Available online at www.sciencedirect.com
SCIENCE DIRECT®

Regional Science and Urban Economics 35 (2005) 445–470

regional SCIENCE & urban ECONOMICS

www.elsevier.com/locate/econbase

The formation and growth of specialized cities: efficiency without developers or Malthusian traps

Alex Anas[a,*], Kai Xiong[b]

[a] *Department of Economics, State University of New York at Buffalo, 405 Fronczak Hall, Amherst, NY 14260, USA*
[b] *Chase Manhattan Bank Plaza, 10th floor, New York, NY 10081, USA*

Received 12 December 2001; received in revised form 29 October 2003; accepted 5 November 2003
Available online 7 February 2004

Abstract

Consider one diversified city initially producing a manufacture with a variety of services used as inputs. Services can be produced locally or in other cities and imported at cost. A new city forms at some point under exogenous population growth. Depending on parameters, it is efficient that this new city either be diversified or specialized, producing either manufactures or services. When it is efficient that it be specialized, then the market, without any help from developers (laissez-faire), correctly times the emergence of such a new city that starts from a tiny size. Contrary to previous results in the literature, it can be efficient for the specialized cities to emerge before the diversified parent city becomes overpopulated. On the growth path, the specialized city remains smaller than its diversified parent and offers a lower wage and rent. Lower inter-city trading cost increases the size of the specialized city. Such a specialized city can emerge and persist even when it is more efficient for a diversified city to have emerged earlier. The specialized city will self-organize if developers do not act just on time to set up diversified cities.

JEL classification: D62; F12; R12; R13
Keywords: Growth; City formation; Specialization; Trade; Externalities

1. Introduction

One of the most important aspects of the economics of cities is the dynamics by which new cities are formed out of existing ones along the growth path of a system of cities. A

* Corresponding author. Chase Manhattan Bank, USA.
E-mail addresses: alexanas@buffalo.edu (A. Anas), Kai.Xiong@chase.com (K. Xiong).

0166-0462/$ - see front matter
doi:10.1016/j.regsciurbeco.2003.11.001

correspondingly important problem is the formation of subcenters or edge cities *within* metropolitan areas (Garreau, 1991; Henderson and Mitra, 1991; Anas and Kim, 1996). In these twin problems, there are several interesting questions. *First*, does the new settlement (city or subcenter) emerge at the efficient time under laissez-faire,[1] or is planned development required to optimize the timing? *Second*, will the efficient and market determined growth paths be stable? *Third*, is it efficient that the new city or subcenter be similar to the old or should it have a unique (more specialized) structure? For example, should edge cities or new cities in a national setting be specialized in manufacturing or in services or should they (can they) evolve as smaller copies of the central business district (or of pre-existing cities)? In addition, with regard to the third question, does laissez-faire foster the creation of the right industry mix, or is planned development required to guide the process?

There are two schools one could draw from in seeking new answers to these questions.[2] One school, the new economic geography (see Fujita et al., 1999) treats pecuniary links among cities as manifested via positive trading costs and differentiated products. This approach emphasizes self-organization driven by atomistic defection as the mechanism for new city formation. Unfortunately, it usually neglects internal city structure and also neglects efficiency. The older approach of Henderson (1974, 1988) treats internal city structure, emphasizes homogeneous goods, treats trade among cities but unfortunately assumes that trading costs are zero. Hence, interactions between cities or links between industries are costless. Henderson emphasizes city developers, not atomistic defection, as the mechanism for city formation and claims that such a mechanism is necessary to achieve efficiency.

In our model, to be presented in the next section, we synthesize the strengths of the two schools by treating trade as costly (as in the NEG) while treating internal urban structure and questions of efficiency as in the Henderson tradition. Like Anas and Xiong (2003), our model has two industries: manufacturing and services allowing treatment of city-industry structure. Manufacturing is competitive and exhibits constant returns, while services require a fixed input and are monopolistically competitive differentiated products. All services are traded between cities at cost for use as intermediate inputs in manufacturing while labor is also used in each industry. Each city experiences a diseconomy from internal commuting. Population growth causes the variety of services to increase. This increased local variety of services imparts an external intra-city urbanization economy on the city's manufacturing sector, known as *the home market effect*. The same external economy is also operative at the inter-city level and is stronger the lower is the cost of importing services from other cities.

We use this model to revisit the following growth scenario. Suppose that there is only one city (or one metropolitan center) initially. All services are necessarily produced there

[1] By "laissez-faire", we refer to the absence of government intervention and the non-existence of large actors such as city-developers in the timing of the creation of new settlements. Under laissez-faire new settlements self-organize by the atomistic defection of agents (a single firm or a very small number of residents) from existing cities.

[2] One of us, Anas (1988, 1992), framed some of the above questions in the context of a homogeneous industry employing only labor. For an extension with the same limitation, see Pines (2000). A purpose of the present paper is to enrich those results by using a more complete two-industry general equilibrium model of a city system described by Xiong (1998) and Anas and Xiong (2003).

together with manufactures. As population grows, a new city or subcenter should be established at some predetermined place. When it is efficient that the new city be set up as a diversified replica of the existing city (as in Henderson's models), the dynamics involve lumpy adjustments and developers would be needed to set it up at the right time. However, we will show that when it is efficient for the new city to be a specialized city, it will emerge at the optimal time under laissez-faire and will be stable. It will start out from a tiny size by atomistic defection and will grow continuously with the existing diversified parent but will remain smaller, always offering a lower wage and intra-city location cost. Depending on parameters, it may specialize in manufactures or in services. Hence, it will trade services with the parent city, importing them if it specializes in manufactures and exporting them if it specializes in services. A lowering of the cost of trading services will increase the size of the specialized city reducing the size of the diversified parent city. In addition, when service transport costs are sufficiently low, it will be efficient to set up the new specialized city even before the existing one reaches autarkic optimal city size. We also show that under laissez-faire, a specialized city can emerge and persist when it is more efficient that a diversified city should have emerged earlier. Therefore, if developers do not act to form diversified secondary cities when it is efficient to do so, specialized cities could emerge later.

Why are our results, described in the previous paragraph, important? The literature, with the exception of the new economic geography, has sought to identify the conditions under which laissez-faire can efficiently time the formation of new cities. Henderson (1974), Anas (1992), Pines (2000) and Henderson and Becker (2000) have all presented different models in which there is strong agreement on two dismal results:

(1) *Malthusian trap*: Under laissez-faire, existing cities get grossly overpopulated before new ones can emerge;
(2) *Developers are needed*: If new cities are to be set up at earlier optimal times to avoid the Malthusian outcome, then a lumpy population adjustment is needed and can only be realized by the action of developers or governments.

Henderson and Becker (2000) put it as follows:

> Starting with an arbitrary given number of cities, population will grow in each city until the $n_{\max}$[maximum city size] limit is hit, causing "bifurcation," where some workers or entrepreneurs deviate and form new cities… Population growth then continues in the new set of cities until the $n_{\max}$bound is again reached, and then more cities form. This depressing "growth" process of repeatedly hitting a Malthusian upper-bound on city sizes continues indefinitely. As solace, in the United States, there is little evidence of a bifurcation process, which implies large drops and cycles in individual city populations…virtually all cities increase in size every decade. Population losses, when they occur, are small. (p. 469).

While the authors cited empirical evidence against these dismal conclusions, they reaffirmed both of them in the context of their theoretical model. Yet, their model—like ours—contains heterogeneous economic activities and, if appropriately modified to permit inter-city interactions, it could be used to refute these conclusions. An informal argument for how this could be done with their model is relegated to our concluding section. The key is to recognize that inter-city interactions, such as trade, make it possible for a small specialized lump of economic activity to relocate away from a large agglomeration while continuing to benefit it and benefit from it. In other words, trade reduces the need for large home–market agglomerations, favoring the emergence of small settlements.

Our next section presents the basic model with costly trade and two industries we are using. Section 3 describes the growth paths that we will study in a two-city economy and Section 4 proves our results. Key technical details are relegated to Appendices A and B. Section 5 is the concluding section.

2. A model with inter-city and intra-city externalities

The basic model structure is first presented for the case of two *identical* cities, each producing manufactures as well as differentiated services with the latter shared with the other city through trade. Then, the case of a single city that uses only its locally produced services is equivalent to setting the cost of trading the services to be infinite. Then manufacturers in each city find it prohibitively costly to import services from the other, and each city is perfectly isolated. In Section 3, we will start from such a single city and we will examine how and when a new city with which services are traded will emerge as population grows.

2.1. A system of two symmetric diversified cities

Because services are traded, the model incorporates both inter-city and intra-city externalities that arise entirely from trading.[3] We will use our model to study structural transitions on the growth path of this two-city system but in this section, we examine equilibrium in a city system when both cities are identical in equilibrium. Each city produces the same manufacture (x) under constant returns and can export it directly to the rest of the world, importing another manufacture (y) from the rest of the world. Hence, manufactures are not traded domestically.[4] Each city also produces domestically tradable services that are intermediate inputs in manufacturing. These are produced under increasing returns, using local labor only. Manufacturing uses labor, supplied locally, and all the services supplied in both cities. Both industries are competitive in the local labor market and pay the same wage.

For manufactures, we use the Ethier (1982) production function: $X = H_x^u \times \left\{ \left(\sum_{i=1}^{2} \sum_{j=1}^{m_i} z_{\mathrm{d}_{ij}}^{(\sigma-1)/\sigma} \right)^{\sigma/(\sigma-1)} \right\}^{1-u}, 0 < u < 1$ and $\sigma > 1$, where X is the aggregate manufactures output of the city, H_x is the local labor input, $z_{\mathrm{d}_{ij}}$ is the demand for the jth service produced in city i, and m_i is the number of service firms in city i. We assume that services sold in the same city incur no transport cost, but manufacturers pay the

[3] As such it differs from that of Abdel-Rahman and Fujita (1993) who impose black-box non-pecuniary economies of scope among industries in order to achieve specialization or diversification.

[4] For a model in which both manufactures and services are traded at cost between cities, see Anas and Xiong (2003).

transportation cost for the services they buy from the other city. This cost takes the iceberg (Samuelson) form: only a fraction τ $(0<\tau<1)$ of the good moved actually arrives at the destination while the rest "melts" in transit. So if q is the price of a service produced and used in city i, the effective (after-transport) price of this service in another city is q/τ.

To simplify further, we will assume that the service industry is in symmetric Nash equilibrium. Then, let m be the number of services produced in a city, let z_{d_i} be city i's demand for each service produced in city i, and $z_{\mathrm{d}_{-i}}$ be city i's demand for each service produced in the other city. The production function of each service firm is $H_z=f+cz_s$, where H_z is total labor input, f is fixed labor input, c is marginal labor input, and z_s is output.

The equilibrium production configuration of each city is determined by Eqs. (2.1)–(2.8) below, which must be solved for w, q, m, H_x, H_z, z_s, z_{d_i}, $z_{\mathrm{d}_{-i}}$. Exogenous parameters are P_x, τ, u, σ, f, c. Aggregate manufactures output is $X=m^{\sigma(1-u)/(\sigma-1)}H_x^u\times\{[z_{\mathrm{d}_i}^{(\sigma-1)/\sigma}+z_{\mathrm{d}_{-i}}^{(\sigma-1)/\sigma}]^{\sigma/(\sigma-1)}\}^{1-u}$. For now, we also take H_c, the total labor supplied to each city, as given. For the two cities to be in equilibrium, we will later see that aggregate population will be allowed to migrate freely between cities, supplying labor to one or the other city, until utilities are equalized.

$$P_x\frac{uX}{H_x}=w, \tag{2.1}$$

$$P_x\frac{(1-u)X}{mz_{\mathrm{d}_i}^{\frac{\sigma-1}{\sigma}}+mz_{\mathrm{d}_{-i}}^{\frac{\sigma-1}{\sigma}}}z_{\mathrm{d}_i}^{-\frac{1}{\sigma}}=q, \tag{2.2}$$

$$P_x\frac{(1-u)X}{mz_{\mathrm{d}_i}^{\frac{\sigma-1}{\sigma}}+mz_{\mathrm{d}_{-i}}^{\frac{\sigma-1}{\sigma}}}z_{\mathrm{d}_{-i}}^{\frac{-1}{\sigma}}=\frac{q}{\tau}. \tag{2.3}$$

$$q=\frac{\sigma c}{\sigma-1}w \tag{2.4}$$

$$z_\mathrm{s}=\frac{f(\sigma-1)}{c} \tag{2.5}$$

$$H_z=f\sigma \tag{2.6}$$

$$z_{\mathrm{d}_i}+\frac{z_{\mathrm{d}_{-i}}}{\tau}=z_\mathrm{s}, \tag{2.7}$$

$$H_c=H_x+mH_z. \tag{2.8}$$

Eqs. (2.1)–(2.3) are the first order conditions of a competitive manufacturer's profit maximization problem. They state that the inputs (labor, local services and services from the other city) are paid the value of their marginal product. As already noted, these aggregate to the city-industry level because of constant returns. Eqs. (2.4)–(2.6) describe the Chamberlin (1933) equilibrium in the service sector for each service firm. It is characterized by marginal revenue equaling marginal cost (resulting in a markup over marginal cost) while each firm makes no profit after covering fixed costs, because of unrestricted entry. Let E be the price elasticity of manufacturers' demand for a service. The markup condition is $q(1-(1/E))=wc$. Assume that m is large enough such that the effect of a price change by one service provider has a small effect on the prices charged by other service providers, then it can be proved that E is approximately equal to σ.[5] With this assumption, the markup condition gives Eq. (2.4). Substituting q/w from Eq. (2.4) into the zero profit condition, $qz_{\mathrm{s}}-w(f+cz_{\mathrm{s}}=0)$, the service output of the firm can be obtained as Eq. (2.5) and is independent of the output price. Also, substituting from Eq. (2.5) into the firm's production function, $H_z=f+cz_{\mathrm{s}}$, we get the firm's labor demand as a constant given by Eq. (2.6). Meanwhile, Eqs. (2.7) and (2.8) state that the city's service and labor markets must clear. The following is the completed closed form solution where $\delta\equiv 1+\tau^{\sigma-1}$: $z_{\mathrm{d}_i}=f(\sigma-1)/(\delta c)$, $z_{\mathrm{d}_{-i}}=\tau^{\sigma}f(\sigma-1)/(\delta c)$, $z_{\mathrm{s}}=f(\sigma-1)/c$, $m=(1-u)H_c/(f\sigma)$, $H_x=uH_c$, $H_z=f\sigma$, $w=\lambda P_x\delta^{(1-u)/(\sigma-1)}H_c^{(1-u)/(\sigma-1)}$, $q=(\sigma c/(\sigma-1))\lambda P_x\delta^{(1-u)/(\sigma-1)}H_c^{(1-u)/(\sigma-1)}$, and $\lambda\equiv[(1-u)/(f\sigma)]^{(\sigma-u)/(\sigma-1)}\{uc\sigma/[(1-u)(\sigma-1)]\}^u f(\sigma-1)/c$.[6] Substituting for m and the inputs into the city's manufacturing production function, we obtain the city's aggregate manufactures output as a function of the labor supplied to the city: $X=\lambda\delta^{(1-u)/(\sigma-1)}H_c^{(\sigma-u)/(\sigma-1)}$.

Note two important properties of this output supply: (1) *Intra-city pecuniary externality*: the marginal product of labor increases with the labor supplied to the city: $\partial^2X/\partial H_{\mathrm{c}}^2>0$. So even though manufacturing is constant returns, aggregate manufacturing output is a convex increasing function of the labor supplied to the city; (2) *Inter-city pecuniary externality*: the marginal product of labor rises also with the presence of a second city (i.e. as $\delta>1$, because $\tau>0$): $\partial^2X/\partial H_c\partial\delta>0$. The source of both of these external returns to scale is the technological bias for the variety of service inputs: When services have a larger cost share $(1-u)$ or when the price elasticity of input demand for services (σ) is smaller, then manufacturing exhibits a higher technological bias for service variety and the external scale economies are stronger. Also, the larger is τ, the cheaper are imported services and the more cities benefits from each other: $\partial^2X/\partial H_c\partial\tau>0$).

2.2. Internal structure of a city

Cities are assumed circular with all producers located in the CBD (the central point) and using no land.[7] Resident-workers, who have identical tastes, consume the domestic manufacture, x, and another manufacture, y, imported from the rest of the world and selling for the exogenous price P_y. Both x and y can be traded with the rest of the world from any CBD. Hence, there is no reason for x to be traded domestically between cities. Each worker is endowed with one unit of time, and incurs a time-cost when commuting to the CBD to work. A worker who resides at distance r from the CBD, offers $H(r)=1-sr$ units of labor where s is the exogenous unit time-cost of commuting. Each worker

[5] See Dixit and Stiglitz (1977) for the proof.

[6] The result $H_x=uH_c$ is perhaps the trickiest. But it is obviously true recalling that all manufacturing revenue ends up as wages either directly in manufacturing or indirectly in the service sector. So by the symmetry, $P_xX=wH_c$. Plug this into Eq. (2.1) and $H_x=uH_c$.

[7] For a model which treats the clustering of producers in subcenters within a city, see Anas and Kim (1996).

consumes one unit of land. Then, the city's radius is $r_f = N_c^{1/2}\pi^{-1/2}$ where N_c is the city's population. We assume that all who join the city must supply labor. Hence, the maximum possible radius is $1/s$ where labor supply falls to zero. Then, the maximum feasible city population is $N_c^{\max} = \pi/s^2$. For $0 < r_f < 1/s$, the city's labor supply is $H_c = \int_0^{r_f} 2\pi r H(r)\mathrm{d}r = N_c\left(1 - kN_c^{\frac{1}{2}}\right)$ with $k = \frac{2s}{3\sqrt{\pi}}$. The net income of a worker is $I_c(N_c) = (1 - sr)w + (TDR/N_c) - R(r)$ where w is the wage, $R(r)$ is the land rent at radius r and TDR is the aggregate differential land rent shared equally by the workers.[8] Rent falls to zero at the city's edge. In equilibrium, workers are indifferent about residence location. Hence, for any two locations $0 \le r \le r_f$, disposable incomes must be equal: $(1 - sr)w - R(r) = (1 - sr_f)w$, and $R(r) = s(r_f - r)w$. The total location cost (rent plus commuting cost) of any resident is $\mathrm{LC} \equiv R(r) + wsr = wsr_f = \frac{wsN_c^{\frac{1}{2}}}{\sqrt{\pi}}$. The differential rent is $\mathrm{TDR} = \int_0^{r_f} 2\pi r R(r)\mathrm{d}r = \frac{wsN_c^{\frac{3}{2}}}{3\sqrt{\pi}}$. Note that LC rises three times as fast with city population than does per capita TDR. The net income spent on x and y is $I_c(N_c) = w + (TDR/N_c) - \mathrm{LC} = (1 - kN_c^{1/2})w = (1 - kN_c^{1/2})\lambda P_x \delta^{(1-u)/(\sigma-1)} H_c^{(1-u)/(\sigma-1)}$ which follows by recalling from Section 2.1 that $w = \lambda P_x \delta^{(1-u)/(\sigma-1)} H_c^{(1-u)/(\sigma-1)}$. Workers have Cobb–Douglas tastes: $U = x^\alpha y^\beta l$, where x and y are the domestic and imported manufactures and $l = 1$ is the fixed lot size. Then, indirect utility is:

$$U(N_c) = P_x^{-\alpha} P_y^{-\beta} I_c(N_c) = P_x^{-\alpha} P_y^{-\beta}\left(1 - kN_c^{\frac{1}{2}}\right)w = P_x^{-1\alpha} P_y^{-\beta}\left(1 - kN_c^{\frac{1}{2}}\right)\lambda\delta^{\frac{1-u}{\sigma-1}} H_c^{\frac{1-u}{\sigma-1}} \tag{2.9a}$$

Substituting in for H_c and δ from Section 2.1, Eq. (2.9a) becomes:

$$U(N_c) = \lambda P_x^{1-\alpha} P_y^{-\beta}[1 + \tau^{\sigma-1}]^{\frac{1-u}{\sigma-1}} N_c^{\frac{1-u}{\sigma-1}}\left(1 - kN_c^{\frac{1}{2}}\right)^{\frac{\sigma-u}{\sigma-1}}. \tag{2.9b}$$

The intra- and inter-city productivity effects initially cause utility to increase with the city's population. However, the increase in city size causes average location cost to increase also. The autarkic efficient city size occurs at $U'(N_c^*) = 0$. Where these two effects balance at the margin. Solving: $N_c^* = [9\omega^2/(9\omega^2 + 6\omega + 1)N_c^{\max}$, where $\omega = (1 - u)/(\sigma - 1)$. Note that with $\sigma > 1$, as $u \to 1$ service variety plays a vanishing role. Then $\omega \to 0$, and optimally sized cities would be infinitesimal, because there would be no positive externalities from localization. This case corresponds to that of "home manufacturing" (backyard economy) without services.

3. Characterizing the growth paths

As explained at the outset, the aim of this paper is to study the transitional dynamics of how specialized cities evolve, as in the formation of edge cities or sub-centering in

[8] We assume that a local government collects and redistributes the TDR equally among the city residents. If a resident moves to another city, he gives up his ownership of land shares in the city from which he came and starts collecting his shares in the city in which he arrived. Owning equal shares of land in all cities would create a second type of inter-city pecuniary externality that is ignored here but is treated in the work of Pines (2000), in a different context. But he ignores the trading externality we treat here.

metropolitan areas. To do so, we will assume a *benchmark* initial condition: the existence of a single diversified city with both manufactures and services. We are also assuming that there is only one predetermined second location (e.g. a trading node) where, under laissez-faire, atomistic defectors can set up new activity. We will show that such atomistic defection can result in the new secondary city being specialized in either manufacturing or services and that such a specialized city can self-organize at the optimal time without government intervention or developer action. More precisely, as the exogenous aggregate population, $N(t)$, grows monotonically and continuously from zero, there is a moment in time when the new city should be set up. Let $N_a(t)$ and $N_b(t)$ denote the populations of the two cities (with a the first and b the new) such that $N_b(t)=N(t)-N_a(t)$ and $U_a(N_a(t), N_b(t))=U_b(N_a(t), N_b(t))$. Since initially there is one city, there is a stretch when $U_a(N(t), 0)>U_b(N(t), 0)$, until some t^* when $U_a(N_a(t^*), N_b(t^*))=U_b(N_a(t^*), N_b(t^*))$ with $N_b(t^*)\geq 0$. Thereafter, for all $t>t^*$ $U_b(N_a(t), N_b(t))>U_a(N_a(t), N_b(t))$ with $N_b(t)>0$.

3.1. Laissez-faire

Under laissez-faire, the emergence of the second city starts by the atomistic defection of a small mass of agents from the existing city, namely $N_b(t^*)\cong 0$. If the initial defection involves the manufacture industry, and since firm size under constant returns is indeterminate, the small mass consists of an infinitesimally small "firm mass" and its labor/population. If the initial defection involves the services industry, the small mass is a single firm and its associated population. Since each service provider employs $H_z=f\sigma$ units of labor, the initial size of such a service city is the population that solves $f\sigma=N_b(t^*)(1-kN_b(t^*)^{1/2})$, in order to clear that city's labor market according to Section 2.2. Because we are interested in a service industry with many firms, we will make $N_b(t^*)$ small enough by choice of the fixed cost, f, so that any lumpiness is rendered uninteresting.

3.2. Planning

Under this process, a planner/developer is assumed to setup the new city whenever doing so yields higher utility than does the one-city system. The planner/developer has the advantage that, if necessary, he can make an optimized lumpy relocation of economic activity from city 1 to city 2. Therefore, the planner/developer is needed when $N_b(t^*)$, the initial population of the new city, is lumpy (greater than the small mass or single small service firm that defects under laissez-faire). We assume that the planner is concerned only with the efficient timing issue and not with the market failure that arises from monopolistic competition in the service sector. Hence, our efficiency results are second-best. However, correcting for the pricing markup due to the imperfect competition would not change any of our qualitative results with respect to the timing of city formation.[9]

Table 1 describes the alternative growth patterns that will be studied in this paper for at most two cities. Extension to more than two cities is straightforward intuitively and the results follow from the two city case.

[9] Xiong (1998) shows that a sales subsidy per unit of service output sold will correct this market failure in the present model. It would be easy to incorporate this correction but it is a nuisance.

Table 1
The four examined growth patterns of the two-city system

	Industry structure in old city (Diversified)	Industry structure in new city (Diversified or specialized)	Properties
Pattern 1	Manufactures and services	None	
Pattern 2	Manufactures and services	Manufactures and services (Dominates if inter-city transport cost of services is sufficiently high. i.e. τ is low.)	(1) Emerges too late under laissez-faire. (2) Optimal setup is unstable. (3) Critical mass migration. (4) Identical with old city.
Patterns 3 and 4	Manufactures and services	Services in Pattern 3 (If $u<0.5$ or $u \geq 0.5$ and σ is close enough to 1.); Manufactures in Pattern 4 (If $u>0.5$ and σ is sufficiently larger than 1.)	(1) Emerges at the optimal time under laissez-faire. (2) Grows from infinitesimal. (3) Smaller than old city. (4) Lower wage than old city.

Under Pattern 1, a single diversified city persists forever and a new city never emerges. We will examine three more possible patterns of growth. We assume that under Pattern 2, when a new city emerges, it is also diversified (containing both manufactures and services) and identical to the first. This pattern will serve to illustrate the lumpy adjustment that can only be realized by having developers. Under Pattern 3, the new city specializes in services exporting them to the older diversified (parent) city, and under Pattern 4, it specializes in manufactures, importing services from the parent. We will denote the equalized utility of consumers under each pattern as $U^1(N(t))$, $U^2(N(t))$, $U^3(N(t))$, $U^4(N(t))$, where $N(t)$ is aggregate not city population. City populations will be denoted as $N_a(t)$, $N_b(t)$. It is a major technical contribution of this paper (see also Anas and Xiong, 2003) that the utility path of a population partition $N_a(t)$, $N_b(t)$ is derived as a function of total population, $N(t)$. As can be seen from the Appendices A and B, the algebra is daunting but helps establish results analytically that, otherwise, could only be established numerically. For any total population $N(t)$, one can observe directly which pattern's utility path dominates the others, when in time it begins to do so and under what parameter values.

3.3. Pattern 1 (One city)

Under Pattern 1, the production configuration corresponds exactly to that derived in Section 2 (modified only by setting $\tau=0$, since there is no trade). The utility path, from Eq. (2.9b), is our benchmark:

$$U^1(N(t)) = \lambda P_x^{1-\alpha} P_y^{-\beta} N(t)^{\frac{1-u}{\sigma-1}} \left[1 - kN(t)^{\frac{1}{2}}\right]^{\frac{\sigma-u}{\sigma-1}}. \tag{3.1}$$

Trading cost is irrelevant since there is no other city to trade with. We know from Section 2 that there is an autarkic efficient city size. Hence, as population grows, the single diversified city eventually becomes overpopulated and utility declines as shown in Fig. 1 until the maximum total population $N(t)=\pi/s^2$ is reached (see Section 2.2).

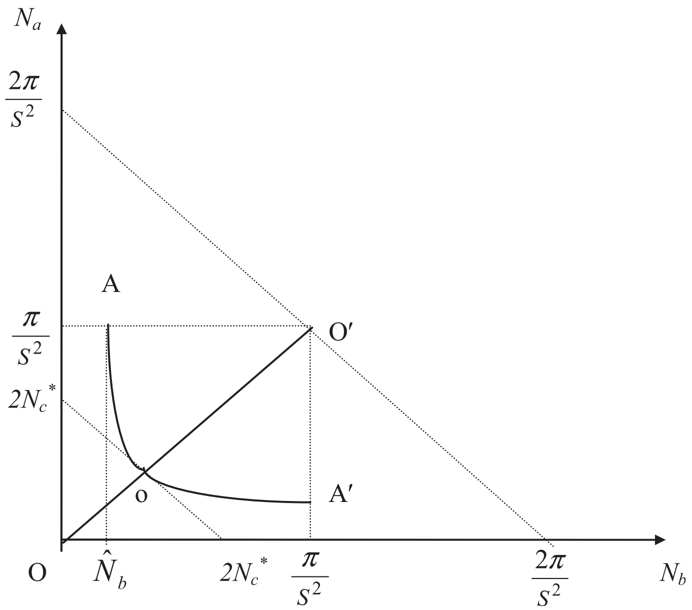

Fig. 1. Possible equilibria of a two-city system when $\tau = 0$. Oo is the unstable symmetric locus, oO′ is the stable symmetric locus and AA′ is the unstable asymmetric locus.

3.4. Pattern 2 (Two identical and diversified cities)

When both cities are diversified, there are two possibilities: at equilibrium, the new and old cities can be identical in size or unequal (asymmetric) in size. The latter case is Pattern 5 (see below). Fig. 1 is a diagrammatic illustration for any total population $0 < N(t) < 2\pi/s^2$. In the figure, $N_a(t)$ and $N_b(t)$ denote the populations of cities a and b, respectively. The vertical and horizontal axes correspond to equilibria under Pattern 1 in which all population is concentrated in one city. The locus OO′ is that of equal-utility equilibria in which the two diversified cities are identical (as assumed in this pattern) and the locus AA′ is that of equal-utility equilibria in which the two diversified cities are of unequal sizes. For any $N(t)$, the equal-utility partitions correspond to the intersections between $N_a(t) + N_b(t) = N(t)$ and AA′ or OO′ or the two axes. Assume that the one-city equilibria of Pattern 1 are locally stable. Then, for $N(t) > 2N_c^*$, the symmetric equilibrium (on oO′) will be locally stable, while the asymmetric equilibria (on AA′) will be unstable.[10] In the case shown in Fig. 1, when city a is maximally sized (i.e. $N_a(t) = \pi/s^2$), equilibrium requires that $N_b(t) = \hat{N}_b > 0$. This means that the transition from a one-city equilibrium to a two-city equilibrium requires a lumpy adjustment of population and cannot be smooth. Under atomistic defection, city b cannot emerge because

[10] This follows Anas (1992), except that in that model, there was no maximum city size, because lot size decreased with city population. Here, the fixed lot size assumption forces a maximum size as we saw.

$N_b(t) = \hat{N}_b > 0$ and a developer is needed for b to emerge at the optimal time as is always assumed in Henderson's papers.

The trade relationship on OO′ is symmetric: each city imports the same amount of services from the other at the same price. Substitute $N_c(t)$=$N(t)$/2 into Eq. (2.9b) and we get the equilibrium utility path under Pattern 2:

$$U^2(N(t)) = \lambda P_x^{1-\alpha} P_y^{-\beta} \left(\frac{1+\tau^{\sigma-1}}{2}\right)^{\frac{1-u}{\sigma-1}} N(t)^{\frac{1-u}{\sigma-1}} \left[1 - \frac{k}{\sqrt{2}} N(t)^{\frac{1}{2}}\right]^{\frac{\sigma-u}{\sigma-1}}. \tag{3.2}$$

Note that while $U^1(N(t))$ does not depend on inter-city transportation, $U^2(N(t))$ is an increasing function of τ. Comparing Eqs. (3.1) and (3.2), it is easy to see that when $\tau = 1$ (services can be transported without cost), then the utility of two diversified cities is above that of a single city for any $N(t)$. However, with sufficiently high inter-city transport costs for services (τ close enough to 0), the utility of a single city is above that of two cities for low $N(t)$. From Fig. 2, we can also see that the maximum utility level achieved under Pattern 2 is higher than under Pattern 1, and that the peak occurs when each city is of autarkic optimal size. Under Pattern 2, cities can achieve a higher utility by exploiting the inter-city externalities. Having one city is better when $N(t)$ is low, because the commuting externality will be low while a single city will not need to import services. However, once city size becomes sufficiently large, the city can be split into two, reducing the commuting externalities while still capturing most of the productivity externalities through the inter-city trade of services. The higher is τ, the earlier the point in time when such a splitting of the one city becomes optimal. Assuming $0 \le \tau < 1$, the time when Pattern 2 first reaches the same utility level as Pattern 1 can be solved from $U^1(N(t_2)) = U^2(N(t_2))$.From this, the total population at which the planner should set up the new diversified city is: $N(t_2) = \frac{1}{k^2}\left(\frac{1-\theta}{1-2^{-\frac{1}{2}}\theta}\right)^2$ where, $\theta \equiv [(1+\tau^{\sigma-1})/2]^{(1-u)/(\sigma-u)}$. Note also that under this pattern, growth would continue until $N(t)$=$2\pi/s^2$: each city would become maximally sized, unless a third city could be established which is beyond our scope here.

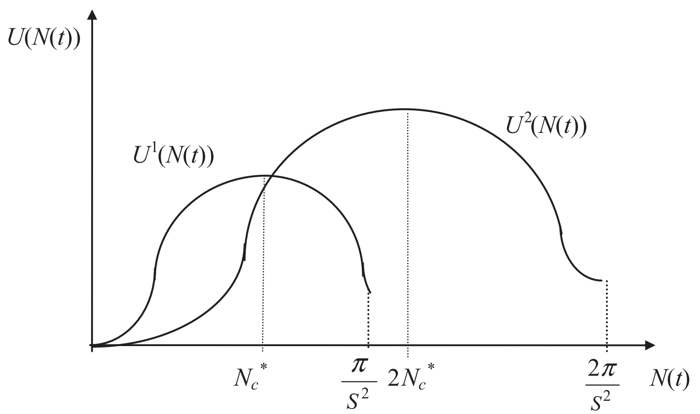

Fig. 2. Utility paths of Patterns 1 and 2.

3.5. Pattern 3 (New city specialized in services)

Under Pattern 3, the old city continues to be diversified, while the new city specializes in services. Edge cities that specialize in services are a good example. They are common in most large metropolitan areas all over the world. In this pattern, in our model, the new city sells all of its services to the old and imports both manufactures from the rest of the world through its CBD. Given the size of the old city N_a and the size of the new city N_b, we can derive the utility levels of both cities in the two-city system (see Appendix A for the algebra):

$$U_a^3(N_a, N_b) = \lambda P_1^{1-\alpha} P_2^{-\beta}\left(1 - kN_a^{\frac{1}{2}}\right)\left[N_a\left(1 - kN_a^{\frac{1}{2}}\right) + \tau^{\frac{\sigma-1}{\sigma}} N_b\left(1 - kN_b^{\frac{1}{2}}\right)\right]^{\frac{1-u}{\sigma-1}} \tag{3.3}$$

$$U_b^3(N_a, N_b) = \lambda P_1^{1-\alpha} P_2^{-\beta}\tau^{\frac{\sigma-1}{\sigma}}\left(1 - kN_b^{\frac{1}{2}}\right)\left[N_a\left(1 - kN_a^{\frac{1}{2}}\right) + \tau^{\frac{\sigma-1}{\sigma}} N_b\left(1 - kN_b^{\frac{1}{2}}\right)\right]^{\frac{1-u}{\sigma-1}} \tag{3.4}$$

Then, the ratio of utilities as a function of the city populations N_a and N_b is:

$$\frac{U_a^3}{U_b^3} = \frac{\left(1 - kN_a^{\frac{1}{2}}\right)w_a}{\left(1 - kN_b^{\frac{1}{2}}\right)w_b} = \frac{\left(1 - kN_a^{\frac{1}{2}}\right)}{\left(1 - kN_b^{\frac{1}{2}}\right)\tau^{\frac{\sigma-1}{\sigma}}} \tag{3.5}$$

Given $N(t)$, equilibrium city sizes satisfy $N_a(t) + N_b(t) = N(t)$ at each t, while the utility ratio given by Eq. (3.5) is 1. The important thing to note here is that—unlike Pattern 2—the new city can start out as infinitesimally small and match the utility of the old city but not before total population $N(t)$ reaches a critical level. To verify this, set $N_b = 0$ and $N_a = N(t)$ in Eq. (3.5) and see that it will satisfy $U_a^3/U_b^3 = 1$, when $N(t_3) = (1 - \tau^{(\sigma-1)/\sigma})^2/k^2$. A tiny new service city emerges at time t_3 by atomistic defection under laissez-faire and without any help from planners. From Eq. (3.5), we can see the properties of the growth path, recalling that $0 < \tau < 1$:

(1) After t_3, both cities grow continuously as the national population grows.
(2) The new specialized city always remains smaller than the old diversified city.
(3) When inter-city transportation cost falls (τ increases), the new specialized city grows while the parent city shrinks.
(4) Since $N_b(t) < N_a(t)$, we need $w_b < w_a$ for Eq. (3.5) to hold. Namely, in the less crowded city where average location cost is lower, producers can afford to pay a lower wage and still attract workers.

As shown in Appendix A, we can get the sizes of the two cities and the utility path:

$$N_b^3(t) = \left\{\frac{[(1+\phi^2)k^2N(t) - (1-\phi)^2]^{\frac{1}{2}}}{(1+\phi^2)k} - \frac{\phi(1-\phi)}{(1+\phi^2)k}\right\}^2 \text{ and}$$
$$N_a^3(t) = N(t) - N_b^3(t), \tag{3.6}$$

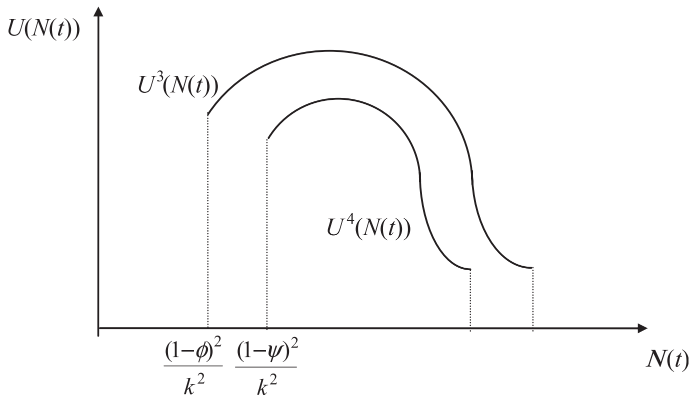

Fig. 3. Utility paths of Patterns 3 and 4.

where $\phi = \tau^{(\sigma-1)/\sigma}$, and

$$U^3(N(t)) = \lambda P_x^{1-\alpha} P_y^{\beta} N(t)^{\frac{1-u}{\sigma-1}} \left\{ \frac{\phi(\phi+1)}{\phi^2+1} - \frac{\phi[(\phi^2+1)k^2 N(t) - (1-\phi)^2]^{\frac{1}{2}}}{\phi^2+1} \right\}^{\frac{\sigma-u}{\sigma-1}} \tag{3.7}$$

Note that as in the previous two patterns, $U^3(N(t))$ eventually decreases with $N(t)$ because the commuting diseconomies at the city level eventually dominate the economies from service variety. $U^3(N(t))$ is plotted in Fig. 3.

3.6. Pattern 4 (New city specializes in manufactures)

In Pattern 4, the old city is again diversified, while the new one produces only manufactures. In the US, Silicon Valley (Bay area) or Schaumburg (Chicago area) are possible examples of manufacturing satellites attached to more diversified cities. Under this pattern, the new city imports services (e.g. financial) from the old. Given the sizes of the old and the new city, N_a and N_b, we can derive the utility levels of both cities in the two-city system (see Appendix B for the algebra):

$$U_a^4(N_a, N_b) = \lambda P_1^{1-\alpha} P_2^{-\beta} \left(1 - kN_a^{\frac{1}{2}}\right) \left[N_a\left(1 - kN_a^{\frac{1}{2}}\right) + \tau^{\frac{1-u}{u}} N_b \left(1 - N_b^{\frac{1}{2}}\right)\right]^{\frac{1-u}{\sigma-1}} \tag{3.8}$$

$$U_a^4(N_a, N_b) = \lambda P_1^{1-\alpha} P_2^{-\beta} \tau^{\frac{1-u}{u}} \left(1 - kN_b^{\frac{1}{2}}\right) \left[N_a\left(1 - kN_a^{\frac{1}{2}}\right) + \tau^{\frac{1-u}{u}} N_b \left(1 - N_b^{\frac{1}{2}}\right)\right]^{\frac{1-u}{\sigma-1}} \tag{3.9}$$

The utility ratio of the two cities is now:

$$\frac{U_a^4}{U_b^4} = \frac{\left(1 - kN_a^{\frac{1}{2}}\right) w_a}{\left(1 - kN_b^{\frac{1}{2}}\right) w_b} = \frac{\left(1 - kN_a^{\frac{1}{2}}\right)}{\left(1 - kN_b^{\frac{1}{2}}\right) \tau^{\frac{1-u}{u}}} \tag{3.10}$$

The critical total population at which the manufacturing satellite can start out as infinitesimally small (at time t_4) is now $N(t_4)=(1-\tau^{(1-u)/u})^2/k^2$. The four properties seen for Pattern 3 again hold on the equilibrium growth path:

(1) After t_4, both cities grow continuously with total population.
(2) The new manufacturing satellite remains smaller than the diversified parent city.
(3) A lowering of service transport cost increases the size of the satellite while shrinking the parent.
(4) Since $N_b(t)<N_a(t)$, we need $w_b<w_a$ for Eq. (3.10) to hold. Namely, in the less crowded city where average location cost is lower, producers can afford to pay a lower wage and still attract workers.

From Appendix B, the city sizes and the utility path are:

$$N_b^4(t)=\left\{\frac{[(1+\psi^2)k^2N(t)-(1-\psi^2)]^{\frac{1}{2}}}{(1+\psi^2)k}-\frac{\psi(1-\psi)}{(1+\psi^2)k}\right\}^2\text{, and}$$
$$N_a^4(t)=N(t)-N_b^4(t), \tag{3.11}$$

where $\psi\equiv\tau^{(1-u)/u}$, and

$$U^4(N(t))=\lambda P_x^{1-\alpha}P_y^{\beta}N(t)^{\frac{1-u}{\sigma-1}}\left\{\frac{\psi(\psi+1)}{\psi^2+1}-\frac{\psi[(\psi^2+1)k^2N(t)-(1-\psi)^2]^{\frac{1}{2}}}{\psi^2+1}\right\}^{\frac{\sigma-u}{\sigma-1}} \tag{3.12}$$

Note again that $U^4(N(t))$ eventually decreases with $N(t)$ because of the growing diseconomies of commuting.

The equilibrium utility paths of Patterns 3 and 4 are plotted in Fig. 3. $U^4(N(t))$ and $U^3(N(t))$ have the same shape. The differences between them are that the utility levels and the starting populations are different depending only on the values of u and σ. If $u<1/(2-\sigma^{-1})$, Pattern 3 dominates Pattern 4. However, if $u<1/(2-\sigma^{-1})$, Pattern 4 dominates Pattern 3. Recall that $\sigma\in(1,\ \infty)$. Hence, the right side of these inequalities is in the interval (0.5,1). Suppose that a specialized new city is set up at the first opportunity. Then, if the share of services in manufacturing, $1-u$, is larger than 0.5 ($u<0.5$), then the new city always specializes in services. Alternatively, for any $1>u\geq0.5$, if the price elasticity of the demand for services, σ, is sufficiently close to its lower limit 1, then the new city again specializes in services. The intuition is that the lower price elasticity makes it possible for service firms (which have pricing power) to defect from the parent city, thus raising the delivered price but still surviving. At city b's birth, the ratio of the delivered price paid at a for services from b to the price of local services at a is $(q_b/\tau)/q_a=(w_b/\tau)/w_a=\tau^{-1/\sigma}>1$. Hence, as as $\sigma\to1$ (less elastic demand), the ratio increases. If $1>u>0.5$ and σ is sufficiently bigger than 1, then the new city specializes in manufactures. All this is shown in Fig. 4.

3.7. Pattern 5 (Two asymmetrically diversified cities)

Under this pattern, which we will not examine formally, the second city is diversified but (unlike Pattern 2) is smaller than the first. The emergence of such a city *b* at the optimal time under laissez-faire would require the coordinated defection of a small mass of manufacturers together with a single service firm (a small coalition). For this to be possible, the asymmetric locus AA′ in Fig. 1 must cut the two axes at some $N_a = N_b < \pi/s^2$, and be concave to the origin. The algebra needed to derive the utility path for this pattern is extremely daunting. Fortunately, examination of such a pattern would only reinforce our results, since it is a natural extension of either Pattern 3 or 4 which are sufficient to rule out the two dismal results noted by Henderson and Becker (2000). To see this, reason as follows. Suppose that initially one service firm defects from *a* to *b*. Once this happens, a small mass of manufacturers can immediately join *b*, making it diversified. By moving to *b*, the service firm saves commuting cost for its workers and can thus pay a lower wage to entice them to *b*. Now a small mass of manufacturers should move to *b* to take advantage of the cheaper labor. However, the manufacturers would have to import the service varieties still produced at *a*, and would thus incur higher service importation costs. If the unit cost of trading a service is low, then the benefits of the move should exceed the costs and the two-city asymmetric diversified pattern will be stable. Under Pattern 4, a small manufacturing mass starts city *b*, paying a lower wage but incurring higher service

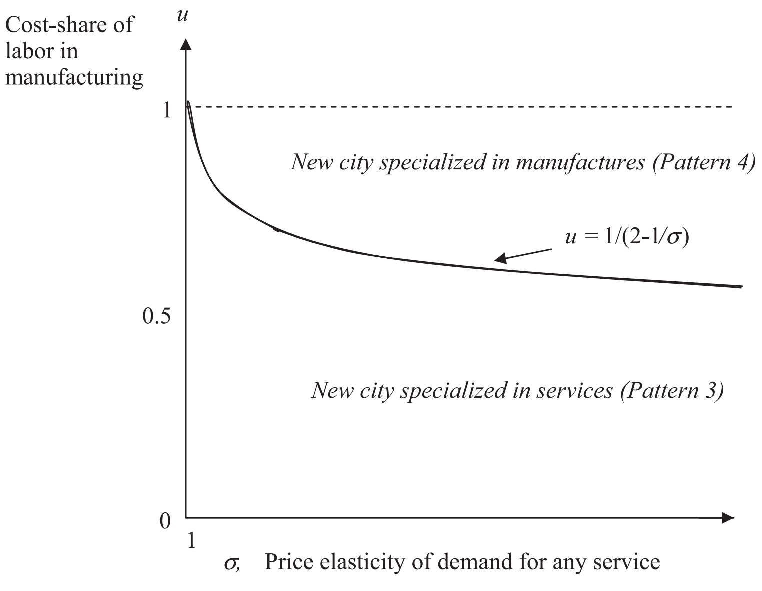

Fig. 4. Specialization regimes in parameter space.

importation costs. More services can join b if the cost of trading with a is low enough. City b should again be stable.

4. Transitions on the growth path

In this section, we compare the utility levels of the four patterns derived above and decide how the two-city system would switch among these patterns under a monotonically and continuously growing $N(t)$. Since there are no adjustment costs, a planner/developer will do the switch at the instant that the new pattern offers higher utility. That is the only action required for timing efficiency. As we will see, a planner is needed only when efficiency calls for a switch from Patterns 1, 3 or 4 to Pattern 2. When efficiency calls for a switch from Pattern 1 to Pattern 3 or 4, this can be achieved by atomistic defection under laissez-faire. As explained earlier, we are not concerned with the correction of the market failure due to the imperfect competition.

Beginning with Pattern 1, we only need to identify which of the last three patterns first passes the utility level of Pattern 1. In other words, which of $N(t_2)$, $N(t_3)$ and $N(t_4)$, derived in Section 3, is the smallest?[11]

Proposition 1 (Existence). *The new city can exist as diversified and identical to the old one, or as smaller, specializing either in manufacturing or in services.*

Proof. We can show that each of $N(t_2)$, $N(t_3)$ or $N(t_4)$ can be the smallest. The smallest of these total populations indicates the earliest opportunity to raise utility by switching patterns. Which pattern first exceeds the utility of Pattern 1 depends on the values of σ and u, and on $1/\tau$, the inter-city unit transportation cost for services. The value of the unit commuting cost s (which, in this model, does not increase with city size) does not affect the relative timing of the three patterns. We can identify three cases.

(a) *New city diversified and identical to first* (Pattern 1 $\rightarrow$ Pattern 2; see Fig. 5): When the inter-city transportation cost of services is sufficiently high (τ close enough to its lower limit) the new city should be set up as a diversified city. This is easy to verify by seeing that $N(t_2) < N(t_3) = N(t_4)$ when τ is zero. The intuition is as follows: a specialized new city must rely heavily on the inter-city trade of services. Hence, a high transport cost for services does not favor specialization. But a diversified city has half of all services locally and can heavily substitute these for services imported from the other city. Hence, a high transport cost for services favors diversification in the second city.

Now, to move on to the proofs of the next two cases, assume that transportation cost is sufficiently low (τ close enough to 1), so that the new city will be specialized.

(b) *New city specializes in services* (Pattern 1 $\rightarrow$ Pattern 3; see Fig. 6): When manufacturers rely heavily on services (small enough u), or the price elasticity of

[11] In what follows below, we will assume that pattern switching occurs before Pattern 1 reaches maximum size. This requires that trading cost not be too high: $\tau > \left[2\left(\frac{\sqrt{2}}{3\sqrt{2}-2}\right)^{\frac{\sigma-u}{1-u}}-1\right]^{\frac{1}{\sigma-1}}$ for Pattern 2 to occur before Pattern 1 reaches maximum size, $\tau > (1/3)^{\sigma/(\sigma-1)}$ for Pattern 3 and $\tau > (1/3)^{u/(u-1)}$ for Pattern 4. If these conditions do not hold, pattern switching occurs immediately after the first city reaches maximum size and involves jumps in utilities. We ignore this situation of jumps.

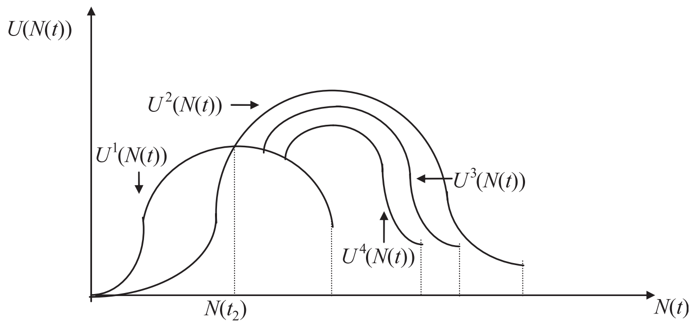

Fig. 5. New city is diversified (Pattern 1 → Pattern 2).

services is small enough (σ close to 1), then it is optimal for the new city to specialize in services. As we saw in Section 3 and illustrated in Fig. 4 the precise condition is $u<1/(2-1/\sigma)$. The intuition was that when services are important inputs for manufactures (small u) given σ, the price elasticity of the demand for services, then service producers can defect from the old diversified city but still experience strong demand from it. Alternatively, given u, the inequality holds by lowering σ. A lower price elasticity of demand for the service increases pricing power and makes the firm more able to defect to the specialized city. Also, as we saw, a specialized new city realizes cost savings because, being smaller, it entails lower average commuting cost, lowering wages.

(c) *New city specializes in manufacturing* (Pattern 1 → Pattern 4): When the elasticity of demand for services is high (large enough) or manufactures do not heavily rely on services (large enough u), it is optimal for the new city to specialize in manufactures. The precise condition is $u>(1/(2-1/\sigma))$. The intuition is the reverse of that for (b). Being away from the old diversified city is unfavorable for manufacturers. However, by being less reliant on services (more reliant on labor), manufacturers can defect to the smaller satellite, pay lower wages and use the savings to import services from the parent city. □

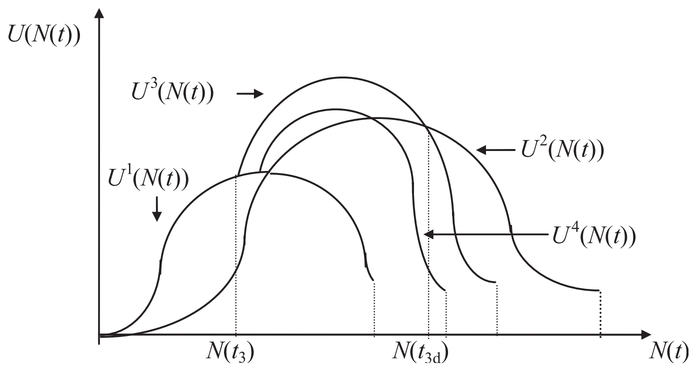

Fig. 6. New city specializes in services (Pattern 1 → Pattern 3).

Proposition 2 (Stability and efficiency of specialization). *The new city is always stable at its establishment when it is a specialized city. The new city is unstable at its establishment when it is a diversified city identical to the old unless the cost in inter-city trade is low enough (τ close to 1).*

(a). Instability of symmetric diversified cities: *When τ is small, two identical diversified cities are unstable when they are set up. In this case, the symmetric equilibrium of Pattern 2 is locally stable only when the total population is sufficiently large.*

Proof. To formally see whether the symmetric equilibrium is locally stable, we perturb it so that one city is slightly larger while the other is smaller than $N(t)/2$. Then, we see whether this perturbed configuration converges back to Pattern 2. The procedure is similar to that of Anas (1992). Here, it suffices to use an informal argument focused more on the underlying intuition. Let us check the stability of Pattern 2 in two special cases.

(1) Prohibitively costly trade ($\tau = 0$): In this case, services are not tradable and the two cities are autarkic. Now we have a case that is formally similar to that examined by Anas (1992). As proved there, the optimal time to switch patterns is when $N_c^* < \tilde{N}(t) < 2N_c^*$ so that $U^1(\tilde{N}(t)) = U^1(1/2\ \tilde{N}(t))$. Pattern 2 is unstable when it is optimal to switch to it. It becomes stable when $N(t) > 2N_c^*$. Fig. 1 illustrates the loci of multiple equilibria in this case of two autarkic cities. The asymmetric and unstable locus (AA′) has been drawn as convex to the origin, which depends on the value of $\omega = (1-u)/(\sigma-1)$. It has also been assumed in Fig. 1 that the maximum city size is more than twice the autarkic optimum size. Note from the figure that two diversified cities cannot exist and be stable before $N(t) = 2N_c^*$. (The portion Oo of the symmetric locus in Fig. 1 is unstable.) The second city emerges later and a relatively big perturbation (a critical size migration) is needed to jog the city-system to the stable (oO′) part of the symmetric locus. Most likely, the second city will emerge exactly when the first city has reached maximum size. At that point, the existing city cannot accommodate more people and any additional growth must go to the new city. Once the number of such migrants just exceeds $\hat{N}_b$ (where $U^1(\hat{N}_b) = U^1(\pi/s^2)$), they achieve utility higher than that of the maximally sized first city and, under laissez-faire, a large catastrophic migration equalizes the sizes of the two cities.

(2) Costless trade ($\tau = 1$): In this case, services are tradable without cost and inter-city externalities are as high as possible. Pattern 2 dominates Pattern 1 for any $N(t)$ and the planner should start with Pattern 2 at time $t = 0$. Now consider the symmetric equilibrium of Pattern 2 for *any* city size. If one worker moves from city a to b, average commuting cost in a is lowered. Meanwhile, there is no disadvantage for a in importing more services from b since this can be done without cost and utility in a rises. Utility in b falls since its average commuting cost increases without gaining any advantage in importing services from a. So if $\tau = 1$, Pattern 2 is always locally stable.

From these two extreme cases, we can see that whether Pattern 2 is locally stable or not depends on the value of τ. Generally speaking, when τ is large enough, the city from which people move out enjoys net efficiency gains and the city people move to suffers net efficiency losses. Therefore, Pattern 2 is likely to be stable with lower inter-city transport costs. □

The above proof expresses the fact that in our present model, laissez-faire cannot engineer a switch at the time when it is optimal to set up a new diversified city identical to the old. As in the work of Anas (1992), planners or developers are desirable to set up the diversified city when it is unstable. In the absence of such planning, the Henderson result holds and the emergence of the new diversified city under laissez-faire would normally be later than is optimal and would occur after a long Malthusian decline. However, as the next part of the proposition claims, this dismal limitation does not exist when it is optimal to set up the new city as a specialized city.

(b). Stability and efficiency of specialization: *Specialized cities are stable, including at the time when they emerge as tiny. (The asymmetric equilibria of Pattern 3 or 4 are always stable.) Hence, the optimal and laissez-faire (or atomistic) timings of the emergence of a new specialized city coincide and developers are not needed.*

Proof. Define the ratio of the two cities' utilities as $Q=((U_a^3)/(U_b^3))$. Note that $Q=\frac{1-k[N(t)-N_b]^{\frac{1}{2}}}{\tau^{\frac{\sigma-1}{\sigma}}\left(1-kN_b^{\frac{1}{2}}\right)}$. It is easy to verify that $\partial Q/\partial N_b>0$. So under Pattern 3, a small increase in the population of city b (at the expense of city a) always increases the utility of city a relative to that of b. Therefore, any migration from a to b will correct itself restoring the initial equilibrium. For Pattern 4, the exponent of τ in Q is $((1-u)/u)$. Steps are the same. Now recall that as $N(t)$ crosses $N(t_3)$ at time t_3 or $N(t_4)$ at time t_4, the first resident of the new city achieves a higher utility than if he went to the old city. Then, no mass migration is required and the new city under Pattern 3 or 4 will emerge at the efficient time as tiny, growing continuously with $N(t)$ thereafter. □

We will now examine the time when the new city *should be* set up (efficient timing). Proposition 3 claims that in the present two-industry model with inter-city trade, it can be optimal to break up an old city before it reaches its peak utility.

Proposition 3 (Specialization avoids Malthusian Traps). *If inter-city transport cost is sufficiently small, the new specialized or diversified city will (should) be set up even before the old diversified city reaches its efficient size.*

Proof. $N(t_2)$, $N(t_3)$ and $N(t_4)$, derived earlier, are each decreasing functions of τ and are each zero when $\tau=1$. So if the inter-city transportation cost is sufficiently small (τ sufficiently close to 1), $N(t_i)<N_c^*=(9\omega^2\pi s^{-2})/(9\omega^2+6\omega+1)$ is possible for $i=2,3,4$ because N_c^* does not depend on τ. Then, it can be optimal to establish the new city under pattern i, before the old diversified city has reached efficient size. □

The intuition for this result is as follows. Breaking up the old diversified city reduces positive intra-city externalities. However, in the present model, these losses can be kept low by the positive inter-city externalities (since services can be imported cheaply by each city) and by the gains from the reduction of commuting diseconomies in each city. Hence, Henderson's claim that growth follows Malthusian cycles need not hold except for very high inter-city trading costs (low inter-city spillovers).

Proposition 4 (Inefficient specialization). *Under laissez-faire (atomistic migrations), the specialized city of Pattern 3 or 4 can emerge even when the economy should have switched to Pattern 2 at an earlier time. Then, the economy gets stuck on Pattern 3 (or 4) and becomes diversified too late.*

Proof. Suppose that $N(t_2) < N(t_3)$ (or $N(t_4)$) as in Fig. 5. Recall that the optimal switch at t_2 requires a critical-mass migration (see Proposition 2(a) and Fig. 1). Since such a lumpy adjustment is unlikely to occur at time t_2 under atomistic laissez-faire, it will be missed unless a developer exists and can act at exactly t_2. The one-city economy will continue to grow sub-optimally until t_3 (or t_4). At that time, a specialized city will emerge and Pattern 3 (or 4) will be followed as a sub-optimal path. □

5. Concluding remarks

Our model reverses the long-standing dismal conclusions reached in the earlier literature and summarized in the introduction. *First*, we explained how specialized small cities can be spawned out of big diversified existing cities without any action by developers. The phenomenon is clearly of historical importance since industries have self-organized out of older cities into smaller rural towns causing local economic growth. A similar dynamic has also occurred at the metropolitan level with the proliferation of employment subcenters or edge cities specialized in manufacturing or in services. *Second*, we showed that atomistic market agents unguided by planners or developers can correctly time the emergence of such specialized secondary cities and subcenters. *Third*, we demonstrated that the asymmetric equilibrium of a large old diversified city with a new small specialized city is stable and can persist for a long time even when the eventual optimum calls for two identical diversified cities to have been established much earlier. *Fourth*, there need not be a Malthusian trap: existing cities do not have to become overpopulated before new cities can emerge.

As promised in the Introduction, we can now present our informal argument about how the results established here can also be derived in a modified model of Henderson and Becker (2000). In their model, cities consist of two types of residents: entrepreneurs who own and operate firms and workers hired to labor in those firms. This is analogous to having manufacturers and services in our model. Their basis for optimal city size is the Marshallian non-pecuniary spillover of trade secrets among entrepreneurs. In our case, the analogous pecuniary externality is the variety of service inputs in manufacturing. Their Marshallian spillover externality provides an incentive for entrepreneurs to locate together raising their productivity. The authors restrict their attention to the case where entrepreneurs and workers *must* locate together in the same city. Under this setup, new cities must be identical to existing ones. This limitation causes them to conclude that the creation of a new city at the optimal time requires lumpy adjustment. Hence, they call for developers. Under laissez-faire, the new city comes too late when the existing city is much larger than the autarkic efficient size (the Malthusian trap). From the perspective of our model, this result is an artifact of the lack of inter-city externalities (trade) in the work of Henderson and Becker (2000). Their model could be modified to our context. To see how, consider as we do only two sites and allow either some entrepreneurs or some workers the choice of locating at the second site and commuting to the first. With such a modification, analogous to trading services in our model, the Henderson and Becker model should (under some parameter values) generate the emergence of a tiny specialized city on the growth path at the efficient time, before the initial city became overpopulated. While the first city would continue to include both entrepreneurs and workers, the specialized city would include only entrepreneurs or only workers.

The empirical significance of our results is that they help explain urban growth as a smooth process (without phases of long Malthusian declines) as supported by the data cited by various authors, notably Dobkins and Ioannides (1995), Eaton and Eckstein (1997) and Black and Henderson (1999). From the perspective of our results, the smoothness of the process is made possible by the spawning of specialized cities much before any Malthusian decline begins to set in. This process was not operative in the earlier analyses. In the works of Anas (1988, 1992) and Pines (2000), models of homogeneous economic activity were used in which case the smooth process cannot arise because there is no inter-city interaction. In models with heterogeneous activity (e.g. Henderson and Becker, 2000), the heterogeneous activities were not allowed to locate separately and still be traded. In the works of Henderson (1974), heterogeneous activities without interindustry externalities were traded at zero cost causing complete specialization from the outset, requiring separate cities of substantial sizes to exist throughout the growth phase, hence requiring developers to set up such cities. Trade as we have used it, and as it can be used in a modified Henderson and Becker model (see previous paragraph), allows the existence of highly asymmetric situations in which viable tiny specialized cities can emerge without developer action. These remain viable because they can trade with the larger and older cities.

Our results do not mean that developers are never needed. We have also shown that there are parameter values under which the second city must be diversified and large from its birth. This can arise when inter-city trading cost is sufficiently high and a lumpy adjustment is needed to set up the new city at the efficient time. If this time is missed—because developers are not allowed to operate or do not exist—then the Malthusian trap arises again, or a specialized city may again emerge but, in this case, inefficiently. Historically, situations where developers are needed to set up cities at the efficient times may be largely anachronisms. In the last century, railroads, the telegraph, the telephone, interstate highways and the Internet have brought down the cost of inter-city transport and communication. Hence, it would seem that modern conditions increasingly favor the self-organization of new settlements under laissez-faire.

Acknowledgements

An earlier version of this paper (“Formation and Growth of Specialized Cities”) was presented at the North American meetings of the Regional Science International Association held at Montreal, Canada, November 10–13, 1999. We thank Marcus Berliant, discussant of that paper, David Pines and two anonymous referees of this journal for their helpful comments.

Appendix A. Derivation of Eqs. (3.3), (3.4), (3.6) and (3.7) for Pattern 3

We show how to derive Eqs. (3.3), (3.4), (3.6) and (3.7) for Pattern 3, in which the specialized city produces services. First, we rewrite Eqs. (2.1)–(2.8), taking into account the asymmetry of the two cities under Pattern 3.

The profit maximization conditions of manufacturers in city a become:

$$w_a = P_x \frac{uX_a}{H_{xa}}, \tag{A.1}$$

$$P_x \frac{(1-u)X_a}{m_a z_{\mathrm{d}a}^{\frac{\sigma-1}{\sigma}} + m_b z_{\mathrm{d}b}^{\frac{\sigma-1}{\sigma}}} z_{\mathrm{d}a}^{-\frac{1}{\sigma}} = q_a, \tag{A.2}$$

$$P_x \frac{(1-u)X_a}{m_a z_{\mathrm{d}a}^{\frac{\sigma-1}{\sigma}} + m_b z_{\mathrm{d}b}^{\frac{\sigma-1}{\sigma}}} z_{\mathrm{d}b}^{-\frac{1}{\sigma}} = \frac{q_b}{\tau}, \tag{A.3}$$

where $z_{\mathrm{d}a}$ is the demand for a locally produced service and $z_{\mathrm{d}b}$ the demand for a service imported from city b and $X = m^{\sigma(1-u)/(\sigma-1)} H_x^u \{[z_{\mathrm{d}a}^{(\sigma-1)/\sigma} + z_{\mathrm{d}b}^{(\sigma-1)/\sigma}]^{\sigma/(\sigma-1)}\}^{1-u}$.

The conditions for a Chamberlin equilibrium in the service sector of each city are:

$$q_i = \frac{\sigma c}{\sigma - 1} w_i; i = a, b, \tag{A.4}$$

$$z_{\mathrm{s}} = \frac{f(\sigma-1)}{c}, \tag{A.5}$$

$$H_z = f\sigma. \tag{A.6}$$

The market-clearing conditions in each service output and in the labor markets are:

$$z_{\mathrm{d}a} = z_{\mathrm{s}} = \frac{f(\sigma-1)}{c}, \tag{A.7}$$

$$z_{\mathrm{d}b} = \tau z_{\mathrm{s}} = \tau \frac{f(\sigma-1)}{c}, \tag{A.8}$$

$$H_a = H_{xa} + m_a H_z, \tag{A.9}$$

$$H_b = m_b H_z \tag{A.10}$$

Recall also the relationship between city population and city labor supply:

$$H_i = N_i\left(1 - kN_i^{\frac{1}{2}}\right); i = a, b. \tag{A.11}$$

These 13 equations are solved for H_{xa}, H_z, z_s and H_i, w_i, q_i, m_i, z_{di}; $i=a,b$ as follows. Substitute into Eq. (A.10) for H_b from Eq. (A.11) and for H_z from Eq. (A.6) to get:

$$m_b = \frac{N_b\left(1 - kN_b^{\frac{1}{2}}\right)}{f\sigma}. \tag{A.12}$$

From Eqs. (A.7) and (A.8), $z_{db} = \tau z_{da}$. Plug this for z_{db} into Eq. (A.2). Also plug into Eq. (A.2) for q_a from Eq. (A.4), for z_{da} from Eq. (A.7), for w_a from Eq. (A.1), for H_{xa} from Eq. (A.9) and then for H_a from Eq. (A.11) and finally for H_z from Eq. (A.6). The result is:

$$m_a + \phi m_b = \frac{1-u}{uf\sigma} N_a\left(1 - kN_a^{\frac{1}{2}}\right) - \frac{1-u}{u} m_a, \mathit{where}\ \phi = \tau^{\frac{\sigma-1}{\sigma}}. \tag{A.13}$$

Now, plug into Eq. (A.13) for m_b from Eq. (A.12) and solve for m_a:

$$m_a = \frac{1-u}{f\sigma}\left[N_a\left(1 - kN_a^{\frac{1}{2}}\right) - \frac{u}{1-u}\phi N_b\left(1 - kN_b^{\frac{1}{2}}\right)\right]. \tag{A.14}$$

Next, forming the ratio of Eqs. (A.2) to (A.3) and substituting into this from Eqs. (A.4), (A.7) and (A.8), we can show that $w_b = \phi w_a$ which appeared in Eq. (3.5). Now to find w_a, use Eq. (A.1) to get:

$$w_a = \frac{uP_x(H_{xa})^u\left[m_a z_{da}^{\frac{\sigma-1}{\sigma}} + m_b z_{db}^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma(1-u)}{\sigma-1}}}{H_{xa}}. \tag{A.15}$$

Now make all the substitutions into the right side of Eq. (A.15) from previously derived expressions, to get:

$$w_a\ \lambda P_x\left[N_a\left(1 - kN_a^{\frac{1}{2}}\right) + \phi N_b\left(1 - kN_b^{\frac{1}{2}}\right)\right]^{\frac{1-u}{\sigma-1}}. \tag{A.16}$$

Substituting w_a and w_b from the above, into the indirect utility function given in Eq. (2.9a), using the facts that $I_a(N_a)=(1 - kN_a^{1/2})w_a$ and $I_b(N_b)=(1 - kN_b^{1/2})w_b$, we get the utility levels of city a and city b given by Eqs. (3.3) and (3.4).

Now, from Eq. (3.5), we know that $(1 - kN_a(t)^{1/2}) = \phi(1 - kN_b(t)^{1/2})$, which is the equilibrium condition guaranteeing that utilities at every time t are equal in the two cities. Plug into this the population partition constraint $N_a(t) = N(t) - N_b(t)$ and manipulate it algebraically to obtain $k^2(1+\phi^2)N_b(t) + 2k\phi(1-\phi)N_b(t)^{1/2} + (1-\phi)^2 - k^2N(t) = 0$. Substitute $Y \equiv N_b(t)^{1/2}$ and $Y^2 \equiv N_b(t)$, to get a quadratic polynomial of Y. Applying the quadratic formula to solve for Y and then for $N_b(t)$, one gets Eq. (3.6). To get Eq. (3.7), recall from Eq. (2.9a) that indirect utility is $U^3(N_a(t), N_b(t)) = P_x^{-\alpha} P_y^{-\beta} (1 - kN_a(t)^{1/2})w_a$. Substitute the expression for w_a from Eq. (A.16) into the right side. Also substitute $N_a(t) = N(t) - N_b(t)$ and substitute for $N_b(t)$ from Eq. (3.6). The result is Eq. (3.7) which gives the equilibrium utility level of Pattern 3 as a function of the total population $N(t)$.

Appendix B. Derivation of Eqs. (3.8), (3.9), (3.11) and (3.12) for Pattern 4

The derivations for Pattern 4, in which the specialized city produces manufactures, are quite similar to those for Pattern 3. We first solve the profit maximization problems of manufacturers in cities a and b and for service producers in city a. Then we impose the market clearing conditions. The profit maximization of manufacturers in city a gives:

$$P_x m_a^{\frac{\sigma(1-u)}{\sigma-1}} u H_{xa}^{u-1} z_{\mathrm{d}a}^{1-u} = w_a \tag{B.1}$$

$$P_x m_a^{\frac{\sigma(1-u)}{\sigma-1}} (1-u) H_{xa}^{u} z_{\mathrm{d}a}^{-u} = m_a q_a. \tag{B.2}$$

Profit maximization of manufactures in city b gives

$$P_x m_a^{\frac{\sigma(1-u)}{\sigma-1}} u H_{xb}^{u-1} z_{\mathrm{d}b}^{1-u} = w_b \tag{B.3}$$

$$P_x m_a^{\frac{\sigma(1-u)}{\sigma-1}} (1-u) H_{xb}^{u} z_{\mathrm{d}b}^{-u} = m_a \frac{q_a}{\tau}. \tag{B.4}$$

Profit maximization of service producers in city a gives

$$\frac{w_a}{q_a} = \frac{\sigma-1}{c\sigma}, \tag{B.5}$$

$$z_{\mathrm{s}} = \frac{f(\sigma-1)}{c}, \tag{B.6}$$

$$H_z = f\sigma. \tag{B.7}$$

The market-clearing conditions for services and for the labor markets in cities a and b are:

$$z_{\mathrm{d}a} + \frac{z_{\mathrm{d}b}}{\tau} = z_{\mathrm{s}} \tag{B.8}$$

$$m_a H_z + H_{xa} = H_a, \tag{B.9}$$

$$H_{xb} = H_b, \tag{B.10}$$

We also have:

$$H_i = N_i\left(1 - kN_i^{\frac{1}{2}}\right); i = a, b. \tag{B.11}$$

These 12 equations can be solved for H_{za}, m_a, q_a, z_s, and H_i, H_{xi}, w_i, z_{di}; $i = a,b$. Following a procedure like that for Pattern 3, the solutions for w_a, w_b and m_a are as follows:

$$w_a = \lambda P_x \left[N_a\left(1 - kN_a^{\frac{1}{2}}\right) + \psi N_b\left(1 - N_b^{\frac{1}{2}}\right)\right]^{\frac{1-u}{\sigma-1}}, \tag{B.12}$$

$$w_b = \psi w_a \tag{B.13}$$

$$m_a = \frac{1-u}{f\sigma}\left[N_a\left(1 - kN_a^{\frac{1}{2}}\right) + \psi N_b\left(1 - N_b^{\frac{1}{2}}\right)\right], \tag{B.14}$$

The steps to derive Eqs. (B.12)–(B.14) above, are as follows. Dividing Eq. (B.1) by Eq. (B.3), we get $w_a/w_b = (H_{xb}^{1-u} z_{da}^{1-u})/(H_{xa}^{1-u} z_{db}^{1-u})$. Then dividing Eq. (B.2) by Eq. (B.4), we get $(H_{xb}^{-u} z_{da}^{-u})/(H_{xa}^{-u} z_{da}^{-u}) = \tau$. Combining these two equations, we get Eq. (B.13), where $\psi \equiv \tau^{(1-u)/u}$. To get Eqs. (B.12) and (B.14), we will first find z_{da}, z_{db} and m_a. First, note from the foregoing that $(H_{xb} z_{da})/(H_{xa} z_{db}) = \tau^{-1/u}$. Substituting into this for H_{xa} and H_{xb} using Eqs. (B.7), (B.7), (B.9), (B.10) and (B.11), we get:

$$z_{db} = z_{da}\tau^{\frac{1}{u}} \frac{N_b\left(1 - kN_b^{\frac{1}{2}}\right)}{N_a\left(1 - kN_b^{\frac{1}{2}}\right) - m_a f\sigma}. \tag{B.15}$$

Now dividing Eq. (B.1) by Eq. (B.2), and using Eq. (B.5) to substitute for w_a/q_a, Eqs. (B.9) and (B.11) to substitute for H_{xa}, and Eq. (B.7) to substitute for H_{za}, and rearranging we get:

$$z_{da} = \frac{1-u}{u}\,\frac{\sigma-1}{c\sigma}\,\frac{N_a\left(1 - kN_a^{\frac{1}{2}}\right) - m_a f\sigma}{m_a}. \tag{B.16}$$

Using Eq. (B.6) in Eq. (B.8), we get:

$$z_{db} = \left(\frac{f(\sigma-1)}{c} - z_{da}\right)\tau \tag{B.17}$$

Setting the two z_{db} from Eqs. (B.15) and (B.17) equal to each other, substituting for z_{da} from Eq. (B.16) into the resulting equation and solving it for m_a we get Eq. (B.14). To get Eq. (B.12), we use Eqs. (B.9), (B.7) and (B.11) to get H_{xa}. Then, we substitute the expression derived for H_{xa}, for z_{da} from Eq. (B.16) and for m_a from Eq. (B.14), into Eq. (B.1).

Now, following the procedure we used in Appendix A, we substitute w_a and w_b from the above, into the indirect utility function given in Eq. (2.9a), using the facts that $I_a(N_a) = (1 - kN_a^{1/2})w_a$ and $I_a(N_b) = (1 - kN_b^{1/2})w_b$, and we get the utility levels of city a and city b for Pattern 4 given by Eqs. (3.8) and (3.9). Finally, Eqs. (3.11) and (3.12) are derived by the following procedure similar to that described in the last paragraph of Appendix A.

From Eq. (3.10), we know that $(1-kN_a(t)^{1/2})=\psi(1-kN_b(t)^{1/2})$, which is the equilibrium condition guaranteeing that utilities at every time t are equal in the two cities. Plug into this the population partition constraint $N_a(t)=N(t)-N_b(t)$ and manipulate it algebraically to obtain $k^2(1+\psi^2)N_b(t)+2k\psi(1+\psi)N_b(t)^{1/2}+(1-\psi)^2-k^2N(t)=0$. Substitute $Y\equiv N_b(t)^{1/2}$ and $Y^2\equiv N_b(t)$, to get a quadratic polynomial of Y. Applying the quadratic formula to solve for Y and then for $N_b(t)$, one gets Eq. (3.11). To get Eq. (3.12), recall from Eq. (2.9a) that indirect utility is $U^4(N_a(t), N_b(t))=P_x^{-\alpha}P_y^{-\beta}(1-kN_a(t)^{1/2})w_a$. Substitute the expression from Eq. (B.12) into the right side. Also substitute $N_a(t)=N(t)-N_b(t)$ and substitute for $N_b(t)$ from Eq. (3.11). The result is Eq. (3.12) which gives the equilibrium utility level of Pattern 4 as a function of the total population $N(t)$.

References

Abdel-Rahman, H.M., Fujita, M., 1993. Specialization and diversification in a system of cities. Journal of Urban Economics 33, 189–222.

Anas, A., 1988. Agglomeration and taste heterogeneity: equilibria, stability, welfare and dynamics. Regional Science and Urban Economics 18, 7–35.

Anas, A., 1992. On the birth and growth of cities: laissez faire and planning compared. Regional Science and Urban Economics 22, 243–258.

Anas, A., Kim, I., 1996. General equilibrium models of polycentric urban land use with endogenous congestion and job agglomeration. Journal of Urban Economics 40, 217–232.

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

Black, D., Henderson, J.V., 1999. A theory of urban growth. Journal of Political Economy 107, 252–284.

Chamberlin, E.H., 1933. The Theory of Monopolistic Competition Harvard Univ. Press, Cambridge, MA.

Dixit, A.K., Stiglitz, J.E., 1977. Monopolistic competition and optimum product diversity. American Economic Review 67 (3), 297–308.

Dobkins, L.H., Ioannides, Y., 1995. Evolution of the USA size distribution of cities. Manuscript, Department of Economics, Tufts University.

Eaton, J., Eckstein, Z., 1997. Cities and growth: theory and evidence from France and Japan. Regional Science and Urban Economics 27, 443–474.

Ethier, W., 1982. National and international returns to scale in the modern theory of international trade. American Economic Review 72, 389–405.

Fujita, M., Krugman, P., Venables, A., 1999. The Spatial Economy: Cities, Regions and International Trade. MIT Press, Cambridge, MA.

Garreau, J., 1991. Edge City: Life on The New Frontier. Doubleday, New York.

Henderson, J.V., 1974. The sizes and types of cities. American Economic Review 64, 640–656.

Henderson, J.V., 1988. Urban Development: Theory, Fact and Illusion Oxford Univ. Press, Oxford.

Henderson, J.V., Becker, R., 2000. Political economy of city sizes and formation. Journal of Urban Economics 48, 453–484.

Henderson, J.V., Mitra, A., 1991. The new urban landscape: developers and edge cities. Journal of Urban Economics 26, 613–643.

Pines, D., 2000. On alternative urban growth patterns. Unpublished manuscript, The Eitan Berglas School of Economics, Tel-Aviv University.

Xiong, K., 1998. Intercity and Intracity Externalities in a System of Cities: Equilibrium, Transient Dynamics and Welfare Analysis. Unpublished PhD Dissertation. State University of New York at Buffalo.