice set defined by Π encompasses all possible candidates for an optimal city hierarchy. In other words, any sequence $(\boldsymbol{\ell}, \boldsymbol{z})$ that satisfies all constraints in (5) is included in $\Pi(\ell_1, z_1)$. If one would like to make the constraint (4) explicit, one could redefine Γ_i by

---

[9]We allow $z_{i+1,2k-1} = z_{i,k}$ (or, $z_{i+1,2k} = z_{i,k}$) as a choice to keep the choice set of $\alpha_{i,2k-1}$ (or, $\alpha_{i,2k}$) a connected interval. To see this, imagine that we are given $\ell_{i,2k-1}$ and $z_{i,k}$, and we have to choose an $\alpha_{i,2k-1}$ and $z_{i+1,2k-1}$. Suppose we want to choose $\alpha_{i,2k-1} = 1/2$, but according to (4), this can give a $z_{i+1,2k-1} \geq z_{i,k}$ if $\ell_{i,2k-1}$ is very large. We know that $z_{i+1,2k-1} > z_{i,k}$ is not optimal, but we can relabel things in this case. We also know that $z_{i+1,2k-1} = z_{i,k}$ is not optimal, but if we do not even allow this, then there is a neighborhood of $1/2$ that we cannot choose for $\alpha_{i,2k-1}$.

12

replacing $z_{i+1,2k-1}, z_{i+1,2k} \in [0, z_{i,k}]$ with

$$z_{i+1,2k-1} = \min\left\{\phi^{-1}\left(\frac{t}{2}\ell_{i+1,4k-1}\ell_{i+1,4k-2}\right), z_{i,k}\right\}, \quad z_{i+1,2k} = \min\left\{\phi^{-1}\left(\frac{t}{2}\ell_{i+1,4k-3}\ell_{i+1,4k-4}\right), z_{i,k}\right\}.$$

Suppose the social planner has two z-cities with distance ℓ and nothing in between them. The total cost in this interval of ℓ is

$$A(\ell, z) \equiv \Phi(z) + \frac{zt\ell^2}{4}.$$

Note that only one setup cost of a z-city is counted in this definition. When a z'-city divides an interval of ℓ bounded by two cities producing at least up to z, the total cost for the range of goods $(z', z]$ is given by $A(\ell, z) - A(\ell, z') = \Phi(z) - \Phi(z') + (z - z') t\ell^2/4$. We can view the per capita cost for the goods $[0, z_1]$ on ℓ_1 as the sum of the per capita cost of different ranges of goods on different market areas within ℓ_1. Namely, the sequence problem is

$$C^*(\ell_1, z_1) \equiv \inf_{\substack{(\boldsymbol{\ell}, \boldsymbol{z}) \in \Pi(\ell_1, z_1), \\ z_1 > 0 \text{ given.}}} \frac{1}{\ell_1} \left[\begin{array}{l} A(\ell_1, z_1) - A(\ell_1, z_2) \\ + \sum_{i=2}^{\infty} \sum_{k=1}^{K_{i-1}} \left[\begin{array}{l} A(\ell_{i,2k-1}, z_{i,k}) - A(\ell_{i,2k-1}, z_{i+1,2k-1}) \\ + A(\ell_{i,2k}, z_{i,k}) - A(\ell_{i,2k}, z_{i+1,2k}) \end{array} \right] \end{array} \right]. \tag{SP}$$

Let us examine (SP) with reference to Figure 3. Suppose in the definition of (SP), for any $A(\ell_{i,2k-1}, z_{i,k}) - A(\ell_{i,2k-1}, z_{i+1,2k-1})$ (or, $A(\ell_{i,2k}, z_{i,k}) - A(\ell_{i,2k}, z_{i+1,2k})$), we count the setup costs incurred at the city on the left end, but not those at the city on the right-end. Then one sees that (SP) includes all of the setup costs incurred on $[0, \ell_1)$, leaving out the setup costs at ℓ_1. Of course, the city at ℓ_1 is the left-end city of another interval, and hence we do not miss any setup cost over the entire space. Solving (SP) gives the infimum of the per capita cost on the half-open interval of length ℓ_1.

Denote the objective function in (SP) as $f(\boldsymbol{\ell}, \boldsymbol{z}) = \lim_{n \to \infty} f_n(\boldsymbol{\ell}, \boldsymbol{z})$, where $f_n(\boldsymbol{\ell}, \boldsymbol{z})$ is the objective function with ∞ replaced by n. Because the partial sum $f_n(\boldsymbol{\ell}, \boldsymbol{z})$ is bounded in $[0, A(\ell_1, z_1)/\ell_1]$ and nondecreasing in n, it converges for any given $(\boldsymbol{\ell}, \boldsymbol{z}) \in \Pi(\ell_1, z_1)$. As the value of the objective is bounded in $[0, A(\ell_1, z_1)/\ell_1]$, C^* is uniquely defined with $C^*(\ell_1, z_1) \in [0, A(\ell_1, z_1)/\ell_1]$ for all (ℓ_1, z_1).

3.1.2   The dynamic programming problem

Given state variables ℓ and z, the social planner needs to decide the size and location of the immediate sub-city, z'-city. Denote the length of the intervals to the left/right of z'-city

as ℓ_l/ℓ_r. Then, $\ell_l + \ell_r = \ell$. Alternatively, let $\ell_l = \alpha\ell$ and $\ell_r = (1 - \alpha)\ell$ for $\alpha \in (0, 1)$. We present the following dynamic programming problem.

$$
\begin{aligned}
C(\ell, z) &= \inf_{\ell_l, \ell_r \in (0,\ell), \ell_l + \ell_r = \ell, z' \in [0,z]} \frac{1}{\ell} \left[A(\ell, z) - A(\ell, z') + \ell_l C(\ell_l, z') + \ell_r C(\ell_r, z') \right] \\
&= \inf_{\alpha \in (0,1), z' \in [0,z]} \frac{1}{\ell} [A(\ell, z) - A(\ell, z')] + \alpha C(\alpha\ell, z') + (1 - \alpha) C((1 - \alpha)\ell, z') \quad (DP)
\end{aligned}
$$

The solution to the above problem is a *cost function* C and *policy functions* $z' = g_z(\ell, z)$ and $\alpha = g_\alpha(\ell, z)$, which entail the next state variables for each side of the division, given current (ℓ, z): $(\ell_l, z') = (g_\alpha(\ell, z)\ell, g_z(\ell, z))$ and $(\ell_r, z') = ((1 - g_\alpha(\ell, z))\ell, g_z(\ell, z))$. The problem (DP) is much more compact than the sequence problem (SP), as the recursive nature allows the terms $\alpha C(\alpha\ell, z')$ and $(1 - \alpha)C((1 - \alpha)\ell, z')$ to subsume the per capita cost for all goods in $[0, z']$ on the intervals of length $\alpha\ell$ and $(1 - \alpha)\ell$, respectively.

3.1.3 Total rather than per capita cost

It is often useful to look at a transformation of (SP) and (DP) by letting $D^*(\ell_1, z_1) = \ell_1 C^*(\ell_1, z_1)$, and $D(\ell, z) = \ell C(\ell, z)$:

$$
D^*(\ell_1, z_1) = \inf_{\substack{(\boldsymbol{\ell}, \boldsymbol{z}) \in \Pi(\ell_1, z_1), \\ z_1 > 0 \text{ given.}}} \begin{array}{c} A(\ell_1, z_1) - A(\ell_1, z_2) \\ + \sum_{i=2}^{\infty} \sum_{k=1}^{K_{i-1}} \left[\begin{array}{c} A(\ell_{i,2k-1}, z_{i,k}) - A(\ell_{i,2k-1}, z_{i+1,2k-1}) \\ + A(\ell_{i,2k}, z_{i,k}) - A(\ell_{i,2k}, z_{i+1,2k}) \end{array} \right] \end{array}, \quad (SP^D)
$$

and

$$
D(\ell, z) = \inf_{\alpha \in (0,1), z' \in [0,z]} A(\ell, z) - A(\ell, z') + D(\alpha\ell, z') + D((1 - \alpha)\ell, z'). \quad (DP^D)
$$

D^* is the infimum of total cost, rather than per capita cost, for all of the goods $[0, z_1]$ on the interval of length ℓ_1. For any solution C to DP, $D = \ell C$ is a solution to (DP^D), and vice versa.

3.2 Characterization theorems in a more general setting

As intuitive as it is, the equivalence between the sequence problem and its corresponding dynamic programming problem, i.e., the *principle of optimality,*[10] requires a proof. This is because, for all we know so far, (DP) or, equivalently, (DP^D), may have zero, one, or

---

[10]See Lucas and Stokey (1989) for an exposition of the principle of optimality in a time sequence problem.

many solutions, and some kind of transversality condition is needed for a solution of (DP) to be the infimum function in (SP).

As there is no discount factor in (DP) as in a typical time sequence problem, it is not obvious whether the mapping defined by (DP) is a contraction. Nevertheless, (DP) can still be characterized in a similar way to what is stated in a contraction mapping theorem. That is, the mapping defined by the right-hand side of (DP) is a self-mapping with C^* as the unique fixed point. Moreover, this fixed point can be approximated by iterating the mapping. These results, along with the principle of optimality, apply to a more general setting with a recursive division structure. In this subsection, we show these results in a general setting, while in the next subsection we show how the conditions of these theorems are satisfied in the city hierarchy problem (SP^D) and (DP^D). The results of the original (SP) and (DP) then follow.

It is worth noting that these characterization results, especially the convergence theorem, are potentially useful in various settings involving recursive division, which often occurs in a hierarchical structure. For example, it can be used to study a firm hierarchy in which a CEO would like to determine the best division of labor among different posts at different layers of the hierarchy, which often invokes a recursive structure.[11] Other possible venues for application include outsourcing in trade and the structure of fiscal decentralization.

3.2.1 A general setting with q-furcation and without discount factor

Consider a generalized setting in which the law of motion stipulates multi-furcation, or, q-furcation, where q is a positive integer. The city hierarchy problem is simply one in which $q = 2$. We first formulate a general sequence problem. Let $X \subseteq \mathbb{R}^N$ denote the domain of state variables and $\Gamma : X \to X^q \subseteq \mathbb{R}^{qN}$ be a nonempty correspondence. Let $K_i = q^{i-1}$, $i \in \mathbb{N}$. For the i-th q-*furcation*, let index $k = 1, 2, ..., q^{i-1}$ be grouped into q^{i-2} sets of indices such that for $j = 1, 2, ..., q^{i-2}$,

$$k_j = (j-1)\, q + 1, (j-1)\, q + 2, ..., jq.$$

That is, $k_1 = 1, 2, ..., q$, $k_2 = q + 1, q + 2, ..., 2q$, and so on. Denote $\mathbf{x}_{i,j} \equiv \left\{ x_{i,k_j} \right\}_{k_j} \in \mathbb{R}^{qN}$, and $\mathbf{x}_i \equiv \{\mathbf{x}_{i,j}\}_{j=1}^{q^{i-2}}$. For $i \geq 2$, define

$$\Gamma_i\, (\mathbf{x}_i) \equiv \left\{ \mathbf{x}_{i+1} \in X^{q^i} \subseteq \mathbb{R}^{q^i N} | \mathbf{x}_{i+1,k} \in \Gamma\, (x_{i,k}) \text{ for } k = 1, 2, ..., K_i \right\}. \tag{6}$$

---

[11] An interesting firm hierarchy problem is studied by Garicano and Rossi-Hansberg (2006), who show in their Appendix 1 that the problem becomes recursive when the decision-making process is decentralized.

In words, Γ_i recursively defines the set of feasible $\mathbf{x}_{i+1,k}$ qN-dimensional vectors, given a $x_{i,k}$ N-dimensional vector. A sequence $\mathbf{x} = \{\mathbf{x}_i\}_{i=1}^{\infty}$ is *feasible* if it satisfies $\mathbf{x}_{i+1} \in \Gamma_i(\mathbf{x}_i) \subseteq \mathbb{R}^{q^i N}$ for all i. Let the set of all feasible sequences starting with $x_1 \in X$ be denoted as $\Pi(x_1)$.

Let F be a real-valued function on $\mathbb{R}^N \times \mathbb{R}^{qN}$. The sequence problem is then defined as

$$V^*(x_1) \equiv \inf_{\substack{\mathbf{x} \in \Pi(x_1), \\ x_1 \in X \text{ given.}}} F(x_1, \mathbf{x}_2) + \sum_{i=2}^{\infty} \sum_{k=1}^{K_i} F(x_{i,k}, \mathbf{x}_{i+1,k}), \qquad (SP^V)$$

The superscript V is reserved for the general sequence and dynamic programming problems. The corresponding dynamic programming problem is

$$V(x) \equiv \inf_{\substack{\mathbf{y} \in \Gamma(x), \\ x \in X \text{ given.}}} F(x, \mathbf{y}) + \sum_{k=1}^{q} V(y_k). \qquad (DP^V)$$

We design the pattern of the notation such that the superscript * indicates it is the optimum value of the sequence problem (C^*, D^*, and V^*), whereas the capital letter without this superscript (C, D, and V) indicates the solution to the corresponding dynamic programming problem.

This problem also applies to $q = 1$, and in this case, the sequence and dynamic programming problems are simply $V^*(x_1) \equiv \inf \sum_{i=1}^{\infty} F(x_i, \mathbf{x}_{i+1})$, and $V(x) \equiv \inf[F(x, \mathbf{y}) + V(y)]$.

3.2.2 Principle of optimality

The following two propositions establish the principle of optimality in the general setting. As both proofs follow steps similar to those in Lucas and Stokey (1989, pp. 67-76), they are relegated to the separate appendix. The first concerns the equivalence between V^* and the solution V to (DP^V).

Proposition 1. *(i) V^*, the infimum function defined in the sequence problem (SP^V), is a solution to the dynamic programming problem (DP^V). (ii) If a function V is a solution to (DP^V) and if*

$$\lim_{i \to \infty} \sum_{k=1}^{K_i} V(x_{i,k}) = 0, \qquad (7)$$

then $V = V^*$.

Proof. See the separate appendix. ∎

We offer a brief account of the proof of Proposition 1. Let u_n be the partial sum of the objective in (SP^V), i.e., the objective with n in place of ∞. Denote the objective in (SP^V) as $u(\mathbf{x}) = \lim_{n \to \infty} u_n(\mathbf{x})$. For all $\mathbf{x} \in \Pi(x_1)$, let $\mathbf{y}_k$ be the k-th part of the sequence starting from the second round of q-furcation. That is, $\mathbf{y}_k \in \Pi(y_k)$, where $\{y_k\}_{k=1}^q \in \Gamma(x_1)$. It is easily verified that for any $\mathbf{x} \in \Pi(x_1)$,

$$u(\mathbf{x}) = F(x, \mathbf{y}) + \sum_{k=1}^{q} u(\mathbf{y}_k).$$

That is, for an arbitrary feasible sequence, the objective in (SP^V) can be written in a recursive way, as in (DP^V). The proof for V^* solving (DP^V) therefore involves carefully showing why we can replace u with V^*. In the reverse direction, we show in the separate appendix that, by induction, any V that satisfies (DP^V) must satisfy

$$V(x_1) \leq \inf_{\{\mathbf{x}_i\}_{i=1}^{n+1}} \left[F(x, \mathbf{x}_2) + \sum_{i=2}^{n} \sum_{k=1}^{K_i} F(x_{i,k}, \mathbf{x}_{i+1,k}) + \sum_{k=1}^{K_{n+1}} V(x_{n+1,k}) \right].$$

We thus need (7), i.e., the limit of the residual term is 0, so that $V(x_1) \leq u(\mathbf{x})$ for all $\mathbf{x} \in \Pi(x_1)$. The rest of the proof involves showing $V(x_1) + \epsilon \geq u(\mathbf{x})$ for any $\epsilon > 0$ and some $\mathbf{x} \in \Pi(x_1)$, and this residual term also appears.

Part (i) of Proposition 1 says that V^* is a solution to (DP^V), and hence there is at least one solution to (DP^V). Part (ii) says that any solution to (DP^V) that satisfies (7) must be V^*.

The second proposition concerns the equivalence between an optimal sequence in (SP^V) and a sequence that satisfies the functional equation in (DP^V) recursively.

Proposition 2. *(i) If a feasible sequence* $\mathbf{x}^* \in \Pi(x_1)$ *attains the infimum in* (SP^V), *then it satisfies*

$$V^*(x_{i,j}^*) = F\left(x_{i,j}^*, \mathbf{x}_{i+1,j}^*\right) + \sum_{k_j} V^*\left(x_{i+1,k_j}^*\right). \tag{8}$$

(ii) If a feasible sequence $\mathbf{x}^* \in \Pi(x_1)$ *satisfies (8), and if it satisfies (7) with* V^* *in place of* V, *then it attains the infimum in* (SP^V).

Proof. See the separate appendix. ∎

If a feasible sequence attains the infimum in (SP^V), i.e., a feasible sequence as a minimizer exists so that the infimum is indeed a minimum given by this sequence, then it also solves (DP^V) recursively with $V = V^*$. The reverse is also true, provided that V^* satisfies (7). We have not yet proven the existence of such a sequence, but this will be addressed in the next proposition.

3.2.3 Dynamic programming mapping and the convergence of iterates

For any continuous, real-valued function v on X, let the mapping T be given by the right-hand side of (DP^V), i.e.,

$$Tv(x) \equiv \inf_{\substack{\mathbf{y} \in \Gamma(x), \\ x \in X \text{ given.}}} F(x, \mathbf{y}) + \sum_{k=1}^{q} v(y_k). \tag{9}$$

A fixed point of the mapping is a solution to (DP^V), and vice versa. The first of the two following propositions connects the mapping T to both (DP^V) and (SP^V), and the second shows that the sequence of iterates converges to V^*, provided that the initial v is in certain space. Namely, denote the sequence of iterates of the mapping T as T^n, then $\lim_{n \to \infty} T^n v = V^*$. Assume that the following hold.

A1. Γ is compact valued.

A2. For each $x_1 \in X$ and each feasible sequence $\mathbf{x} \in \Pi(x_1)$, $\lim_{i \to \infty} x_{i,k} = 0$.

A3. $F : X \times X^q \to \mathbb{R}_+$ is a continuous function with nonnegative values, and $F(\mathbf{0}) = 0$.

Proposition 3. *Suppose that A1, A2, and A3 hold. Also suppose that there exists a continuous and strictly increasing function $M : X \to \mathbb{R}_+$ with $M(\mathbf{0}) = 0$ such that if any continuous real-valued function v satisfies $0 \le v(x) \le M(x)$, then $0 \le Tv(x) \le M(x)$. Denote the set of continuous functions v satisfying $0 \le v(x) \le M(x)$ as $\mathcal{V}(X)$. Then, the following hold.*

(i) Tv is continuous. Hence, T is a self-mapping on $\mathcal{V}(X)$.

(ii) The minimum is attained; so $\inf$ in the definition of T in (9) can be replaced with $\min$. Moreover, the set of minimizers is an upper hemi-continuous correspondence on X.

(iii) V^* is the unique solution to (DP^V) in $\mathcal{V}(X)$ and hence the unique fixed point of the mapping T on $\mathcal{V}(X)$.

Proof. That the minimum is obtained follows directly from the facts that both F and v are continuous and that $\Gamma(x)$ is compact for any $x \in X$. That Tv is continuous and the set of minimizers given $x \in X$ is an upper hemi-continuous correspondence on X follows from the Theorem of the Maximum (see Lucas and Stokey 1989, p. 62). Since Tv is continuous and $0 \le Tv(x) \le M(x)$ for any continuous v such that $0 \le v(x) \le M(x)$, T is a self-mapping on $\mathcal{V}(X)$. Because $M(\mathbf{0}) = 0$ and for all k, $\lim_{i \to \infty} x_{i,k} = 0$, any $v \in \mathcal{V}(X)$ satisfies (7) with v in place of V. Then, according to Proposition 1, if any $v \in \mathcal{V}(X)$ is a

18

solution to (DP^V), in which case we denote this particular v as V, then it equals of V^*. As V^* is uniquely defined and a solution to (DP^V), V^* is the unique solution to (DP^V) in $\mathcal{V}(X)$, and hence the unique fixed point of T on $\mathcal{V}(X)$. ■

Proposition 4. *Suppose all of the conditions in Proposition 3 hold. Then, for any $v \in \mathcal{V}(X)$, the sequence $\{T^n v\}$ converges to V^*.*

Proof. The complete proof is relegated to the separate appendix, and we provide a sketch here. In the separate appendix, we show that $\{T^n v\}$ is Cauchy, and hence $\{T^n v\}$ converges. The intuition behind $\{T^n v\}$ being Cauchy is briefly explained as follows. For an arbitrary $v \in \mathcal{V}(X)$, Tv is a minimization problem with one $F(x, \mathbf{x}')$. To get $T^2 v$, one replaces v with Tv, and hence $T^2 v$ becomes a problem of $1 + q$ min operators with q minimization problems embedded in an overall one. When repeating this process to get $T^n v$, there are numerous terms similar to $F(x, \mathbf{x}')$ with $x, \mathbf{x}'$ properly replaced by the sequence notation. Even though the problem implied by $T^n v$ is not the same as (SP^V), the difference between them diminishes as n gets large. This is mainly because v appears only at the very end of the (expanded) $T^n v$ problem, and when examining $|T^{n+1} v - T^n v|$, it is easy to verify that the difference is a matter of two multi-furcations at the end, i.e., for $i = n, n+1$. The fact that any $v \in \mathcal{V}(X)$ satisfies (7) with v in place of V implies that the difference eventually disappears as n goes to infinity.

Although $T^n v(x)$ as a *hierarchy* of minimization problems is different from the partial sum version of (SP^V), it is shown in the separate appendix that for all $x \in X$,

$$T^n v(x) \le \min_{\mathbf{x} \in \Pi(x)} \left[u_{n-1}(\mathbf{x}) + \sum_{k=1}^{K_n} Tv(x_{n,k}) \right]. \tag{10}$$

Since $Tv \in \mathcal{V}(X)$ so that (7) holds with Tv in place of V, take n to infinity and we have

$$\lim_{n \to \infty} T^n v(x) \le \min_{\mathbf{x} \in \Pi(x)} u(\mathbf{x}) = V^*(x).$$

Now, denote the optimal sequence that solves this $T^n v(x)$ problem as $\{\tilde{\mathbf{x}}_i\}_{i=1}^{n+1}$. The existence of such a sequence is guaranteed by Proposition 3 because each of the minimization problems in the expanded problem implied by $T^n v(x)$ has a minimizer. Obviously, $\tilde{\mathbf{x}} \equiv \{\tilde{\mathbf{x}}_i\}_{i=1}^{\infty} \in \Pi(x_1)$. Taking n to infinity and by definition of V^*,

$$\begin{aligned} \lim_{n \to \infty} T^n v(x) &= \lim_{n \to \infty} \left\{ u_{n-1}(\tilde{\mathbf{x}}) + \sum_{k=1}^{K_n} Tv(\tilde{x}_{n,k}) \right\} \\ &= u(\tilde{\mathbf{x}}) \ge V^*(x). \end{aligned} \tag{11}$$

The result that $\lim_{n \to \infty} T^n v(x) = V^*(x)$ follows from (10) and (11). $\blacksquare$

3.3 Characterization in the city hierarchy problem

To see how the city hierarchy problem is a special case of (SP^V) and (DP^V), let $q = 2$ and $X = [0, \ell_1] \times [0, z_1] \subseteq \mathbb{R}^2$. Let $F(x, \mathbf{y}) = A(\ell, z) - A(\ell, z')$, where $x = (\ell, z)$ and Γ is given by (2) such that $\mathbf{y} = \{(\alpha \ell, z'), ((1 - \alpha) \ell, z')\}$, where $\alpha \in (0, 1)$, and $z' \in [0, z]$. For the sequence notation, $x_1 = (\ell_1, z_1)$, and for $i \geq 2$, $\mathbf{x}_{i,j} = \{x_{i,2j-1}, x_{i,2j}\} = \{(\ell_{i,2j-1}, z_{i,j}), (\ell_{i,2j}, z_{i,j})\}$, for $j = 1, 2, ..., q^{i-2}$. Hence, for $i \geq 2$, $F(x_{i,2j-1}, \mathbf{x}_{i+1,2j-1}) = A(\ell_{i,2j-1}, z_{i,j}) - A(\ell_{i,2j-1}, z_{i+1,2j-1})$, and $F(x_{i,2j}, \mathbf{x}_{i+1,2j}) = A(\ell_{i,2j}, z_{i,j}) - A(\ell_{i,2j}, z_{i+1,2j})$. One obtains (SP^D) and (DP^D) by substituting all of these into (SP^V) and (DP^V).

In this subsection, we explain how the conditions of Propositions 1 to 4 hold. We start with the principle of optimality, i.e., Propositions 1 and 2.

Corollary 1. *For any two positive real numbers ℓ_1 and z_1, let $X = [0, \ell_1] \times [0, z_1]$, and let $\mathcal{D}(X)$ denote the set of all real-valued continuous functions $d : X \to \mathbb{R}_+$ such that*

$$0 \leq d(\ell, z) \leq A(\ell, z). \tag{12}$$

Then, the following hold.

(i) D^* is the unique solution to (DP^D) in $\mathcal{D}(X)$.

(ii) A feasible sequence $(\boldsymbol{\ell}^*, \mathbf{z}^*) \in \Pi(\ell_1, z_1)$ attains the infimum in (SP^D) if and only if it satisfies (DP^D) recursively, i.e.,

$$D^*(\ell_{i,k}^*, z_{i,k}^*) = A(\ell_{i,k}^*, z_{i,k}^*) - A(\ell_{i,k}^*, z_{i+1,k}^*) + D^*(\ell_{i+1,2k-1}^*, z_{i+1,k}^*) + D^*(\ell_{i+1,2k}^*, z_{i+1,k}^*).$$

Proof. First note that as ℓ_1 and z_1 are positive, the constraint Γ given by (2) is nonempty. Note that **A2** holds as a result of Lemmas 1 and 2. Then, for $d \in \mathcal{D}(X)$ and any feasible sequence $(\boldsymbol{\ell}, \mathbf{z}) \in \Pi(\ell, z)$,

$$0 \leq \sum_{k=1}^{K_n} [d(\ell_{n+1,2k-1}, z_{n+1,k}) + d(\ell_{n+1,2k}, z_{n+1,k})] \leq \sum_{k=1}^{K_n} [A(\ell_{n+1,2k-1}, z_{n+1,k}) + A(\ell_{n+1,2k}, z_{n+1,k})],$$

of which the right-hand side goes to 0 as n goes to infinity because $A(\mathbf{0}) = 0$. Hence, if any D is a solution to (DP^D) and $D \in \mathcal{D}(X)$, then

$$\lim_{n \to \infty} \sum_{k=1}^{K_n} [D(\ell_{n+1,2k-1}, z_{n+1,k}) + D(\ell_{n+1,2k}, z_{n+1,k})] = 0, \tag{13}$$

which is the version of (7) in the city hierarchy problem. Because there are positive savings from building smaller cities, $D^*(\ell_1, z_1) \in [0, A(\ell_1, z_1)]$, and hence by part (i) of Proposition 1, D^* is a solution to (DP^D) in $\mathcal{D}(X)$. According to (13) and part (ii) of Proposition 1, any solution D to (DP^D) in $\mathcal{D}(X)$ equals D^*. Hence, D^* is the unique solution to (DP^D) in $\mathcal{D}(X)$. As the conditions needed for Proposition 2 are the same, the second result follows. ∎

To show that Propositions 3 and 4 hold in the city hierarchy problem, we need the following lemma.

Lemma 3. *Let $\mathcal{D}(X)$ be given by Corollary 1. Let $T : \mathcal{D}(X) \to \mathcal{D}(X)$ be given by, for each $d \in \mathcal{D}(X)$,*

$$Td(\ell, z) \equiv \inf_{\alpha \in (0,1), z' \in [0,z]} A(\ell, z) - A(\ell, z') + d(\alpha \ell, z') + d((1-\alpha)\ell, z'). \tag{14}$$

Then,

(i) $0 \leq Td(\ell, z) \leq A(\ell, z)$.

(ii) Given $(\ell, z) \in X$, there exists an $\epsilon > 0$ such that every optimal choice of $\alpha \in [\epsilon, 1 - \epsilon]$ for all d. That is, the optimal choice of α cannot be arbitrarily close to 0 or 1.

Proof. That $Td \geq 0$ is trivial. Use (12) to write

$$
\begin{aligned}
Td(\ell, z) &\leq \inf_{\alpha \in (0,1), z' \in [0,z]} A(\ell, z) - A(\ell, z') + A(\alpha \ell, z') + A((1-\alpha)\ell, z') \\
&= \inf_{\alpha \in (0,1), z' \in [0,z]} A(\ell, z) + \Phi(z') - \frac{z' t \ell^2}{2} \alpha(1 - \alpha).
\end{aligned}
\tag{15}
$$

By (4),

$$\Phi(z') - \frac{z' t \ell^2}{2} \alpha(1 - \alpha) = \int_0^{z'} \left[\phi(y) - \frac{t \ell^2}{2} \alpha(1 - \alpha) \right] dy < 0.$$

Hence, Td satisfies (12). The proof of part (ii) is relegated to the separate appendix. The intuition is that when α is arbitrarily close to either 0 or 1, z' tends to 0, and the city hierarchy so built is close to nonexistent. Therefore, such α cannot be optimal because it must fare worse than $\alpha = 1/2$, which guarantees positive savings. ∎

The following corollary shows how Propositions 3 and 4 hold in the city hierarchy problem.

Corollary 2. *Let $\mathcal{D}(X)$ and T be given by Corollary 1 and Lemma 3. Then,*

(i) Td is continuous. Hence, T is a self-mapping on $\mathcal{D}(X)$.

21

(ii) The minimum is attained; so inf *in the definition of T in (14) can be replaced with* min. *Moreover, the set of minimizers is an upper hemi-continuous correspondence on X.*

(iii) D^* is the unique solution to (DP^D) in $\mathcal{D}(X)$ and hence the unique fixed point of the mapping T on $\mathcal{D}(X)$.

(iv) For any $d \in \mathcal{D}(X)$, the sequence $\{T^n d\}$ converges to D^*.

Proof. The four points (i)-(iv) are the results in Propositions 3 and 4. We must show that **A1**, **A2**, and **A3** hold, and that there exists an M that satisfies the condition described in Proposition 3. First, **A2** holds as a result of Lemmas 1 and 2. Recall that for $i \geq 2$, $F(x_{i,2j-1}, \mathbf{x}_{i+1,2j-1}) = A(\ell_{i,2j-1}, z_{i,j}) - A(\ell_{i,2j-1}, z_{i+1,2j-1})$, and $F(x_{i,2j}, \mathbf{x}_{i+1,2j}) = A(\ell_{i,2j}, z_{i,j}) - A(\ell_{i,2j}, z_{i+1,2j})$. Hence, **A3** holds. Lemma 3 shows that although the constraint Γ given by (2) is not compact because $\alpha \in (0,1)$, the effective constraint set is compact, and hence **A1** holds. It also shows that we can simply set $M = A$, which is continuous and strictly increasing with $A(\mathbf{0}) = 0$. $\blacksquare$

We have written Matlab programs implementing this iterative method of finding a solution with any ϕ and/or any initial guess. In particular, because we have the analytical solution when ϕ is a power function (see the next subsection), we compare the numerical solution with the analytical one in this case. The approximation works well, and the convergence is achieved quickly.[12] These programs are available to interested readers upon request.

3.4 Central place property

The beauty of Corollary 2 is that it allows the solution to be found numerically for any arbitrary ϕ that satisfies the basic assumptions. Nevertheless, there is an interesting and empirically relevant case of ϕ in which we can obtain the solution analytically. It turns out that the central place property holds in this case.

Suppose that the setup cost is a power function: $\phi(z) = az^b$, for $a > 0$ and $b > 0$. Under this functional form, $\Phi(z) = \frac{a}{b+1} z^{b+1}$. The power function assumption of ϕ, in fact, means that the distribution of setup costs across goods is also a power function. Let Y denote the random variable of setup cost for a good. Then, for $y \in [0, \phi(z_1)]$,

$$\Pr[Y \leq y] = \frac{\phi^{-1}(y)}{z_1} = \frac{1}{z_1} \left(\frac{y}{a}\right)^{\frac{1}{b}}.$$

---

[12]For the power ϕ case, depending on the level of tolerance, $T^n c$ converges in about 10 to 15 iterations, and, when plotted, the limit function is visually indistinguishable from the true analytical C^* given by (21) except near the boundary of the domain.

As shown in Hsu's (2012) equilibrium model of central place hierarchy, this distribution of setup cost is a prototype of a class of distributions that leads to a power law distribution of city size.[13] This class encompasses several well-known, commonly used distributions. See Hsu (2012) for more details.

Recall that it is possible that the optimal $z_2 = z_1$ if ℓ_1 is too large. Note from (4) that savings $s^1(\ell_1, \alpha)$ is bounded by $s^1(\ell_1, 1/2) = t\ell_1^2/8$. Define $\bar{\ell}(z)$ by the solution of ℓ in the following equation.

$$\frac{t\ell^2}{8} = \phi(z). \tag{16}$$

Then, for any $\ell_1 < \bar{\ell}(z_1)$, optimal $z_2 < z_1$, and thus the two z_1-cities with distance ℓ_1 are neighboring. For the rest of the analyses in this paper, we impose the condition that $\ell_1 < \bar{\ell}(z_1)$.

Proposition 5. *Suppose that $\ell_1 < \bar{\ell}(z_1)$, where $\bar{\ell}(z)$ is defined as the solution to (16). Suppose that the setup cost function $\phi(y) = az^b$, for positive constants a and b. Then, the central place property holds.*

Proof. For ease of presentation, let $a = 1$. A general $a > 0$ does not change the result. From (4),

$$z' = \left(\frac{t\ell^2}{2}\alpha(1-\alpha)\right)^{\frac{1}{b}} = \left(\frac{t}{2}\alpha(1-\alpha)\right)^{\frac{1}{b}} \ell^{\frac{2}{b}} \equiv \kappa(\alpha)\ell^{\frac{2}{b}}. \tag{17}$$

Equation (17) implicitly assumes that $z' < z$. Recall that Lemma 2 rules out $z_{i+1,2k-1} = z_{i,k}$ or $z_{i+1,2k-1} = z_{i,k}$ as an optimal solution, and hence (17) is necessarily true for all optimal choices of $z_{i,k}$, except possibly for $i = 2$. However, the constraint $\ell_1 < \bar{\ell}(z_1)$ ensures that optimal $z_2 < z_1$.

Recall from (5) that there is a sequence $\boldsymbol{\alpha} = \{\alpha_{i,k}\}$ associated with any sequence $(\boldsymbol{\ell}, \boldsymbol{z}) \in \Pi(\ell_1, z_1)$. The fact that the optimal solution of z' is separable in ℓ and α implies that, except for z_1, we can write $z_{i,k} = \ell_1^{2/b}h_{i,k}(\boldsymbol{\alpha})$ and $\ell_{i,k} = \ell_1 g_{i,k}(\boldsymbol{\alpha})$, for some functions $h_{i,k}$ and $g_{i,k}$. Thus, both $A(\ell_{i,2k-1}, z_{i,k}) - A(\ell_{i,2k-1}, z_{i+1,2k-1})$ and $A(\ell_{i,2k}, z_{i,k}) - A(\ell_{i,2k}, z_{i+1,2k})$ are multiplicatively separable in $\ell_1^{2(b+1)/b}$ and some functions of $\boldsymbol{\alpha}$. Thus,

---

[13]As we mention in footnote 6, there are deviations from the power law when smaller cities and towns are included. For example, Eeckhout (2004) shows that lognormal distribution fits better than the power law in this case. In fact, it can be verified by following the procedure in Hsu (2012) that when $\phi(0) > 0$, the Zipf's plot of city size (log of rank vs. log of size) is concave, which would be the case under log-normal distribution. The larger the value of $\phi(0)$, the larger the concavity. The reason for such concavity is that when there are finite layers, the entire hierarchy is less of a fractal structure and deviation from the power law is observed.

for some function H, (SP) can be rewritten as

$$C^*(\ell_1, z_1) = \frac{z_1^{b+1}}{(b+1)\,\ell_1} + \frac{z_1 t \ell_1}{4} + \ell_1^{\frac{b+2}{b}} H(\boldsymbol{\alpha}^*).$$

By Corollarys 1 and 2, an optimal $\boldsymbol{\alpha}^*$ exists, and as such, $H(\boldsymbol{\alpha}^*)$ is well defined. Note that $H(\boldsymbol{\alpha}^*) < 0$, and $\ell_1^{(b+2)/b} |H(\boldsymbol{\alpha}^*)|$ is the per capita savings from building the optimal city hierarchy. Given the equivalence between (SP) and (DP), observe that the negative of per capita savings from having an optimal city hierarchy in an interval of ℓ is given by

$$\tilde{S}(\ell, z) \equiv C(\ell, z) - \frac{A(\ell, z)}{\ell} = C(\ell, z) - \frac{z^{b+1}}{(b+1)\,\ell} - \frac{zt\ell}{4} = \ell^{\frac{b+2}{b}} H(\boldsymbol{\alpha}^*). \tag{18}$$

This says that the $\tilde{S}$ function is homogenous of degree $(b+2)/b$ in ℓ and independent of z. With a little abuse of notation, we write $\tilde{S}(\ell) = \tilde{S}(\ell, z)$. Given (17) and (18), (DP) can be rewritten as

$$\tilde{S}(\ell) = \min_{z' \in (0,z), \alpha \in (0,1)} \frac{A(\alpha\ell, z') + A((1-\alpha)\ell, z') - A(\ell, z')}{\ell} + \left[\alpha^{\frac{2(b+1)}{b}} + (1-\alpha)^{\frac{2(b+1)}{b}}\right] \tilde{S}(\ell) \tag{19}$$

Thus,

$$\tilde{S}(\ell) = \frac{-b}{b+1} \left(\frac{t}{2}\right)^{\frac{b+1}{b}} \ell^{\frac{b+2}{b}} \max_{\alpha \in (0,1)} \frac{[\alpha(1-\alpha)]^{\frac{b+1}{b}}}{1 - \alpha^{\frac{2(b+1)}{b}} - (1-\alpha)^{\frac{2(b+1)}{b}}}. \tag{20}$$

We show in the separate appendix that the unique solution to the maximization problem in (20) is $\alpha = 1/2$. ∎

Observe that the optimal sequence $\boldsymbol{\alpha}^*$ does not depend on ℓ_1. The recursive nature implies that for all i, k, the optimal sequence in the interval of $\ell_{i,k}$, i.e., $\{\alpha_{i',k}\}_{i' \geq i}$, does not depend on the magnitude of $\ell_{i,k}$. Thus, under this power function distribution of setup costs, the optimal city hierarchy in any interval of $\ell_{i,k}$ resembles that of the entire one in ℓ_1. As Hsu (2012) shows that this *scale-free* property gives the city hierarchy a *fractal structure*; specifically, the structure of the smaller part of the hierarchy resembles that of the larger.

With optimal $\alpha = 1/2$, using (18) and (20), we obtain the cost function

$$C(\ell, z) = \frac{z^{b+1}}{(b+1)\,\ell} + \frac{zt\ell}{4} - \frac{b}{b+1} \frac{1}{2^{\frac{3(b+1)}{b}} - 2^{\frac{2b+1}{b}}} t^{\frac{b+1}{b}} \ell^{\frac{b+2}{b}}. \tag{21}$$

One can verify Proposition 5 by applying the guess-and-verify technique to (DP). That is, if one plugs the functional form of C given by (21) into the right-hand side of (DP) and solves the minimization problem, one will find that the unique minimizer is $\alpha = 1/2$.

Then, the value of the right-hand side, given (ℓ, z), will be exactly (21), which verifies that the C given by (21) satisfies the functional equation. Since $\ell C(\ell, z) < A(\ell, z)$, Corollary 1 and the equivalence between (DP^D) and (DP) imply that the C given by (21) is C^*.

The tractability under the power function ϕ allows us to see the intuition behind $\alpha = 1/2$. Observe that there are two conflicting forces in choosing α. If one simply builds a z'-city without further building cities smaller than z', then the optimal choice is $\alpha = 1/2$. This is obvious upon observing (3) and noting that the savings in transport costs reaches its maximum when $\alpha = 1/2$. However, Lemma 1 also says that we must build smaller cities in each of the two intervals split by the z'-city. To maximize the total savings in the two intervals from building these smaller cities, it is optimal to split the intervals unevenly, i.e., to choose an α relatively close to 0 or 1. This is characterized in equation (19) in the term $\left[\alpha^{\frac{2(b+1)}{b}} + (1 - \alpha)^{\frac{2(b+1)}{b}} \right] \tilde{S}(\ell)$. Obviously, the bracket term reaches its maximum at 0 or 1. This means that the benefit of having one large interval so that more savings can be attained outweighs the loss of having a smaller one. The two conflicting forces are nicely summarized in (20), the first in the numerator and the second in the denominator. Note that the two conflicting forces operate on different levels of $y \in [0, z']$, and that how ϕ changes across y is critical. The first force, by having α closer to $1/2$, increases z' and saves on these higher levels of goods, whereas the second force is concerned with savings on goods relatively close to 0. Imagine two ϕ functions which take the same value at z with one ϕ diminishing faster than the other when y moves down from z. For the ϕ function that $\phi(y)$ diminishes quickly when y goes down, like the power function, the first force dominates the second because there are more savings from the higher level goods as the setup costs for these goods are lower than the case in which $\phi(y)$ diminishes slowly. This intuition is consistent with the example in Section 2.4 where $\alpha \neq 1/2$. In that example, the smooth approximation of a non-smooth ϕ means that ϕ does not diminish very quickly for $y \in [0, e]$ until it gets very close to 0. In other words, if ϕ diminishes at a fast enough rate, the first force dominates and $\alpha = 1/2$ becomes the optimal outcome.

3.5 Optimal distance between largest cities

In this subsection, we characterize the optimal distance between largest cities conditional on the central place property. We use the superscript * to denote an optimal solution. The optimal distance between two neighboring z_1-cities is the social planner's first stage problem:

$$\ell_1^* = \arg \min_{\ell_1 \in \left(0, \bar{\ell}(z_1)\right)} C^*(\ell_1, z_1). \tag{22}$$

When ℓ_1 increases to $\bar{\ell}(z_1)$, $z_2^* = z_1$, and the distance between two neighboring z_1-cities drops to $\bar{\ell}(z_1)/2$. Hence, to look for the optimal distance between two neighboring z_1-cities, we need only focus on $\ell_1 \in (0, \bar{\ell}(z_1))$. The next proposition guarantees the existence of ℓ_1^* and provides a sharper lower bound for ℓ_1^*, which is useful for the welfare analysis conducted in Section 4.

Proposition 6. *Suppose that the central place property holds in an optimal solution. Then, there exists an $\ell_1^* \in (0, \bar{\ell}(z_1))$ that solves (22). That is, an optimal distance between two neighboring z_1-cities exists. Moreover, $\ell_1^* \in [\bar{\ell}(z_1)/2, \bar{\ell}(z_1))$.*

Proof. See the separate appendix. ■

The intuition is that (22) can be viewed as a minimization of a continuous function C^* on a compact set $[\delta, \bar{\ell}(z_1)]$ for some small $\delta > 0$. It is intuitive that for any given z_1, $\lim_{\ell_1 \to 0} C^*(\ell_1, z_1) = \infty$ because the per capita setup cost explodes when $\ell_1 \to 0$, and hence $(0, \delta)$ for some δ are excluded as possible minimizers. To search for optimal ℓ_1, we can include $\bar{\ell}(z_1)$ as candidate because if it attains the minimum, then we set $\ell_1^* = \bar{\ell}(z_1)/2$. To see why $\ell_1^* \geq \bar{\ell}(z_1)/2$, suppose that $\ell_1^* < \bar{\ell}(z_1)/2$, i.e., the optimal distance between the two largest cities is less than half the distance that entails $z_2^* = z_1$. Then, at $\ell_1 = 2\ell_1^*$, it is optimal to set $z_2^* = z_1$ so that the effective distance between two neighboring z_1-cities is exactly ℓ_1^*, which by definition minimizes the per capita cost C^*. However, this contradicts the fact that the optimal z_2^* at ℓ_1 must be less than z_1 since $\ell_1 = 2\ell_1^* < \bar{\ell}(z_1)$.

In general, ℓ_1^* need not be unique, although the uniqueness of ℓ_1^* in the case of a power function ϕ can be verified by studying the shape of C given in (21). For any optimal solution ℓ_1^*, it is optimal to let all z_1-cities be evenly spaced with a distance of ℓ_1^*, although z_1-cities do not have to be evenly spaced when there are multiple solutions to (22). Figure 4 shows how $C(\ell_1, z_1)$ depends on $\ell_1 < \bar{\ell}(z_1)$ when ϕ is a power function.

Next, we utilize an envelope argument to illustrate the economics behind the determination of ℓ_1^*. Of course, the underlying economics do not depend on the differentiability of C^* in ℓ_1, but it is convenient to proceed assuming such differentiability.[14] Observe that from (DP),

$$\frac{\partial C(\ell, z)}{\partial \ell} = \frac{t(z - z'^*)}{4} - \frac{\Phi(z) - \Phi(z'^*)}{\ell^2},$$

where z'^* is the optimal size of the immediate sub-city, given (ℓ, z). The optimal choice of

---

[14] The differentiability of C^* in z_1 is guaranteed if $\ell_1 < \bar{\ell}(z_1)$ so that optimal $z_2 < z_1$. However, the differentiability of C^* in ℓ_1 is not obvious. For example, Lucas and Stokey (1989) prove the differentiability of a value function by the concavity of the function. Here, such an argument does not work since C^* is not necessarily convex (we are looking at a minimization problem, whereas a maximization problem is examined in Lucas and Stokey). As Figure 4 shows, C^* in ℓ_1 is neither convex nor concave.
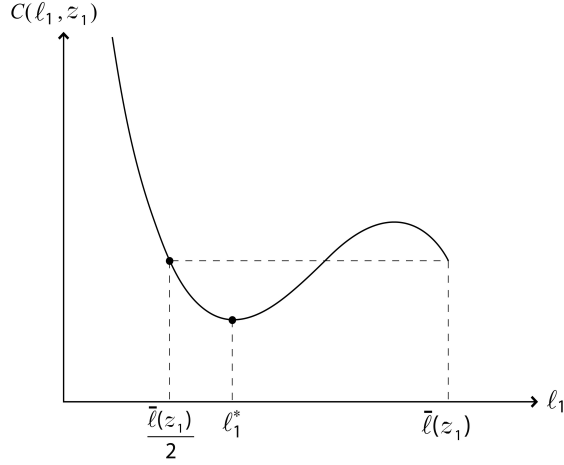
Figure 4: How cost C depends on the distance ℓ_1 between largest cities.

ℓ_1 satisfies the necessary condition: $\partial C(\ell_1, z_1)/\partial \ell_1 = 0$. Hence,

$$\frac{t(z_1 - z_2^*)\ell_1^*}{4} = \frac{\Phi(z_1) - \Phi(z_2^*)}{\ell_1^*}.$$

The above equation means that the choice of ℓ_1^* is such that the per capita transport cost for the goods $(z_2^*, z_1]$ equals the per capita setup costs for this range of goods. As the per capita transport/setup cost increases/decreases in ℓ_1 with z_2^* fixed at an optimal level, the scale economies of having a larger ℓ_1 to share the setup costs, at ℓ_1^*, should be exactly offset by the additional transport costs.

4   Welfare Analysis and Extension to the Plane

The environment in Hsu's (2012) model is essentially the same as that in this paper. Thus, we can compare the equilibrium allocation in his model with our optimal solution. In this section, we first introduce Hsu's equilibrium setting and the main result. Then, we determine whether an optimal solution can be decentralized, and, if there is a discrepancy between an equilibrium and the optimal allocation, what the pattern of deviation is in terms of entry. Note that as there are infinitely many production sites for each good in both models, we measure entry for each good as the number of production sites per unit distance. Hence, the smaller the distance between production sites, the larger the entry. In this section, we also briefly discuss how Lemmas 1 and 2 apply to a two dimensional space.

4.1 Hierarchy equilibrium in Hsu (2012)

The environment in Hsu (2012) is the same as that in Section 2.1, except that the hierarchy property is not imposed. As an equilibrium model, Hsu (2012) must specify the interactions of agents through which an equilibrium arises. The related details are as follows.

4.1.1 Firm entry and one-good equilibrium

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).

1. Entry and location stage
   The potential firms simultaneously decide whether to enter. Upon entering, each entrant chooses a location and pays the setup cost for the good it produces. Assume the tie-breaking rule: if a potential firm sees a zero-profit opportunity, then it enters.

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

Consider the subgame perfect Nash equilibrium (SPNE) of any particular good z. Consider two neighboring 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 $MC_A = \gamma + tx$ and $MC_B = \gamma + t(\ell - x)$. 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 $[0, \ell]$ can be written as

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

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 $t\ell^2/4$. Consider any entrant's strategy in the first stage. Let this entrant be named C. If C were to enter into the interval between A and B, then it is straightforward that C's profit-maximizing location will 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 nonnegative.

Therefore, firms must be an equal distance apart, and the gross profit of any firm with a market area of ℓ is $t\ell^2/2$. Now, for any z, define $\underline{\ell}(z)$ as the solution to the zero-profit condition $t\left[\underline{\ell}(z)\right]^2/2 = \phi(z)$. Thus, $\underline{\ell}(z) = \sqrt{2\phi(z)/t}$. The foregoing derivation of an SPNE for an arbitrary good implies that there is a continuum of equilibria in which one firm is located at every point in $\{x + n\ell\}_{n=-\infty}^{\infty}$, where $\ell \in [\underline{\ell}(z), 2\underline{\ell}(z))$ and $x \in [0, \underline{\ell}(z))$. The continuum of equilibria exists because any distance ℓ in the interval $[\underline{\ell}(z), 2\underline{\ell}(z))$ is an equilibrium distance; $\ell \geq \underline{\ell}(z)$ implies that all firms earn a nonnegative profit (no exit), whereas $\ell/2 < \underline{\ell}(z)$ implies that any new entrant between any two existing firms must earn a negative profit (no entry).

4.1.2 Hierarchy equilibrium and central place property

An equilibrium is a collection of firm locations, delivered price schedules, and farmers' consumption choices such that the allocation for each good is an SPNE. A *hierarchy equilibrium* is an equilibrium in which, at any production location, the set of goods produced takes the form $[0, z]$ for some $z \in (0, z_1]$. In a hierarchy equilibrium, there exists a decreasing sequence $z_1 > z_2 > \ldots > z_i > \ldots$ such that any production location is z_i-city for some z_i. Obviously, a hierarchy equilibrium satisfies the central place property if the market area of the firms producing $(z_{i+1}, z_i]$ is half that of the firms producing $(z_i, z_{i-1}]$.

Proposition 1 in Hsu (2012) states that some equilibria have the hierarchy property and that all such hierarchy equilibria satisfy the central place property. Hence, every hierarchy equilibrium is a central place hierarchy characterized as follows. Fix an $x \in \mathbb{R}$ and set the grid for $(z_{i+1}, z_i]$ as $\{x + n\ell_i\}_{n=-\infty}^{\infty}$, where $\ell_1 \in [\underline{\ell}(z_1), 2\underline{\ell}(z_1))$, $\ell_i = \ell_1/2^{i-1}$, and the cutoff z_i is given by the zero-profit condition

$$\phi(z_i) = \frac{t\ell_i^2}{2} \qquad \text{for all } i \geq 2. \tag{23}$$

Without loss of generality, let $x = 0$. Then, the location configuration so constructed, which obviously satisfies both the hierarchy and the central place property, is precisely that given in Figure 1, except that only four layers are depicted in the figure.

The characterization above indicates that there is a continuum of central place hierarchy equilibria, each of which is characterized by an $\ell_1 \in [\underline{\ell}(z_1), 2\underline{\ell}(z_1))$. Note that, depending on ϕ, an optimal hierarchy may or may not have the central place property, as shown by Proposition 5 and the example in Section 2.4. In contrast, if an overall equilibrium is a hierarchy, it will have the central place property, and such equilibria exist for any ϕ.

4.2 Decentralization?

In this and the next subsection, we compare hierarchy equilibria, i.e, those equilibria that are hierarchies (and hence have the central place property by Proposition 1 of Hsu 2012), with optimal hierarchies that satisfy the central place property. As the superscript $*$ denotes an optimal allocation, we use the superscript e to denote the allocation in a hierarchy equilibrium. For $\ell_1^* \in \left(0, \bar{\ell}(z_1)\right)$, $z_2^* < z_1$. Since $\alpha = 1/2$, $\ell_{i,k}^*$ is the same across k, and we denote $\ell_i^* \equiv \ell_{i,k}^*$ for all $i \geq 2$. Obviously, $\ell_i^* = \ell_1^*/2^{i-1}$. Using (4) recursively, for $i \geq 2$,

$$\phi\left(z_i^*\right) = \frac{t\left(\ell_{i-1}^*\right)^2}{8} = \frac{t\left(\ell_i^*\right)^2}{2}. \tag{24}$$

By comparing (23) and (24), we see that the $\{z_i^*\}$ and $\{z_i^e\}$ sequences would be identical if and only if $\ell_1^* = \ell_1^e$.[15] Hence, the optimal allocation can be decentralized if and only if $\ell_1^* \in [\underline{\ell}(z_1), 2\underline{\ell}(z_1))$; that is, the optimal distance between two neighboring z_1-cities is within the continuum of the distance between two neighboring z_1-cities in a hierarchy equilibrium.

Corollary 3. *Suppose that the central place property holds in an optimal solution. Then, $\ell_1^* \in [\underline{\ell}(z_1), 2\underline{\ell}(z_1))$. That is, the optimal solution can be decentralized.*

Proof. Observe that by combining (16), (22), and (23), we have

$$\frac{t\left(\ell_1^*\right)^2}{8} < \frac{t\bar{\ell}(z_1)^2}{8} = \phi(z_1) = \frac{t\underline{\ell}(z_1)^2}{2},$$

which implies that $\bar{\ell}(z_1) = 2\underline{\ell}(z_1)$. The results follow from Proposition 6. ∎

4.3 Entry comparison

As there is a continuum of hierarchy equilibria, some may be suboptimal. For example, with a power function ϕ, it can be verified that ℓ_1^* is unique (see Figure 4), and thus there must be suboptimal hierarchy equilibria. Here, we compare the entry for goods between a hierarchy-equilibrium allocation and an optimal one. In a spatial competition model,

---

[15]To see the intuition for why the two sequences are identical, recall that the gross profit of a firm with market area ℓ is $t\ell^2/2$. The gross profit of an entrant that enters at the middle of a market area of ℓ is thus $t\ell^2/8$, which is exactly the savings in transport costs created by the entrant locating in the middle. Thus, the zero-profit condition (23) is the same as (24). Note that under Bertrand competition, gross profits are created from the difference between two competing firms' delivery costs. If an entrant were not to enter, then all the potential gross profits of an entrant would be part of the transport costs incurred by some incumbents. In other words, the benefits to consumers brought by an entrant, i.e., the savings in transport costs, equal the entrant's own benefits.
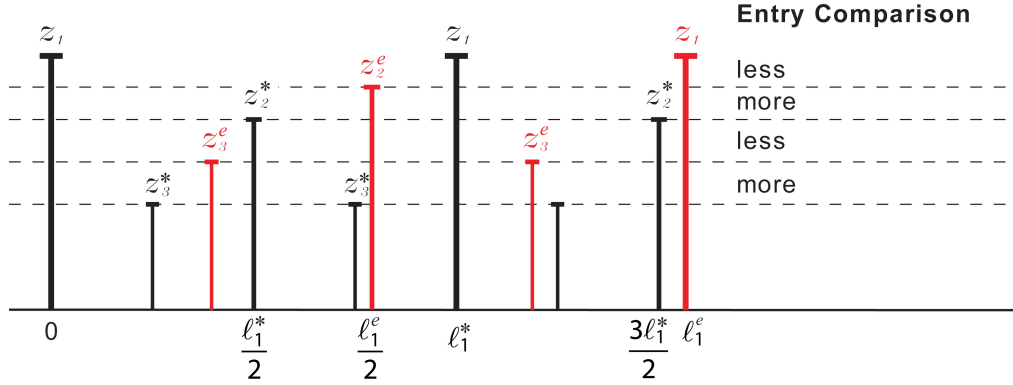
Figure 5: An equilibrium allocation can alternately have greater or smaller entry than the optimal allocation. Depicted is the case of $\ell_1^e > \ell_1^*$.

Salop (1979), using a one-good model, shows that there is always greater equilibrium entry than what is optimal. We show that whenever $\ell_1^e \neq \ell_1^*$, the directions of deviation for different goods are different, in contrast with Salop's result. More specifically, the directions of deviation alternate across sets of goods. That is, the whole range of goods can be partitioned into sets such that the first set has an equilibrium entry that is less (more) than that in the optimal solution, the second set has one that is more (less), the third set less (more), and so on. Figure 5 illustrates such a case where $\ell_1^e > \ell_1^*$. The words "more/less" in the figure mean that the equilibrium entry is more/less than the optimal one.

Proposition 7. *With the central place property, the following hold.*

1. *If $\ell_1^e = \ell_1^*$, then entry for each good is identical in both the equilibrium and the optimal solution.*

2. *If $\ell_1^e > \ell_1^*$, then*

   (a) *$z_{i+1}^* < z_{i+1}^e \leq z_i^*$ for all $i \geq 1$. The $[0, z_1]$ continuum can be partitioned into sets of the form $(z_{i+1}^e, z_i^*]$ and $(z_{i+1}^*, z_{i+1}^e]$, for all $i \in \mathbb{N}$.*

   (b) *For all $y \in (z_{i+1}^*, z_{i+1}^e]$, equilibrium entry is weakly more than the optimal one.*

   (c) *For all $y \in (z_{i+1}^e, z_i^*]$, equilibrium entry is less than the optimal one.*

3. *If $\ell_1^e < \ell_1^*$, then the result in (b) holds with the superscripts of * and e exchanged.*
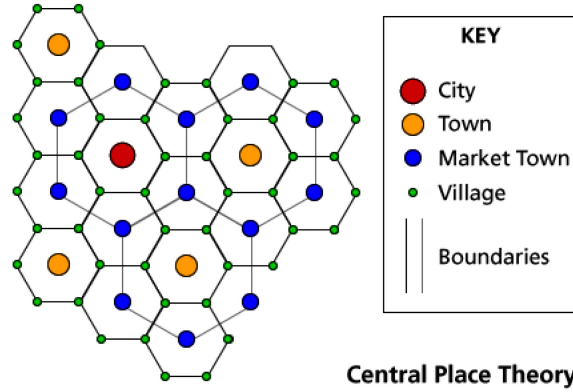
Proof. See the separate appendix. ■

Figure 6: Central Place Hierarchy on the Plane (Christaller, 1933).

4.4 Extension to the plane

Suppose that the geographic space is the infinite plane, instead of the real line. One difficulty with two-dimensional space is that for a polygon formed by a set of neighboring cities at the vertices, the market area of smaller cities inside the polygon may actually extend outside the polygon, and some cities outside the polygon can have their market areas inside it. It is thus unclear how the sequence and dynamic programming problems should be formulated. Even when they are formulated, it is conceivably difficult to generalize Corollary 2 because while there is just a one-parameter family of intervals in one-dimensional space, there is an intractible infinitude of two-dimensional regions. Nevertheless, it is interesting to observe that Lemmas 1 and 2 still hold on the plane.[16]

Lemma 1' It is never optimal to have an area without any city in it. That is, cities are dense.

Proof. Suppose the contrary is true. That is, suppose that there is a polygon formed by cities that produce at least up to $z \in (0, z_1]$ at the vertices, and that there is no city inside the polygon. Consider having a z'-city inside the polygon with $z' \in (0, z]$. Given the location of the z'-city, the savings in transport costs for each good are fixed at some number $s > 0$, and the total savings for all goods $[0, z']$ equals sz'. Then, for each z' such that $\phi(z') \leq s$, the setup cost of the z'-city is $\Phi(z') \leq sz'$. Because ϕ is strictly increasing and continuous with $\phi(0) = 0$, such z' must exist, and hence there must be positive net savings. Thus, a polygon without any smaller cities in it is never optimal. ∎

---

[16]It is also interesting to note that if the plane is divided according to Christaller (1933) as in Figure 6, then with optimal z_i^* given by $\phi(z_i^*)$ equaling the savings in transport costs for the good z_i^*, it is a straightforward exercise to follow Hsu (2012) to verify that the city size distribution follows the power law. In fact, the power law result also holds for regular square and triangular regions.

Lemma 2′ It is never optimal to have two cities of the same size z' with a point closer to both of them than to any other city with $z \geq z'$, where $z \leq z_1$.

Proof. Observe that Lemma 2 holds precisely because the savings in transport costs of increasing z' at two z'-cities simultaneously are less than the sum of the savings of increasing z' at each z'-city separately. This is because there is some point (consumer) in the interval that benefits from increasing z' at either city because it is closer to both of the z'-cities than it is to any of the cities producing at least $z \geq z'$. The same logic applies to two-dimensional space here. ∎

5 Conclusions

This paper presents an analysis of optimal city hierarchy. The model formalizes central place theory via an efficiency rationale. It takes the hierarchy property as given and provides the conditions under which the central place property is optimal. In this sense, this paper complements Quinzii and Thisse (1990), who model the hierarchy property. It remains to be seen whether the optimality of both properties can be obtained in one concise model.

As mentioned, as long as there are immobile consumers, regardless of their fraction in the economy, spreading across the entire geographic space, there exists the need for cities and towns to spread out to serve these consumers. From this perspective, central place theory is still very relevant to the modern-day economy, even though it may be true that as the economy becomes more industrialized, the location patterns may become more biased compared with the "ideal" central place pattern. The uniform distribution of consumers is, of course, a rough approximation of real spatial distribution. Nevertheless, it is useful to clarify in theory what happens under the uniform distribution to take advantage of its tractability. One take-home message is that even under uniform distribution, the optimal locational pattern does not necessarily conform to the central place pattern. As Propositions 3 and 4 can in principle be applied in the more general environments of spatial problems, simulations of what would happen under a more realistic distribution may be a desirable direction for future research.

We conclude by summarizing two methodological messages. First, there are benefits to using dynamic programming to study spatial problems, as demonstrated by Lucas and Rossi-Hansberg (2002) on internal city structure and this paper on city hierarchy. Second, the techniques developed in this paper may also be useful for various hierarchical problems (not limited spatial problems), as Propositions 1 to 4 are all applicable in a general

setting with recursive divisions.

References

Beckmann, M. J. (1958), "City hierarchies and the distribution of city size," *Economic Development and Cultural Change*, 6(3): 243-248.

Beckmann, M. J. (1970), "City size distribution in a central place hierarchy: An alternative approach," *Journal of Regional Science*, 10(1): 25-33.

Berliant, M. (2008), "Central place theory," dictionary entry in *The New Palgrave Dictionary of Economics*, 2nd Edition, Palgrave Macmillan.

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

Córdoba, J.-C. (2008), "On the distribution of city sizes," *Journal of Urban Economics,* 63: 177-197.

Duranton, G. (2006), "Some foundations for Zipf's law: Product proliferation and local spillovers," *Regional Science and Urban Economics,* 36(4): 543-563.

Duranton, G. (2007), "Urban evolutions: The fast, the slow, and the still," *American Economic Review,* 97 (1): 197-221.

Eaton, B. C. and R. G. Lipsey (1982), "An economic theory of central places," *The Economic Journal,* 92(365): 56-72.

Eeckhout, J. (2004), "Gibrat's law for (all) cities." *American Economic Review,* 94(5): 1429-1451.

Fujita, M., P. Krugman, and T. Mori (1999), "On the evolution of hierarchical urban systems," *European Economic Review,* 43: 209-251.

Fujita, M., P. Krugman, and A. J. Venables (1999), *The Spatial Economy: Cities, Regions, and International Trade,* The MIT Press.

Gabaix, X. (1999) "Zipf's law for cities: An explanation," *The Quarterly Journal of Economics,* 114(3): 739-767.

Henderson, J. V. (1974), "The types and size of cities," *American Economic Review,* 64: 640-656.

Garicano, L. (2000), "Hierarchies and the organization of knowledge in production," *Journal of Political Economy*, 108(5): 874-904.

Garicano, L. and E. Rossi-Hansberg (2006), "Organization and inequality in a knowledge economy," *Quarterly Journal of Economics*, 121(4): 1383-1435.

Hsu, W.-T. (2012), "Central place theory and city size distribution," forthcoming, *Economic Journal*.

King, L. J. (1984), *Central Place Theory*, Beverly Hills, CA: Sage Publications.

Lederer, P. J. and A. P. Jr. Hurter (1986), "Competition of firms: Discriminatory pricing and location," *Econometrica*, 54 (3): 623-640.

Lösch, A. (1940), *The Economics of Location*, translated by W. Woglom (1956), New Haven, CT: Yale University Press.

Lucas, R. E. Jr. and E. Rossi-Hansberg (2002), "On the internal structure of cities," *Econometrica*, 70 (4): 1445-1476.

Lucas, R. E. Jr. and N. Stokey, with E. Prescott (1989), *Recursive Methods in Economic Dynamics*, Cambridge, MA: Harvard University Press.

Mirrlees, J. A. (1972), "The optimum town," *Swedish Journal of Economics*, 74(1): 114-135.

Mori, T., K. Nishikimi and T. E. Smith (2008), "The Number-Average Size Rule: A new empirical relationship between industrial location and city size," *Journal of Regional Science*, 48: 165-211.

Mori, T. and T. E. Smith (2011), "An industrial agglomeration approach to central place and city size regularities," *Journal of Regional Science*, 51(4), pp. 694-731.

Qian, Y. (1994), "Incentives and loss of control in an optimal hierarchy," *Review of Economic Studies*, 61(3): 527-544.

Quinzii, M. and J.-F. Thisse (1990), "On the optimality of central places," *Econometrica*, 58(5): 1101-1119.

Rossi-Hansberg, E., and M. L. J. Wright (2007), "Urban structure and growth," *Review of Economic Studies*, 74 (2): 597-624.

Rozenfeld H., D. Rybski, X. Gabaix, and H. Makse (2011), "The area and population of cities: New insights from a different perspective on cities," *American Economic Review*, 101(5): 2205-25.

Salop, S. C. (1979), "Monopolistic competition with outside goods," *The Bell Journal of Economics*, 10(1): 141-156.

Simon, H. (1955), "On a class of skew distribution functions," Biometrika, 44:425-440, reprinted in: (1957) *Models of Man: Social and Rational. Mathematical Essays on Rational Human Behavior in a Social Setting*, New York: Wiley and Sons.

Tabuchi T. and Thisse, J.-F. (2006). "Regional specialization, urban hierarchy, and commuting costs," *International Economic Review*, 47(4): 1295-1317.

Tabuchi T. and Thisse, J.-F. (2011). "A new economic geography model of central places", *Journal of Urban Economics*, 69: 240-52.