Urban growth and subcenter formation: A trolley ride from the Staples Center to Disneyland and the Rose Bowl

Marcus Berliant [a,*], Ping Wang [a,b]

[a] *Washington University in St. Louis, St. Louis, MO, USA*
[b] *NBER, USA*

Abstract

The long-term trends of urbanization suggest: not only have more cities formed, but the leading metropolises have grown larger, with a number of peripheral subcenters developing over time. Conventional models of urban growth are limited, in that commuting cost and congestion eventually result in decreasing returns in a monocentric city as population becomes very large. We construct a general-equilibrium model with dynamic interactions between spatial agglomeration and urban development, driven by location-dependent knowledge spillovers. Our contribution allows endogenous development of subcenters to capture benefits from knowledge spillovers and offset diminishing returns from urban congestion, thus permitting more sustained city growth.
© 2007 Elsevier Inc. All rights reserved.

1. Introduction

There have been long-term trends of urbanization and city growth across developed and developing countries over the past two centuries. Not only have more cities formed, but the leading metropolises have grown larger over time. Conventional endogenous urban growth models characterize this long-term trend in a canonical infinite-horizon framework with a monocentric city structure.[1] One may inquire: given limited land or increasing total commuting cost in a monocentric city with localization externalities and increasing consumer population, how can we explain continual growth of city populations? After all, one would expect diminishing returns to set in eventually with higher population. Our answer to this question is the formation of satellite subcenters. The primary purpose of this paper is to propose and solve a model with the following features:

- There is an externality in capital usage that affects production and depends on both the overall geographical distribution of firms within a city as well as a particular firm's location in the distribution.
- The model is purely competitive. There are no developers, and cooperative behavior on the part of firms is excluded.

---

* Corresponding author at: Department of Economics, Washington University, St. Louis, MO 63130-4899, USA. Fax: +1 314 935 4156.
 E-mail address: berliant@artsci.wustl.edu (M. Berliant).

[1] See Berliant and Wang (2004) for a detailed review of exogenous and endogenous urban growth models.

- The model is dynamic in the sense that population grows, whereas agents have perfect foresight. This, together with full depreciation of capital, boils down to period-by-period optimization of agents with population change.
- The model is analytically solvable.
- The equilibrium of the model features endogenous subcenter formation as population grows.
- The firms do not own land, so it is unnecessary to reward them with increased land rents as a result of capitalization of the knowledge externality when they form a new subcenter.

A better understanding of the emergence of subcenters is useful for developing transportation, local industrial and urban land use policy. Moreover, it is important to understand the role of geography in economic growth, and vice versa.

Many large metropolises feature a core central city together with a number of peripheral subcenters. For example, McMillen and Smith (2003) provide a careful empirical documentation of such urban structures. They find that most US cities have multiple subcenters, with as many as 46 subcenters in Los Angeles.[2] Although there have been attempts at modeling this core-periphery urban structure in a static set up without population growth (see Abdel-Rahman and Wang, 1995, and papers cited therein), a complete characterization of the competitive formation of subcenters within an endogenous urban boundary framework remains unexplored. Thus, our paper also contributes to the conventional core-periphery literature.

The modeling strategy is as follows. We construct a competitive spatial equilibrium model with households and firms residing in a linear city with continual exogenous growth of population and expansion of city boundaries. Following the pivotal work by Ogawa and Fujita (1980) and Fujita and Ogawa (1982), we incorporate positive inter-firm production externalities that decrease with distance between the firms. Using the structure proposed in the fixed population, closed city setup by Berliant et al. (2002), we measure this geographically diminishing effect by both Euclidean distance to the mean of the distribution of firms and an overall firm dispersion index. In addition to using land, firms employ labor and capital to produce a single homogeneous good, maximizing profits. Households choose

workplaces and residences to maximize utility. Capital is fully depreciated, so the choice variables of agents (all with perfect foresight) have no effect on state variables, which greatly simplifies the analysis. In competitive equilibrium, the allocation of goods and factors, the locational choice of firm sites and household residences, and factor prices and land rents are all endogenously determined.

The results derived for this model are as follows. When commuting cost for additional population exceeds production benefits from additional population, one would expect to see urban growth subsiding. Given the costs and benefits, we could need coordination between producers and consumers to achieve this (localization benefits for producers and commuting cost for consumers is the trade-off). But with a wage gradient, we might not—the producers could be making the decision about a new subcenter in a decentralized way, since they are paying the commuting cost indirectly through the wage. So exogenous population growth, an endogenous city boundary, a simplified localization externality, commuting costs, and a wage gradient are the key ingredients of the model. The idea is that as population expands, the city boundary moves out (given consumer demand for land is inelastic), thus commuting cost gets large, so land rent and the wage are relatively small for locations distant from firms. Eventually, some firms will decide to move into a gap. Formation of a subcenter can be better than the formation of a new city from the point of view of production efficiency, because of the presence of localization externalities. Formally, we prove the existence of symmetric competitive spatial equilibrium, establishing necessary and sufficient conditions to support either a monocentric or a core-periphery urban configuration and deriving analytically the corresponding equilibrium wage and land rent schedules. To us, the most surprising aspect of this project is that one does not have to assume that the firms own shares in land rent in order to induce them to form subcenters endogenously in equilibrium. They will do so willingly, in a noncooperative fashion, without enjoying the benefits of land value capitalization around them. There is no government or developer in the model who can see rents and coordinate agents to take action. The absentee landlord, completely passive in the model, is the only agent that benefits from higher land rents. In the dynamic process of urban development in response to exogenous population growth, eventually the monocentric equilibrium disappears, but the equilibrium with a core-periphery urban configuration appears and survives for a while. This latter result may serve to explain why the monocentric urban structure is rarely seen today in US cities.

---

[2] Other large cities, such as New York, Chicago and Washington, DC, have 10–38 subcenters. St. Louis, where the authors reside, has 5 subcenters. Nashville and Salt Lake City are among the very few cities featuring a monocentric urban structure with no subcenters.

1.1. Related literature

Our work builds on Berliant et al. (2002) in two important respects. First, Berliant et al. is static, whereas the current paper is dynamic. In particular, we allow population to *grow* (at a constant rate) and the city boundary to *expand* accordingly to accommodate the expanding population (at the same rate under the linear city setup). Second, while Berliant et al. examine the possibilities of monocentric, mixed and duocentric urban structures, the current paper considers the formation of a *core-periphery* tricentric urban configuration, that has been largely ignored in the theoretical literature on city formation. In contrast with both our previous work and most of the previous literature, we consider *asymmetric* urban structures where centers and subcenters do not have the same number of firms and do not employ the same number of workers.

Our paper is obviously related to the subcenter formation literature, specifically studies by Henderson and Slade (1993), Henderson and Mitra (1996), and Fujita et al. (1997).[3] In Henderson and Slade, it is the developers who try to form competing subcenters. In Henderson and Mitra and Fujita et al. subcenters are formed in initially peripheral locations due to the establishment of new plants set up by a large (non-competitive) firm that has a dominant employment share.[4] We wish to emphasize, however, that, in contrast with these papers, we use a *competitive* framework. In addition to this key difference in modeling strategies concerning subcenter formation, the firms in our paper do *not* capture the benefits of location capitalized into land rents (particularly when a subcenter forms), whereas such benefits are necessary in the other papers to obtain subcenter formation.

There are several important points to be made in relating our model to the broader literature. First, the Fujita and Ogawa (1982) framework, that was at least 20 years ahead of its time, employs a form of production externality for each firm that relies on the pairwise distance between a given firm and any other firm. Using numerical simulations, they find equilibria with multiple symmetric centers (each with the same number of firms) and conclude that both population and commuting cost play important roles in determining the number of subcenters in equilibrium, though equilibrium is far from unique. In contrast, our model has centers with different numbers of firms, is close to having an *analytical solution* (we employ numerical simulations at the very end to extract a characterization of equilibrium), and takes a more conservative view of the production externality, biased away from subcenter formation. That is, our externality function measures the distance from the mean location of firms (generally the middle of the set of locations) to any particular firm, and the overall dispersion of firms. Thus, subcenters are less likely to form, since the externality gives no weight to the local density of firms. Nevertheless, we obtain endogenous subcenter formation without developers or government. Naturally, due to the production externality, our model yields equilibrium allocations that are not first best.

Relative to the literature on the new economic geography (NEG), as exposited by Krugman (1993), Fujita and Krugman (1995) and Fujita and Mori (1997), our model of course relies on an explicit production externality rather than a pecuniary one with imperfect competition.[5] Naturally, their extended comparative statics exercises are similar in spirit to ours, as is the result obtaining an endogenous core-periphery structure; they must work harder to get this result, often resorting to numerical simulations, due partly to the large number of equilibria in their model. Generally speaking, the NEG model is one of regions rather than the internal structure of a city. No doubt externalities are stronger for firms within the same city than for firms in different cities but in the same region; Rosenthal and Strange (2003) find that production externalities are in reality quite localized. Moreover, our framework can be used to generate subcenters of *different sizes*, in contrast with the population symmetry that is ubiquitous in NEG models.[6] Our model allows analysis of the relationship between the emergence of subcenters and the interactions among firms via knowledge spillovers. Furthermore, in contrast

---

[3] The reader is referred to a survey paper by Anas et al. (1998) that contains a good discussion of subcenters from both the theoretical and empirical points of view.

[4] Both papers use the Fujita (1989) open city framework. While the former uses simulations, the latter produces analytical solutions.

[5] See Baldwin and Forslid (2000) and Fujita and Thisse (2002) for more recent applications of the NEG model. One may relate our study of subcenter formation to NEG models of city formation in a system of cities. In these models, each city is specialized in producing a fixed and limited set of varieties (possibly one) and each firm is specialized in producing one variety. Thus, *by assumption*, as the number of firms grows, the number of cities also grows (see, e.g., Anas and Xiong, 2003). Moreover, as in most variety models, there is a scale effect in the sense that a larger economy with larger employment is associated with a greater number of varieties. So, *by construction*, an increase in population also leads to more cities (see, e.g., Anas, 2004).

[6] For example, the spatial configuration in Anas (2004) is a system of identical cities with the number and size endogenously determined. The configurations in Anas and Xiong (2003) are monocentric, duocentric and multicentric with cities of identical size.

with most of the NEG models, we do not rely exclusively on numerical simulations.

Finally, Lucas and Rossi-Hansberg (2002) is related to our work in terms of the questions asked, but we use a different type of externality (both in functional form and by using capital instead of labor) as well as a different closure of the model, namely we have exogenous population and an endogenous border while their model is the other way around. To be more specific, in our model we have both capital and labor while their model has only labor. We simplify the land market relative to their model by assuming that land demand is inelastic for all agents. Our externality is based on *capital usage* by firms, whereas theirs is based on labor usage. Their externality accounts for pairwise labor usage and pairwise distance between firms, whereas ours relies on distance to the mean location of firms and the distribution of capital use. We use assumptions on *primitives*, whereas they use some assumptions on objects that are not primitive. We solve our model analytically whereas they solve theirs computationally. Their model is static whereas ours is at least *partially dynamic*. Finally, we take population as exogenous and derive the equilibrium utility level of consumers *endogenously*. They take the utility level as exogenous and allow population to migrate in or out of the city. This last difference is important, because in urban models typically indirect utility is not a monotonic function of population. In general, it has at least one peak (usually called a "utility hill") and could have multiple peaks. So there is usually not a unique equilibrium level of population corresponding to an equilibrium utility level. In many other respects, the contrast between their work and ours is similar to the contrast between Fujita and Ogawa (1982) and ours. Generally speaking, equilibrium in all of these models is second best, due to externalities or other forms of market failure.

2. The basic model

Both time and location are continuous; they are indexed by t and z, respectively. The spatial economy features a long, narrow metropolitan area (MSA) over a featureless line $\Omega \equiv [-B, B]$, with uniformly distributed land. The borders, represented by the variable B, are *endogenously determined*. We will use agricultural land rent to determine the city boundary and thus to close the model. This MSA boundary variable B will increase with the population of the MSA, but that will be a result, not an assumption. We restrict our attention to the case of only one core in the spatial economy. Without loss of generality, we let location $z = 0$ indicate the pre-

determined central business district (CBD) of the core prior to the development of satellite subcenters at the fringes of the MSA. One may imagine location $z = 0$ as downtown Los Angeles. The western boundary of the city is limited by the Pacific Ocean, the eastern boundary by mountains. One can visualize locations of satellite subcenters such as Anaheim and Pasadena when the border of the Los Angeles MSA stretched out as its population grew. Limits of such growth to the east were initially given by the coastal mountain range, though recent development has occurred in the range and beyond.

Index all of the variables by time t, denoted by $E(t)$ for any variable E. There is a continuum of infinitely lived firms of mass M and a continuum of infinitely lived households of mass $N > 0$. For simplicity, we assume that the mass of firms is fixed (no entry) and that the mass of households grows at an exogenous constant rate $\nu > 0$:[7]

$$N(t) = N(0)e^{\nu t} \tag{1}$$

where the initial population is normalized to unity: $N(0) = 1$.

There is an absentee landlord who owns all the land (of measure $2B$), collects the total land rent (TLR), and consumes no land. As the population grows, the city boundary will expand. As we will show below, each household or firm occupies a unit density of land in equilibrium. As a consequence, the absence of vacant land yields the following identity:

$$B(t) = \frac{M + N(t)}{2}. \tag{2}$$

Each firm rents land from the landlord, rents capital (K) from households and employs labor (L) to produce a single homogeneous final good (Y) using a Cobb–Douglas technology. The output of final goods can be distributed to workers (wages), capital owners (capital rentals) and land owners (land rents). Each household supplies a unit density of labor endowment inelastically and chooses both workplace and residence to maximize lifetime utility. The absentee landlord's behavior is passive, simply spending the given total land rents, TLR, for final good consumption. Both factor and goods markets are perfectly competitive, and each factor is fully employed in equilibrium.

The urban configuration could be *concentrated* or *dispersed*. A typical concentrated urban configuration is *monocentric*, where the MSA features a core without any subcenters. Several types of dispersed urban

---

[7] One may allow M to grow without altering the main conclusions of the paper.

configurations have been studied in the literature related to urban productivity spillovers (see Fujita and Ogawa, 1982 and Berliant et al., 2002), including (i) completely mixed (firms and households are uniformly mixed at all locations), (ii) incompletely mixed (firms and households are mixed at some locations), and (iii) duocentric (firms are located at clusters in the outskirts of the MSA). Berliant et al. (2002) show that under their location-dependent knowledge-spillover framework, the duocentric urban configuration cannot emerge in competitive general equilibrium.[8] However, whether a *core-periphery* tricentric urban structure may form remains unexplored. This latter urban configuration is the focus of our paper in the dispersed case.

We use subscript τ to denote the type of urban configuration, where the MSA is either monocentric ($\tau = M$) or core-periphery tricentric ($\tau = C$).[9] The densities of households and firms at each location z under an equilibrium configuration τ are denoted by $n_\tau(z)$ and $m_\tau(z)$, respectively. We will find conditions on exogenous parameters such that an initial urban configuration is monocentric, where firms are clustered within an interval $[-q, q] \subset \Omega$, with a uniform density of firms $m_\tau = 1$. Thus, we have:

$$q = \frac{M}{2}. \tag{3}$$

Fix a time t. We inquire whether it is possible for two satellite subcenters to form at locations $-D$ and D, where D is endogenously determined, away from the original center 0 toward the MSA fringes. More specifically, we restrict our attention to the *symmetric* core-periphery urban structure where all satellite subcenters are alike. For illustrative purposes, consider the possibility that the two satellite subcenters of equal size at $-D$ and D together absorb half of the firms. This MSA therefore features an old city center (the core), with more firms ($M/2$) than each of the two satellite subcenters (the periphery, with $M/4$ firms each).[10] The core is now an interval $[-q_1, q_1] \subset \Omega$, whereas the two satellite subcenters are represented by $[-q_4, -q_3]$ and

$[q_3, q_4]$. Workers residing in $[-q_2, -q_1] \cup [q_1, q_2]$ commute to the core and those in $[-B, -q_4] \cup [-q_3, -q_2]$ and $[q_2, q_3] \cup [q_4, B]$ commute to satellite subcenters $-D$ and D, respectively. Under this core-periphery urban structure with uniform firm density $m_\tau = 1$, it is not difficult to derive:

$$D(t) = \frac{3}{4}B(t);$$

$$q_1(t) = \frac{M}{4}; \qquad q_2(t) = \frac{B(t)}{2} = \frac{M + N(t)}{4};$$

$$q_3(t) = \frac{3}{4}B(t) - \frac{M}{8} = \frac{2M + 3N(t)}{8};$$

$$q_4(t) = B(t) - \frac{1}{8}N(t) = \frac{4M + 3N(t)}{8}. \tag{4}$$

For the rest of the paper, we will suppress the time index whenever it does not cause confusion.

2.1. Location-dependent knowledge spillovers

Uncompensated inter-firm knowledge spillovers occur within the MSA. Thus, the aggregate capital stock of all firms located in the city, denoted $\overline{K}$, has positive effects on the individual production of each firm. To capture the stylized facts documented by Jacobs (1969), we allow the magnitude of knowledge spillovers to diminish with distance.[11] Specifically, we assume that such spatially dependent knowledge spillovers take the form constructed by Berliant et al. (2002): the effective spillovers in the MSA are a fraction of $\overline{K}$ with the fraction depending on the mean location and the dispersion of firm sites.

Under a particular MSA configuration τ, define the mean location of firm sites in $S \subseteq \Omega$ as $\mu_\tau = \int_{z \in S} z m_\tau(z)\, dz$, and the dispersion index of firm sites as σ_τ. The dispersion index is measured by two times the absolute deviation: $\sigma_\tau = \frac{2}{M} \int_{z \in S} m_\tau(z)|z - \mu_\tau|\, dz$,[12] that satisfies the desired properties of absoluteness (invariance to adding a constant to every firm's location), decomposability (into subgroups with subgroup consistency), and symmetry (to the mean location).

In the case of a monocentric MSA with uniform firm density $m_\tau = 1$, the set of firm sites is $S = [-q, q]$, the

---

[8] The intuition for the result is that under a possible duocentric urban equilibrium configuration with the externality dependent on the proximity of a firm to the mean location of firms, any firm in one of the two centers will want to move to the mean location, thus disrupting the potential equilibrium.

[9] Note that if M is not a subscript, then it denotes the exogenous mass of firms.

[10] Specifically, we guess an urban configuration and then verify that under some parameter values that this urban configuration arises in equilibrium. Of course, there are other core-periphery urban configurations with different geographical distributions of firms that can arise under different sets of parameters.

[11] Berliant et al. (2006) attempt to provide microstructure for localization externalities, while Saxenian (1994) gives anecdotal evidence of their importance.

[12] As in Berliant et al. (2002), it is convenient to multiply the standard index by 2 in order to keep expressions simple.

mean location of firm sites is $\mu_M = 0$, and the dispersion index of firm sites can be computed as:

$$\sigma_M = \frac{4}{M} \int_0^q \left(\frac{M}{2q}\right) |z - 0| \, dz = q = \frac{M}{2}. \tag{5}$$

Let the subscript C denote the core-periphery tricentric configuration. When the urban structure of the MSA is core-periphery with uniform firm density $m_\tau = 1$, we have: $S = [-q_4, -q_3] \cup [-q_1, q_1] \cup [q_3, q_4]$, $\mu_C = 0$, and noting that $q_3 = q_4 - \frac{M}{4}$,

$$\sigma_C = \frac{4}{M} \left[\int_0^{q_1} m_\tau |z - 0| \, dz + \int_{q_3}^{q_4} m_\tau |z - 0| \, dz \right]$$

$$= q_4 = B - \frac{N}{8}. \tag{6}$$

One can therefore see that our dispersion index values under the two urban structures of interest are quite simple.

We now turn to inter-firm knowledge spillovers. By following the setup in Berliant et al. (2002), we measure the degree of effective spillovers of firm interactions between a particular firm $z \in S$ and the others in the MSA given a configuration of type τ as follows:

$$Q_\tau(z) = \left[B^2 - (z - \mu_\tau)^2 \right] + \left[B^2 - \varepsilon \sigma_\tau^2 \right]$$

$$= 2B^2 - (z - \mu_\tau)^2 - \varepsilon \sigma_\tau^2 > 0 \tag{7}$$

where ε indicates the exogenous degree of penalty on the dispersion of firms within the MSA. We subscript the index by τ to denote the configuration.

Two remarks are in order. First, this Q function should be viewed as a proxy (using only the first and second moments of the distribution of firms) for the overall (locational) distribution of firms. It captures the Shell (1966)–Romer (1986) type of uncompensated spillovers based on an aggregation of individual measures. Its simplicity enables us to obtain analytical results, relying as little as possible on numerical examples. As Q is strictly concave in z, the penalty for firm dispersion must be strictly convex in z. In this regard, our setup is consistent with the transactions cost setup in Ogawa and Fujita (1980) and the locational potential function setup in Fujita and Ogawa (1982), though the form of this external factor and also its interaction with the production factor inputs as specified below are entirely different. Second, this measure of effective spillovers of firm interactions within the particular MSA depends crucially on the configuration of the urban structure, via both the mean location and the disper-

sion of firm sites.[13] In the two urban configurations to be studied in this paper (monocentric and core-periphery), we have urban structures that are locationally symmetric around zero. As a consequence, the mean locations do not play any role in differentiating the two configurations ($\mu_M = \mu_C = 0$). However, since the creation of the two satellite subcenters to form a core-periphery urban structure (from an originally monocentric MSA) enlarges firm dispersion ($\sigma_C = B - \frac{N}{8} > q = \sigma_M$), it reduces the MSA-wide positive production externalities ($Q_C(z) < Q_M(z)$ for all $z \in S$) and gives the core-periphery urban configuration a disadvantage on the basis of uncompensated spillovers.

2.2. A firm's optimization problem

By employing a density $a(z)$ of land with capital K and labor L, a firm located at $z \in S$ in an MSA of configuration type τ can produce a single final good in quantity $Y_\tau(z)$. The production technology takes the Cobb–Douglas form:

$$Y_\tau(z) = A K^\alpha L^\beta \left[Q_\tau(z) \overline{K} \right]^{1-\alpha-\beta} \Theta(z)^{1-\alpha-\beta} \tag{8}$$

where $\alpha, \beta \in (0, 1)$, $\alpha + \beta \in (0, 1)$, and $\Theta(z)$ measures the effective land input given by $\Theta(z) = \min\{1, \theta(z)\}$ with $\theta(z) = a(z)$ for all $a(z) \geqslant 1$ and $\theta(z) = 0$ for all $a(z) < 1$. The uncompensated spillovers in firm production, measured by $Q_\tau(z) \overline{K}$, are taken as given by each individual firm. Moreover, the specification of the use of land implies that whenever a firm is operating (i.e., $Y_\tau(z) > 0$) in equilibrium, its land input must be: $a(z) = 1$. This "fixed lot size" setup not only simplifies the analysis greatly, but also distinguishes our model from others because we allow for factor substitution between labor and capital that is not present in Fujita and Ogawa (1982).[14]

The output of the final good is taken as the numéraire. Denote by $R(z)$ the land rent at location z, and denote by r and $w(z)$ the rental cost of capital and the wage rate, respectively. We assume that capital is freely mobile, implying that capital rent is constant across locations; yet, wages may vary with location. Each

---

[13] Thus, the magnitude of this production externality is based only on the average or aggregation of the capital stocks of individual firms as a consequence of uncompensated knowledge spillovers.

[14] In equilibrium, the production technology is constant returns to scale with respect to all private factor inputs $\{K, L, \Theta\}$. The production technology is also constant returns to scale with respect to labor and capital together, $\{L, K, \overline{K}\}$. Thus, our model is readily extended to a dynamic framework with endogenous growth by following the setup in Romer (1986).

incumbent firm maximizes its value, defined as the integral of flow profits discounted by the market rental rate, given the production technology (8) and market prices $\{R(z), r, w(z)\}$:

$$\max_{K,L,z} \Pi_\tau(z) = AK^\alpha L^\beta \left[Q_\tau(z)\overline{K}\right]^{1-\alpha-\beta} - rK$$
$$- w(z)L - R(z). \tag{9}$$

The profit optimization problem of the firm has several implications. First, equilibrium profit for any active firm must be non-negative, since the firm can always elect not to produce. Second, equilibrium profit must be the same for all active firms. If profits were higher for some firm, then the firm with lower profit can imitate one with higher profit. This implies that equilibrium profit is the same across locations occupied by firms. In fact, there will be a continuum of equilibria, each with a different, non-negative profit for firms that is constant across locations, and generally different values of B. We select the *zero-profit equilibrium* for three reasons. First, it simplifies calculations because we do not have to carry a constant through the equations and worry about the distribution of profits to households. Second, it would represent the equilibrium of a larger model where free entry were allowed. Third, and most importantly, it is a *competitive* spatial equilibrium in which the firms do *not* capture the benefits of location capitalized in land rents (particularly when a subcenter forms). This latter feature differentiates our approach from those in the subcenter development literature cited in the introduction. With zero profit in equilibrium, we can define the bid rent function of a firm as:

$$R_F(z) = \max_{K,L} AK^\alpha L^\beta \left[Q_\tau(z)\overline{K}\right]^{1-\alpha-\beta} - rK - w(z)L. \tag{10}$$

Under perfect competition, the land rent paid by each firm will be equal to the bid rent in equilibrium at all locations occupied by firms. Incumbent firms make their choice of production site facing the trade-offs between the land rent, labor costs and the external benefit from knowledge spillovers.

The first-order conditions with respect to K, L and z, respectively, can be derived as:

$$\alpha \frac{Y}{K} = r, \tag{11}$$

$$\beta \frac{Y}{L} = w(z), \tag{12}$$

$$-2(1-\alpha-\beta)\frac{Y}{Q_\tau(z)}(z-\mu_\tau) = R'_F(z) + w'(z)L_\tau. \tag{13}$$

Obviously, (11) and (12) equate the marginal products of capital and labor with the corresponding factor prices. Equation (13) is a locational equilibrium condition for firms (which is analogous to the Muth–Mills Condition in the context of a household's residential choice). It states that the knowledge-spillover gain from moving marginally closer to a central location is exactly offset by the increased land rent and labor cost.

Over time, firms in an initially monocentric MSA need to make one more locational decision: whether or not to set up two satellite subcenters and to relocate accordingly. Given that all firms are identical and there are no new entrants, this choice boils down to a choice between the existing monocentric MSA and the core-periphery MSA while maintaining zero profit. Implicitly, one may imagine the decision of firms in a two stage game. The first stage is the discrete locational choice between setting up a firm site in the core or the periphery. The second stage is the standard point-in-time profit maximization problem. To satisfy subgame perfection, the second stage problem is solved first, followed by the first stage of discrete locational choice. When all firms choose to be in the core, the equilibrium urban configuration is monocentric; when half of the firms choose to be in the two subcenters, the equilibrium urban configuration is core-periphery.

There are two reasons we use a dynamic game, where agents have no market power. First, we wish to study what happens when population grows exogenously over time. Second, in order to allow firms to coordinate actions in a non-cooperative game structure with perfect foresight, and shift between the monocentric and core-periphery structures, it is useful to have dynamics. Most important, the location of a new subcenter is chosen by a group of firms in a *self-fulfilling* fashion as in the dynamic macroeconomic literature, though it is completely consistent with standard microeconomics. All firms behave noncooperatively and are rational. In anticipating $\frac{1}{4}$ of the firms moving to each of the two subcenters at a particular time and $\frac{1}{2}$ of the firms staying, a mover will find it optimal to move and a firm staying will find it optimal to stay. In equilibrium, it turns out that the configuration is exactly as anticipated and the expectations are self-fulfilling. Under this equilibrium notion, there is no need for collective action or a collective decision to form subcenters. Formally, this is *noncooperative* behavior that can *appear* to be cooperative.

2.3. A household's optimization problem

All households are identical in every respect. Each household is endowed with one unit of labor every instant in time, receiving no disutility from work and thus supplying one unit of labor inelastically. Each household has an instantaneous felicity function given by $\tilde{u}(c, h)$, where c is the final good consumption and h is residential land consumption. Assume $\tilde{u}(c, h) = u(c)$ for all $h \geq 1$ and $\tilde{u}(c, h) = 0$ for all $h < 1$, where $u(c)$ is strictly increasing and concave in c with $u(0) = 0$. As a consequence, each optimizing household would consume exactly one unit of land and we can simply restrict our attention to the effective utility function, $u(c)$.

Given a constant rate of time preference $\rho > 0$, the lifetime utility of an infinitely lived household can be specified as: $U(\{c(t)\}_{t \in [0, \infty)}) = \int_0^\infty u(c(t)) e^{-\rho t} \, dt$. For analytic convenience, assume that new capital is fully depreciated and hence capital (owned by households) does not evolve over time. Since neither labor nor land is reproducible, the household's intertemporal optimization problem can be reduced to point-in-time utility maximization.

To each household, the locational choice includes both the residential site x and the workplace z. This can be conveniently divided into two steps. In the first step, each household residing in x while working at z, indexed by (x, z), maximizes utility for a given location pair (x, z), subject to the budget constraint at every point in time. In the second step, this household maximizes the value (indirect utility) achieved in the first step, $V(x, z)$, to determine the optimal locational choice for the household.

Under our set up, these two steps can be reduced to one—maximization of the "net income", denoted by $I(x, z)$ for a household residing at x while working at z. Specifically, letting $\phi > 0$ denote the commuting cost per unit of distance traveled from residence to workplace, this household earns a wage of $w(z)$, incurs a linear commuting cost of $\phi|x - z|$ and pays land rent $R(x)$ on its one unit of land consumption. Given the utility function specification, its objective is to maximize $u(c)$. Since u is strictly monotonic in its single argument, this optimization problem is equivalent to maximization of the *net income*:

$$\max_{x, z} I(x, z) = w(z) - \phi|x - z| - R(x). \qquad (14)$$

Notice that locational no-arbitrage requires each household to reach a constant net income (I_0) for any pair of occupied residential and work locations, (x, z):

$$I(x, z) = I_0. \qquad (15)$$

Under this condition with no vacant land, we can compute a household's bid rent as:

$$R_C(x) = \max_z \{w(z) - \phi|x - z| \mid I(x, z) = I_0\}. \qquad (16)$$

Under perfect competition, we can use (14)–(16) to obtain:

$$R_C(x) = w(z) - \phi|x - z| - I_0 \qquad (17)$$

for all x where consumers live. This condition implies that no household with location attributes (x, z) would have an incentive to change either residential or employment sites, because the incremental benefit from changing location exactly offsets the incremental cost.

3. Equilibrium

In this section, we will define the concept of competitive spatial equilibrium associated with a particular urban configuration τ.

Each firm earns zero profit in equilibrium. Utilizing (11) and (12), the zero profit condition can be written as:

$$Y_\tau(z) = (1 - \alpha - \beta)^{-1} R_F(z). \qquad (18)$$

This can be combined with (7) and (13) to yield an ordinary first-order differential equation for R_F with respect to z. Integrating this differential equation with respect to z generates firms' bid rent:

$$R_F(z) = \frac{2B^2 - (z - \mu_\tau)^2 - \varepsilon \sigma_\tau^2}{w(z)^{-\beta/(1-\alpha-\beta)}} \Lambda(\bar{z}_\tau) \qquad (19)$$

where $\Lambda(\bar{z}_\tau) = R_F(\bar{z}_\tau) w(\bar{z}_\tau)^{\beta/(1-\alpha-\beta)} / [2B^2 - (\bar{z}_\tau - \mu_\tau)^2 - \varepsilon \sigma_\tau^2]$ depends on a reference location $\bar{z}_\tau$ under a particular urban configuration τ.

We normalize agricultural land rent to be 1.

Definition 1. A **competitive spatial equilibrium** is a list of quantities $\{K(z), L(z), Y_\tau(z)\}$, prices $\{r(z), w(z), R(z)\}$, and population densities $\{m_\tau(z), n_\tau(z)\}$ for $z \in \Omega$, together with the set of firm sites $S \subseteq \Omega$, such that the following conditions are satisfied:

(i) profit maximization: (11) and (12);
(ii) land rent:

$$R(z) = \max \{R_F(z), R_C(z), 1\},$$

$$R(z) = \begin{cases} R_F(z) & \text{if } m_\tau(z) > 0, \\ R_C(z) & \text{if } n_\tau(z) > 0, \end{cases} \qquad (20)$$

$$R(-B) = R(B) = 1$$

where $R_F(z)$ and $R_C(z)$ are determined by (19) and (17), respectively;

(iii) set of firm sites: $S = \{z \mid m_\tau(z) > 0\}$;
(iv) zero profit: (18);
(v) land market equilibrium: $m_\tau(z) + n_\tau(z) = 1$ for all $z \in \Omega$;
(vi) labor market equilibrium: $\int_{z \in S} L(z) m_\tau(z)\, dz = N$;
(vii) population identities: $\int_{z \in S} m_\tau(z)\, dz = M$ and $\int_{z \in \Omega} n_\tau(z)\, dz = N$.

While the other conditions are straightforward, the land rents specified in part (ii) deserve further comment. In particular, (ii) defines the equilibrium market land rent $R(z)$ as the *upper envelope* of the two bid rent functions faced by each firm and household. They are $R_F(z)$ and $R_C(z)$, respectively, at location z. Whenever a firm or a household resides at a particular location z, its bid rent must be equal to the equilibrium market land rent. When a location is populated with positive densities of firms and households, it must be that $R_F(z) = R_C(z) = R(z)$. Since both the demand for and the supply of land at the boundary are completely inelastic, the equilibrium land rent is indeterminate. Thus, one must normalize the boundary land rent in order to obtain a unique equilibrium rent schedule over the MSA.[15]

This notion of equilibrium is essentially static in nature. However, the underlying equilibrium concept is dynamic, representing a time path of endogenous variables over an infinite time horizon. Agents have perfect foresight and there is no uncertainty. They cannot transfer income between periods. We have omitted the time index purely for notational simplicity. There is essentially no connection between time periods aside from population growth (in particular, no state variable is the result of a choice made by agents in an earlier time period). Thus, the equilibrium problem reduces to one of a repeated static problem; the only difference between periods is exogenous population growth.

We call a competitive spatial equilibrium $\{\{K(z), L(z), Y_\tau(z)\}, \{r(z), w(z), R(z)\}, \{m_\tau(z), n_\tau(z)\}, S\}$ *symmetric* if for almost every $z \in S$, $L(z) = L$ (so all firms use the same quantity of labor) and every endogenous variable is symmetric around location 0, namely its values at z and $-z$ are identical. Of course, this implies the same use of labor by all firms. Symmetry is used to reduce the indeterminacy of equilibrium in our model.

Under symmetry and uniform density of firm land use, we can solve from the labor market equilibrium condition the equilibrium density of labor demand:

$$L = \frac{N}{M}. \tag{21}$$

Since M is fixed but N is growing at a constant rate ν, L grows over time at the same rate ν.

In the next two sections, we establish conditions under which a particular urban configuration is supported as a competitive spatial equilibrium. We begin by studying the case of a monocentric MSA. We then examine how satellite subcenters may form away from the original center and toward the outskirts of the MSA as population grows. This would result in a core-periphery urban structure.

4. Equilibrium urban configurations

We now establish conditions under which a monocentric or a core-periphery urban configuration is supported in a symmetric competitive spatial equilibrium.[16]

4.1. The monocentric urban configuration

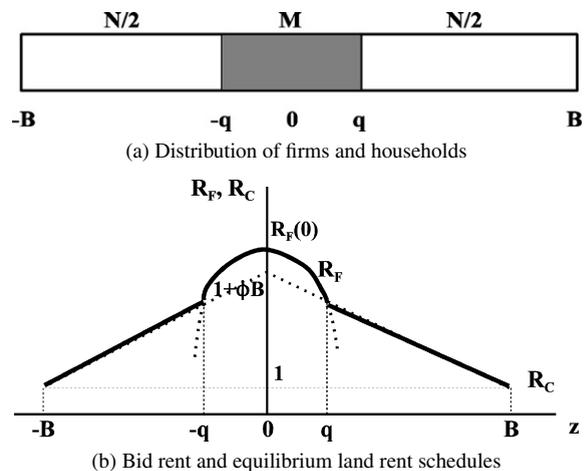As illustrated in Fig. 1, in the case of a monocentric MSA ($\tau = M$), $S = [-q, q]$, $\mu_M = 0$, and

Fig. 1. Monocentric urban configuration.

---

[15] Condition (vi) of the definition is also worthy of discussion. Details are given in Berliant et al. (2002, p. 285). An aggregate labor market clearance condition is used instead of requiring firm demand at each location to be equal to the number of consumers who wish to work there. Briefly, the reason is that due to the wage gradient, in equilibrium consumers are completely indifferent about where their employer is located (as long it is within an area populated by firms), so it is unnecessary to keep track of local labor market clearance.

[16] For ranges of parameters other than those studied in this section, other equilibrium configurations can arise. For example, when commuting cost is high, there is a completely mixed equilibrium configuration where consumers work at their residence and there is no commuting.

$\sigma_M = q = \frac{M}{2}$. All households reside in the outskirts $[-B, -q] \cup [q, B]$ and commute to the firm cluster S to work. In equilibrium, the land rent at the fringe must satisfy: $R_C(-B) = R_C(B) = 1$. In this urban configuration, an important reference point is $\bar{z}_M = q = \frac{M}{2}$, at which $R_F(q) = R_C(q)$ (similar properties hold for location $-q$).

Next, we shall impose two conditions on exogenous parameters so that the monocentric configuration is supported as an equilibrium.

Condition N: $\frac{N(t)}{M} > \frac{\beta}{1-\alpha-\beta}$.

Condition M: $\phi(1 + \frac{N(t)}{M}) < \frac{M(1+t\frac{N(t)}{2})}{2(\frac{M+N(t)}{2})^2-(1+\varepsilon)(\frac{M}{2})^2}$.

We can then establish:

Theorem 1 (*Existence*). *Under Conditions N and M, there is a **symmetric competitive spatial equilibrium with a monocentric urban configuration** in which all firms are clustered around the city center $S = [-\frac{M}{2}, \frac{M}{2}]$ and all households reside at the outskirts of the MSA. The equilibrium wage schedule and the equilibrium land rent are:*

$$w(x) = \frac{\beta}{1-\alpha-\beta} \frac{M}{N}\left[1 + \left(\frac{1-\alpha}{\beta}\right)\phi\frac{N}{2}\right] - \phi|x|, \tag{22}$$

$$R(z) = \begin{cases} R_F(z) & \text{for } z \in S, \\ R_C(z) & \text{for } z \in \Omega \setminus S \end{cases} \tag{23}$$

where $R_F(z)$ and $R_C(z)$ are given by

$$R_F(z) = \frac{[2(\frac{M+N}{2})^2 - \varepsilon(\frac{M}{2})^2 - z^2]\Lambda(q)}{\{\frac{\beta}{1-\alpha-\beta}\frac{M}{N}[1 + (\frac{1-\alpha}{\beta})\phi\frac{N}{2}] - \phi|z|\}^{\beta/(1-\alpha-\beta)}}, \tag{24}$$

$$R_C(z) = 1 + \phi(B - |z|) \tag{25}$$

and $\Lambda(q) = \frac{1+\phi\frac{N}{2}}{2(\frac{M+N}{2})^2-(1+\varepsilon)(\frac{M}{2})^2}\left[\frac{\beta}{1-\alpha-\beta}\frac{M}{N}\left(1+\phi\frac{N}{2}\right)\right]^{\beta/(1-\alpha-\beta)}$.

By examining the equilibrium wage and land rent schedules, $\{w(x)\}$ and $\{R(z)\}$ pinned down in the theorem above, one can see that both are decreasing in the distance from the city center 0. Moreover, an increase in the household population (N) for a given mass of firms reduces the wage rate at each location, while an increase in the degree of penalty on overall firm dispersion (ε) has no effect on wages and an ambiguous effect on land rents.

In order to support the monocentric urban configuration, it is required that each household has an incentive to work away from home because by working for a firm closer to the city center, the incremental gain from a higher wage dominates the induced cost of commuting. Condition N guarantees $R_F(0) > R_C(0)$ while Condition M guarantees $|R'_F(q)| > |R'_C(q)|$. These together ensure a locational equilibrium choice of workplace and residence under the monocentric urban configuration in which the equilibrium land rent is the upper envelope of firms' and households' bid rents. Thus, Conditions N and M are *necessary and sufficient* for supporting a symmetric competitive spatial equilibrium with a monocentric urban configuration.

4.2. Equilibrium with a core–periphery urban configuration

As illustrated in Fig. 2, in the case of a core-periphery MSA ($\tau = C$), $S = [-q_4, -q_3] \cup [-q_1, q_1] \cup [q_3, q_4]$, $\mu_C = 0$, and $\sigma_C = B - \frac{N}{8}$. All households residing in $[-q_2, -q_1] \cup [q_1, q_2]$ commute to the core $S_C = [-q_1, q_1]$ to work, whereas those in $[-B, -q_4] \cup [-q_3, -q_2]$ commute to the satellite subcenter to the left of the core $S_P^L = [-q_4, -q_3]$ and those in $[q_2, q_3] \cup$
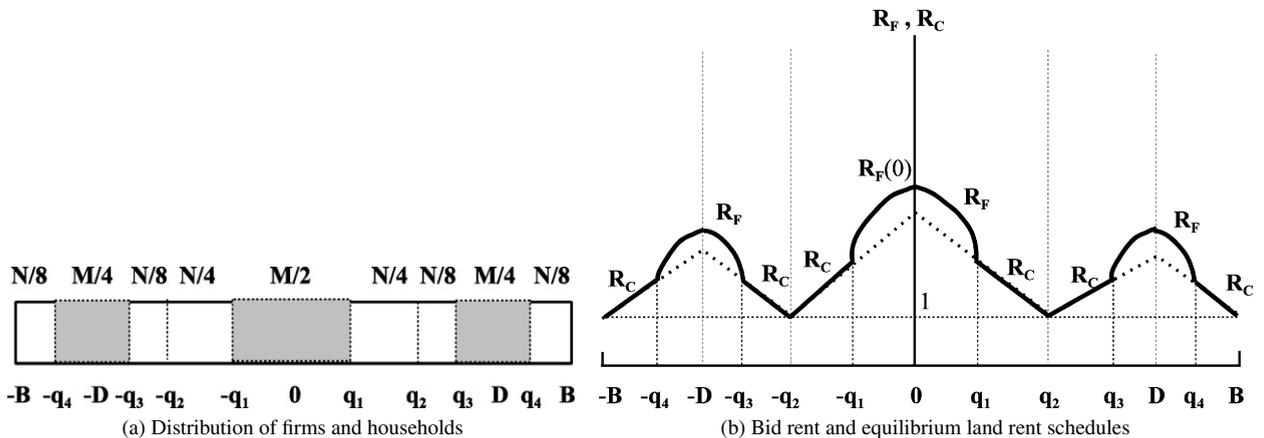

Fig. 2. Core–periphery urban configuration.

(a) Distribution of firms and households

(b) Bid rent and equilibrium land rent schedules

$[q_4, B]$ commute to the right satellite subcenter $S_P^R = [q_3, q_4]$, respectively ($D = \frac{3}{4}B$). At a symmetric equilibrium, $R_C(-B) = R_C(-q_2) = R_C(q_2) = R_C(B) = 1$ and there are now two important reference points $\bar{z}_C \in \{q_1, q_4\}$ defining the boundary between firms and households, at which $R_F(q_1) = R_C(q_1)$ and $R_F(q_4) = R_C(q_4)$ (noting that the properties of bid rents at q_3 are just a mirror image of those at q_4).

We now define $\Gamma(N(t)) = 2(\frac{M+N(t)}{2})^2 - \varepsilon(\frac{4M+3N(t)}{8})^2$. Consider,

Condition C: $\quad \phi\big(1 + \frac{N(t)}{M}\big) < \min\left\{ \frac{\frac{M}{2}(1+\phi\frac{N(t)}{4})}{\Gamma(N(t))-(\frac{M}{4})^2}, \right.$

$\left. \frac{(\frac{4M+3N(t)}{4})(1+\phi\frac{N(t)}{8})}{\Gamma(N(t))-(\frac{4M+3N(t)}{8})^2} \right\}.$

We can then establish:

Theorem 2. *Under Conditions N and C, there is a* ***symmetric competitive spatial equilibrium with a core–periphery urban configuration*** *in which*:

(i) *half of the firms are clustered in the core, $S_C = [-\frac{M}{4}, \frac{M}{4}]$, whereas the remaining half are symmetrically clustered within the two satellite subcenters, $S_P^L = [-\frac{4M+3N(t)}{8}, -\frac{2M+3N(t)}{8}]$ and $S_P^R = [\frac{2M+3N(t)}{8}, \frac{4M+3N(t)}{8}]$;*

(ii) *all households residing in inner locations of the MSA, $[-\frac{M+N(t)}{4}, -\frac{M}{4}] \cup [\frac{M}{4}, \frac{M+N(t)}{4}]$, commute to the core S_C to work; those in outer left locations, $[-\frac{M+N(t)}{2}, -\frac{4M+3N(t)}{8}] \cup [-\frac{2M+3N(t)}{8}, -\frac{M+N(t)}{4}]$, commute to the distance-minimizing satellite subcenter S_P^L to work; whereas those in outer right locations, $[\frac{M+N(t)}{4}, \frac{2M+3N(t)}{8}] \cup [\frac{4M+3N(t)}{8}, \frac{M+N(t)}{2}]$, commute to the distance-minimizing satellite subcenter S_P^R to work.*

The equilibrium wage schedule and the equilibrium land rent are:

$$w(x) = \begin{cases} \frac{\beta}{1-\alpha-\beta}\frac{M}{N}[1 + (\frac{1-\alpha}{\beta})\phi\frac{N}{2}] - \phi(\frac{B}{2} + |x|) \\ \quad for\ x \in [-\frac{3B}{4}, -\frac{M}{4}] \cup [\frac{M}{4}, \frac{3B}{4}], \\ \frac{\beta}{1-\alpha-\beta}\frac{M}{N}[1 + (\frac{1-\alpha}{\beta})\phi\frac{N}{2}] - \phi|x| \\ \quad for\ x \in [-B, -\frac{3B}{4}] \cup [\frac{3B}{4}, B]; \end{cases} \quad (26)$$

$$R(z) = \begin{cases} R_F(z) & for\ z \in S, \\ R_C(z) & for\ z \in \Omega \setminus S \end{cases} \quad (27)$$

where $R_F(z)$ and $R_C(z)$ are given by,

$$R_F(z) = \begin{cases} \frac{[\Gamma(N)-z^2]\Lambda_1(q_1)}{\{\frac{\beta}{1-\alpha-\beta}\frac{M}{N}[1+(\frac{1-\alpha}{\beta})\phi\frac{N}{2}]-\phi(\frac{B}{2}+|z|)\}^{\beta/(1-\alpha-\beta)}} \\ \quad for\ z \in [-\frac{B}{2}, \frac{B}{2}], \\ \frac{[\Gamma(N)-z^2]\Lambda_2(q_4)}{\{\frac{\beta}{1-\alpha-\beta}\frac{M}{N}[1+(\frac{1-\alpha}{\beta})\phi\frac{N}{2}]-\phi|z|\}^{\beta/(1-\alpha-\beta)}} \\ \quad for\ z \in [-B, -\frac{B}{2}] \cup [\frac{B}{2}, B]; \end{cases} \quad (28)$$

$$R_C(z) = \begin{cases} 1 + \phi(\frac{B}{2} - |z|) \\ \quad for\ z \in [-\frac{3B}{4}, -\frac{M}{4}] \cup [\frac{M}{4}, \frac{3B}{4}], \\ 1 + \phi(B - |z|) \\ \quad for\ z \in [-B, -\frac{3B}{4}] \cup [\frac{3B}{4}, B] \end{cases} \quad (29)$$

and

$$\Lambda_1(q_1) = \frac{(1 + \frac{\phi N}{4})[\frac{\beta M(1+\phi N/2)}{(1-\alpha-\beta)N} - \frac{\phi N}{4}]^{\beta/(1-\alpha-\beta)}}{\Gamma(N) - (\frac{M}{4})^2},$$

$$\Lambda_2(q_4) = \frac{(1 + \frac{\phi N}{8})[\frac{\beta M(1+\phi N/2)}{(1-\alpha-\beta)N} - \frac{3\phi N}{8}]^{\beta/(1-\alpha-\beta)}}{\Gamma(N) - (\frac{4M+3N}{8})^2}.$$

Under the core-periphery urban configuration, the equilibrium wage and land rent schedules, $\{w(x)\}$ and $\{R(z)\}$, are decreasing in the distance from the respective centers of the core and the two satellite subcenters. Again, an increase in the household population (N) for a given mass of firms reduces the wage rate at each location, whereas an increase in the degree of penalty on overall firm dispersion (ε) has no effect on wages and an ambiguous effect on land rents.

To support this urban configuration, households need to have an incentive to work away from home. Moreover, by working for a closer firm in the respective clusters instead of a firm in the core, the incremental gain from a higher wage dominates the induced cost of commuting. Under the level and slope conditions (Conditions N and C), a locational equilibrium choice of workplace and residence under the core-periphery urban configuration implies that the equilibrium land rent is the upper envelope of firms' and households' bid rents. Conditions N and C are therefore *necessary and sufficient* for supporting a symmetric competitive spatial equilibrium with a core-periphery urban configuration.

5. A transition from monocentric to core–periphery urban configurations

We are now ready to examine how satellite subcenters may form away from the original center and toward the outskirts of the MSA as population grows. This results in a core-periphery urban structure. Due to the

complexity of the relevant conditions (particularly Conditions M vs. C), we are unable to characterize this transition analytically. Instead we elaborate our findings using numerical exercises.

Normalize $N(0) = 1$ and set $M = 0.1$. We select the following parameter values:

(i) (Production Technology) $\alpha = 0.3$, $\beta = 0.5$, $\varepsilon = 2.25$;
(ii) (Commuting) $\phi = 0.005$;
(iii) (Population Growth) $\nu = 2\%$.

It is easily seen that Condition N always holds. We can further check Conditions M and C to conclude:

(i) $(0 \leqslant t < t_1)$ The MSA features a monocentric urban configuration ($\tau = M$).
(ii) $(t_1 \leqslant t < t_2)$ The monocentric configuration disappears and a core-periphery urban configuration forms in the MSA ($\tau = C$) where the geographical centers of the two satellite subcenters continue to move outward at rate ν.

According to our calculations, $t_1 = 19.84$ and $t_2 = 27.09$. We assume that the initial equilibrium configuration, at time $t = 0$, is monocentric. As population grows, eventually the monocentric equilibrium disappears (at t_1), and the core-periphery urban configuration appears and persists (at least until time t_2). We do not make any claims about uniqueness of symmetric competitive spatial equilibrium. Particularly, the reader should be warned that there are other types of equilibria not considered here; we may easily have more equilibria coexisting with the ones we have studied.[17]

One may then inquire what happens when $t \geqslant t_2$. This is speculation on our part, based on some numerical exercises. In this case, neither of the two urban configurations that we have considered can arise. When t is not too large, the MSA becomes either: (a) mixed with firms and workers both living at some locations or (b) features more than two satellite subcenters.[18] When t is

very large, ultimately diminishing returns to the number of subcenters dominates as the city expands out due to the reduction in overall knowledge spillovers. In this case, new cores will be formed.

These critical times t_1 and t_2 change as the population growth rates vary:

ν	1%	5%	10%
t_1	39.69	7.94	3.97
t_2	54.18	10.84	5.42

Thus, if a city grows slowly (say, at 1%), it may maintain a monocentric configuration for almost 40 years and have two satellite subcenters forming for another 15 years. If a city grows fast (say, at a 10% rate, similar to Las Vegas, Phoenix and San Antonio), the monocentric urban structure may quickly fall in as few as 4 years with two satellite subcenters forming only for about a year and a half before more satellite subcenters develop. This process may be expedited if other urban congestion factors are taken into account. Thus, one may not expect to find many US cities featuring a monocentric urban structure today, corroborating the empirical findings in McMillen and Smith (2003).

Moreover, McMillen and Smith (2003) suggest that the critical population size for subcenters to form (without accounting for congestion factors) is about 2.68 million. By setting $N(0) = 1.8$ million, we can reproduce this critical population as $N(19.84) = 2.68$ million. Thus, our model is rich enough to conduct calibration analysis.

We can analyze numerically comparative statics with respect to commuting cost as well as production-spillover and dispersion-penalty parameters of the model. Our results, which are robust to plausible variations in the parameters, indicate that the speed of forming satellite subcenters increases with the unit commuting cost (ϕ) but decreases with the degree of penalty on the dispersion of firms for knowledge spillovers (ε). While the latter may be difficult to measure in practice, the long commute from the Staples Center to Disneyland and the Rose Bowl may explain why Los Angeles features as many as 46 subcenters.

6. Concluding remarks

We have examined conditions necessary and sufficient for the equilibrium structure of a closed city (with

---

[17] Since we are using the guess and verify method for analyzing equilibrium, it is necessary to guess.

[18] A plausible core-periphery urban structure is one with two larger subcenters centered at $-\frac{3B}{5}$ and $\frac{3B}{5}$, and two smaller subcenters centered at $-\frac{9B}{10}$ and $\frac{9B}{10}$. Then the core, $[-\frac{M}{5}, \frac{M}{5}]$, has firms of mass $\frac{2M}{5}$ and employees residing in $[-\frac{2B}{5}, -\frac{M}{5}]$ and $[\frac{M}{5}, \frac{2B}{5}]$. Each of the two larger satellite subcenters hosts firms of mass $\frac{M}{5}$ and employees residing in $[-\frac{4B}{5}, -(\frac{3B}{5} + \frac{M}{10})] \cup [-(\frac{3B}{5} - \frac{M}{10}), -\frac{2B}{5}]$ and $[\frac{2B}{5}, \frac{3B}{5} - \frac{M}{10}] \cup [\frac{3B}{5} + \frac{M}{10}, \frac{4B}{5}]$. Each of the two smaller satellite subcenters contains firms of mass $\frac{M}{10}$ and employees residing

in $[-B, -(\frac{9B}{10} + \frac{M}{20})] \cup [-(\frac{9B}{10} - \frac{M}{20}), -\frac{4B}{5}]$ and $[\frac{4B}{5}, \frac{9B}{10} - \frac{M}{20}] \cup [\frac{9B}{10} + \frac{M}{20}, B]$.

exogenous but growing population, an endogenous city boundary, and with localization externalities between firms) to be monocentric or to have subcenters. To relate the model to city growth, we have postulated an exogenous population growth process, where population is the only state variable that changes over time. Beginning with a monocentric configuration, there is eventually a transition to an equilibrium with employment subcenters. There is no need for firms to receive a share of rents for this to happen, nor is a government or a developer necessary. We have actually biased the externality against subcenter formation, relying only on distance to the center of firm locations rather than on the local density of firms around any given firm. From a technical viewpoint, for low populations, we select the monocentric equilibrium as an initial condition, to see the evolution of a city. With high populations, the equilibrium with subcenters appears and survives while the monocentric equilibrium disappears.

There are many more questions raised by this work than we have answered. As population rises further, can an equilibrium urban configuration with more subcenters coexist with an equilibrium featuring a mixed configuration in which the firms and consumers co-locate? With an equilibrium featuring subcenters, will subcenters multiply at an accelerated rate as population grows? If more subcenters are formed, will the configuration be consistent with central place theory, and is the theory of the open city model in this context similar? It may be of particular interest to examine whether the main findings obtained under our competitive spatial equilibrium setup differ from those established by Abdel-Rahman and Wang (1995) within the search equilibrium framework. One may study the welfare properties of the model, particularly comparing the level of community welfare achieved under each urban configuration. In our model, all firms earn zero profit and all households gain identical net income in equilibrium. Thus, the total land rent collected by the absentee landlord alone determines the welfare of an MSA. Due to the complexity of the bid rent functions, this welfare comparison cannot be conducted analytically, but numerical exercises may be performed for such purposes. Finally, the model could be useful for examining the recent work on urban growth boundaries, such as Brueckner (2005) and Anas and Rhee (2006) that has congestion. Whether boundaries cause welfare to go up or down in the second-best context (with a congestion externality) seems to depend on whether the city in the model is assumed to be monocentric or polycentric. In our model, we do not need to make an assumption about the urban configuration, but rather allow it to be determined endogenously. One may

thus revisit this welfare issue in the second-best context (in the presence of a production externality), taking into account the endogenous formation of subcenters. We leave all of these topics to future research, particularly for our students and grandstudents in a cluster of urban economists with positive knowledge spillovers.

Acknowledgments

Appendix A

In this appendix, we prove the main theorems derived in the paper.

Proof of Theorem 1. Since the MSA is symmetric with respect to the geographic center, our focus is on the right half of the city, $[0, B]$. Evaluating (17) at B and applying $R_C(B) = 1$, we can solve the wage schedule as:

$$w(z) = 1 + I_0 + \phi(B - z) \tag{30}$$

and the equilibrium bid rent schedule for households (x, z) as:

$$R_C(x) = 1 + \phi(B - z) - \phi|x - z|. \tag{31}$$

Evaluating this at $x = q$ and using equilibrium land rent condition $R_F(q) = R_C(q)$, we get:

$$R_C(q) = R_C(B) + \phi(B - q)$$
$$= 1 + \phi(B - q) = R_F(q). \tag{32}$$

Substituting (30) and (32) into (19) yields the equilibrium bid rent schedule for firms over $S = [0, q]$:

$$R_F(z) = \frac{2B^2 - z^2 - \varepsilon q^2}{[1 + I_0 + \phi(B - z)]^{\beta/(1-\alpha-\beta)}} \Lambda(q) \tag{33}$$

where $\Lambda(q) = \frac{1+\phi(B-q)}{2B^2-(1+\varepsilon)q^2}[1 + I_0 + \phi(B - q)]^{\beta/(1-\alpha-\beta)}$.

Combining (12), (18) and (30), one can evaluate at $z \in [0, q]$ firms' equilibrium land rent payment:

$$R_F(z) = \frac{1 - \alpha - \beta}{\beta}\big[1 + I_0 + \phi(B - z)\big]L. \tag{34}$$

Utilizing (32) and (21), we can solve for the piece of the expression in brackets, namely endogenously determined net income:

$$1 + I_0 + \phi B = \frac{\beta}{1 - \alpha - \beta} \frac{M}{N} \left(1 + \phi \frac{N}{2}\right) + \phi \frac{M}{2},$$

or,

$$1 + I_0 + \phi B = \frac{\beta}{1 - \alpha - \beta} \frac{M}{N} \left[1 + \left(\frac{1 - \alpha}{\beta}\right) \phi \frac{N}{2}\right]. \tag{35}$$

With this and accounting appropriately for the left-half of the city $[-B, 0]$, we can obtain (22), (25), and (24). Moreover, we can substitute (21) and (35) into (34) to derive:

$$R_F(0) = \frac{1 - \alpha - \beta}{\beta} \frac{N}{M} (1 + I_0 + \phi B)$$
$$= 1 + \left(\frac{1 - \alpha}{\beta}\right) \phi \frac{N}{2}. \tag{36}$$

From (31) and (36), one can see that $R_F(0) > R_C(0)$ if and only if Condition N holds.

Further differentiating (33) gives:

$$\frac{R'_F(z)}{R_F(z)} = -\left[\frac{2z}{2B^2 - z^2 - \varepsilon q^2} - \frac{\beta}{1 - \alpha - \beta} \frac{\phi}{1 + I_0 + \phi(B - z)}\right]. \tag{37}$$

Evaluating this expression at q and manipulating, we have:

$$R'_F(q) = -\left[\frac{M(1 + \phi \frac{N}{2})}{2(\frac{M+N}{2})^2 - (1 + \varepsilon)(\frac{M}{2})^2} - \phi \frac{N}{M}\right].$$

To guarantee $Q_M(z) > 0$ for all $z \in S = [0, q]$, we need:

$$2 \left(\frac{M + N}{2}\right)^2 - (1 + \varepsilon) \left(\frac{M}{2}\right)^2 > 0$$

which is implied by Condition M (noting also that this inequality implies that the right hand side of the inequality in Condition M is positive). In competitive spatial equilibrium, the following land rent slope condition must be met: $|R'_F(q)| > |R'_C(q)| = \phi$, which is equivalent to Condition M. When the inequality in Condition M is met, the above inequality must be met as well and the bid rent must be downward-sloping, i.e., $R'_F(z) < 0$ for all $z \in S = [0, q]$. $\quad \square$

Proof of Theorem 2. Following a procedure similar to that in the proof of Theorem 1 and focusing on the right half of the city $[0, B]$, we can evaluate (17) at $q_2 = \frac{B}{2}$ and $q_2 = B$, and then use $R_C(q_2) = R_C(B) = 1$ to obtain:

$$w(z) = \begin{cases} 1 + I_0 + \phi(q_2 - z) & \text{for } z \in [q_1, D], \\ 1 + I_0 + \phi(B - z) & \text{for } z \in [D, B]; \end{cases} \tag{38}$$

$$R_C(q_1) = 1 + \phi(q_2 - q_1) = R_F(q_1), \tag{39}$$

$$R_C(q_4) = 1 + \phi(B - q_4) = R_F(q_4). \tag{40}$$

Substitution of (38), (39) and (40) into (19) gives:

$$R_F(z) = \begin{cases} \dfrac{[2B^2 - z^2 - \varepsilon(B - \frac{N}{8})^2] \Lambda_1(q_1)}{[1 + I_0 + \phi(\frac{B}{2} - z)]^{\beta/(1 - \alpha - \beta)}} \\ \quad \text{for } z \in [0, q_2], \\ \dfrac{[2B^2 - z^2 - \varepsilon(B - \frac{N}{8})^2] \Lambda_2(q_4)}{[1 + I_0 + \phi(B - z)]^{\beta/(1 - \alpha - \beta)}} \\ \quad \text{for } z \in [q_2, B] \end{cases} \tag{41}$$

where

$$\Lambda_1(q_1) = \frac{[1 + \phi(\frac{B}{2} - q_1)][1 + I_0 + \phi(\frac{B}{2} - q_1)]^{\frac{\beta}{1 - \alpha - \beta}}}{2B^2 - (q_1)^2 - \varepsilon(B - \frac{N}{8})^2},$$

$$\Lambda_2(q_4) = \frac{[1 + \phi(B - q_4)][1 + I_0 + \phi(B - q_4)]^{\frac{\beta}{1 - \alpha - \beta}}}{2B^2 - (q_4)^2 - \varepsilon(B - \frac{N}{8})^2}.$$

Using (12), (18), (21) and (30), we get:

$$R_F(z) = \begin{cases} \frac{1 - \alpha - \beta}{\beta} \frac{N}{M}[1 + I_0 + \phi(\frac{B}{2} - z)] \\ \quad \text{for } z \in [q_1, D], \\ \frac{1 - \alpha - \beta}{\beta} \frac{N}{M}[1 + I_0 + \phi(B - z)] \\ \quad \text{for } z \in [D, B]. \end{cases} \tag{42}$$

Under this urban configuration, $1 + I_0 + \phi B$ is still determined by (35). We can thus derive (26), (28) and (29), as well as (36). Since $R_F(0)$ and $R_C(0)$ are the same as in Theorem 1, Condition N still applies to the current case, which is necessary and sufficient to ensure $R_F(0) > R_C(0)$. Additionally, however, we must have: $R_F(D) > R_C(D)$. Manipulations similar to those used in the proof of Theorem 1 yield:

$$R_F(D) = \frac{1 - \alpha - \beta}{\beta} \frac{N}{M} \left[1 + I_0 + \phi(B - D)\right]$$
$$= 1 + \left(\frac{1 - \alpha}{\beta}\right) \phi \frac{N}{8},$$

$$R_C(D) = 1 + \phi(B - D) = 1 + \phi \frac{N}{8}.$$

Since $1 - \alpha - \beta > 0$, we have $\frac{1 - \alpha}{\beta} > 1$ and hence $R_F(D) > R_C(D)$ always holds. That is, there is no additional level condition required for the core-periphery urban configuration to arise.

We must further check slope conditions. Differentiating (33) implies:

$$\frac{R'_F(z)}{R_F(z)} = \begin{cases} -[\frac{2z}{2B^2 - z^2 - \varepsilon(B - \frac{N}{8})^2} - \frac{\beta}{1 - \alpha - \beta} \frac{\phi}{1 + I_0 + \phi(\frac{B}{2} - z)}] \\ \quad \text{for } z \in [0, q_2], \\ -[\frac{2z}{2B^2 - z^2 - \varepsilon(B - \frac{N}{8})^2} - \frac{\beta}{1 - \alpha - \beta} \frac{\phi}{1 + I_0 + \phi(B - z)}] \\ \quad \text{for } z \in [q_2, B]. \end{cases}$$
$$\tag{43}$$

Evaluating this at q_1 and q_4, one obtains:

$$R'_F(q_1) = -\left[\frac{\frac{M}{2}(1 + \phi \frac{N}{4})}{2(\frac{M+N}{2})^2 - (\frac{M}{4})^2 - \varepsilon(\frac{4M+3N}{8})^2} - \phi \frac{N}{M} \right],$$

$$R'_F(q_4) = -\left[\frac{(\frac{4M+3N(t)}{4})(1 + \phi \frac{N(t)}{8})}{2(\frac{M+N(t)}{2})^2 - (1 + \varepsilon)(\frac{4M+3N(t)}{8})^2} - \phi \frac{N}{M} \right].$$

Condition C is not only necessary and sufficient to guarantee both $|R'_F(q_1)| > |R'_C(q_1)| = \phi$ and $|R'_F(q_4)| > |R'_C(q_4)| = \phi$, but also sufficient to guarantee $Q_C(z) > 0$ for all $z \in S$ and $R'_F(z) < 0$ for all $z \in [0, q_1] \cup [D, q_4]$. $\square$

References

Abdel-Rahman, H., Wang, P., 1995. Toward a general-equilibrium theory of a core–periphery system of cities. Regional Science and Urban Economics 25, 529–546.

Anas, A., 2004. Vanishing cities: What does the new economic geography imply about the efficiency of urbanization? Journal of Economic Geography 4, 181–199.

Anas, A., Rhee, H.-J., 2006. Curbing excess sprawl with congestion tolls and urban boundaries. Regional Science and Urban Economics 36, 510–541.

Anas, A., Xiong, K., 2003. Intercity trade and the industrial diversification of cities. Journal of Urban Economics 54, 258–276.

Anas, A., Arnott, R., Small, K., 1998. Urban spatial structure. Journal of Economic Literature 36, 1426–1464.

Baldwin, R., Forslid, R., 2000. The core–periphery model and endogenous growth: Stabilising and de-stabilising integration. Economica 67, 307–324.

Berliant, M., Wang, P., 2004. Dynamic urban models: Agglomeration and growth. In: Capello, R., Nijkamp, P. (Eds.), Urban Dynamics and Growth: Advances in Urban Economics. North-Holland, Amsterdam, pp. 533–581.

Berliant, M., Peng, S., Wang, P., 2002. Production externalities and urban configuration. Journal of Economic Theory 104, 275–303.

Berliant, M., Reed, R., Wang, P., 2006. Knowledge exchange, matching, and agglomeration. Journal of Urban Economics 60, 69–95.

Brueckner, J.K., 2005. Urban growth boundaries: An effective second-best remedy for unpriced traffic congestion? Mimeo. University of California, Irvine, CA.

Fujita, M., 1989. Urban Economic Theory. Cambridge Univ. Press, Cambridge.

Fujita, M., Krugman, P., 1995. When is the economy monocentric? von Thünen and Chamberlin unified. Regional Science and Urban Economics 25, 505–528.

Fujita, M., Mori, T., 1997. Structural stability and evolution of urban systems. Regional Science and Urban Economics 27, 399–442.

Fujita, M., Ogawa, F., 1982. Multiple equilibria and structural transition of non-monocentric urban configurations. Regional Science and Urban Economics 12, 161–196.

Fujita, M., Thisse, J.-F., 2002. Economics of Agglomeration. Cambridge Univ. Press, Cambridge.

Fujita, M., Thisse, J.-F., Zenou, Y., 1997. On the endogenous formation of secondary employment centers in a city. Journal of Urban Economics 41, 337–357.

Henderson, J.V., Mitra, A., 1996. The new urban landscape: Developers and edge cities. Regional Science and Urban Economics 26, 613–643.

Henderson, J.V., Slade, E., 1993. Development games in non-monocentric cities. Journal of Urban Economics 34, 207–229.

Jacobs, J., 1969. The Economy of Cities. Random House, New York, NY.

Krugman, P., 1993. On the number and location of cities. European Economic Review 37, 293–298.

Lucas, R.E. Jr., Rossi-Hansberg, E., 2002. On the internal structure of cities. Econometrica 70, 1445–1476.

McMillen, D.P., Smith, S.C., 2003. The number of subcenters in large urban areas. Journal of Urban Economics 53, 321–338.

Ogawa, M., Fujita, M., 1980. Equilibrium land use patterns in a non-monocentric city. Journal of Regional Science 20, 455–475.

Romer, P., 1986. Increasing returns and long-run growth. Journal of Political Economy 94, 1002–1037.

Rosenthal, S., Strange, W., 2003. Geography, industrial organization, and agglomeration. Review of Economics and Statistics 85, 377–393.

Saxenian, A., 1994. Regional Advantage: Culture and Competition in Silicon Valley and Route 128. Harvard University Press, Cambridge, MA.

Shell, K., 1966. Toward a theory of inventive activity and capital accumulation. American Economic Review 61, 62–68.