Learning investment and industrial diversity in urban growth*

Chung–Yi Tse[†]

May 31, 2007

Abstract

This paper studies an urban growth model where learning through personal contacts could be more effective in a denser locale, whereas the effectiveness of learning through impersonal means of communications depends principally on the technology of communications rather than on the locale in which learning takes place. As a result of advances in communications technology, cities would be larger, and workers would spent more time on learning through personal contacts but may cut time on impersonal learning if the two kinds of learning investment are complements. Otherwise, cities could become smaller, while workers would spend more time on impersonal learning at the expense of time on learning through personal contacts. In a multi-sector economy, urban industrial diversity tends to increase or fall together with city size.

JEL classifications: O41, R11, R12.

Key words: urban growth, human capital investent, industrial diversity.

1. INTRODUCTION

While internal and external increasing returns to scale in production, forward and backward linkages, and the savings of transportation costs must all have paid important roles for the rise and continuing prominence of the modern city, many writers have argued that it is the role of the city as a hotbed for learning spillovers that constitutes the city's primary function in the service economy of the latter part of the 20th century and the time to come.[1] Indeed since Marshall (1890), economists have emphasized how proximity fosters frequent interactions among workers – an important ingredient for the growth and diffusion of knowledge.[2]

We may learn in solitude by reading books and other printed materials, and in recent times and increasingly, by accessing information available in electronic mediums. How well we learn through such impersonal means of communications probably does not depend on where the learning takes place, but principally on the ease with which the knowledge can be accessed as determined by the technology of communications. We may also learn through personal contacts by observing how others perform certain tasks, by taking instructions from more experienced colleagues, and by discussions with others sharing common interests. This paper studies an urban growth model where the engine of growth is human capital investment through time spent on learning as in Lucas (1988). The point of departure is that I distinguish between learning through personal contacts and learning through impersonal means

---

[1]See for instance Chinitz (1999), Swann (1999), Leamer and Storper (2001), and Glaeser and Kohlhase (2004).

[2]Noteble among recent empirical work that study the link between agglomeration and learning include Glaeser and Maré (2001) and Audretsch and Feldman (1996). Jaffe, Trajtenberg and Henderson (1992), using patent citation data, find that spillovers appear to be geographically bounded. Thompson (2006) finds that the results of Jaffe et al. survive amidst a more rigorous identification strategy.

of communications. The basic hypothesis is that learning through personal contacts could be more effective in a denser urban setting because it allows for more frequent and varied interactions among the inhabitants, whereas the effectiveness of learning through impersonal means of communications should not differ across locations but is determined, by and large, by the technology of communications.

The defining difference between this paper and previous papers on urban growth, that include for instance Eaton and Eckstein (1997), Black and Henderson (1999), and Rossi-Hansburg and Wright (2007),[3] is that agglomeration in the present model is solely for facilitating learning investment, whereas it has been conventional in dynamic urban models to assume that the agglomeration economies are production economies that raise the contemporaneous productivity of firms. No doubt, previous authors recognize that the major channel through which agglomeration raises productivity is that it facilitates various kinds of learning among the city's inhabitants from one another.[4] The simple assumption that agglomeration helps raise static productivity is, by all means, merely a reduced form specification of a more elaborated setting in which the higher productivity results from various kinds of learning spillovers. The virtue of proceeding with assuming static production externalities, instead of starting from first principle, is that it results in a highly tractable setting to facilitate the analysis of a multitude of questions pertaining to urban structure, growth, and efficiency.

Nevertheless, an explicit analysis of how agglomeration helps workers learn better in an urban growth model could in itself yield valuable insights. The modeling of the distinction between personal and impersonal learning investments in particular

---

[3]The survey in Berliant and Wang (2004) of the literature provides an interesting perspective in relating the literature to the tradition of neoclassical growth models.

[4]Durnaton and Puga (2004) survey models of the microfoundation of agglomeration economies. Fujita and Thisse (2002, chapter 6) study an explicit model of the interactions among workers in a city.

provides an ideal setting to analyze the tropical question of the future of cities amidst the vast improvements in information and communications technology (ICT) in the recent past and the expected continuing advances in the foreseeable future. The answer depends, rather intuitively, on the elasticity of substitution between the two kinds of learning investment. In case the elasticity is below unity, so that the two kinds of investment are complements, advances in communications technology that help the worker make better use of her time on impersonal learning tend to raise the payoff of time spent on learning through personal contacts. As a result, there would be greater agglomeration in equilibrium. In case the elasticity is above unity, so that the two kinds of investment are substitutes, workers would first for sure find it optimal to raise impersonal learning investment. The increase could possibly be partially at the expense of time on personal learning investment if the latter may be easily substituted away. In equilibrium, cities could become smaller in size and more numerous.

In an extension to the basic model, I turn to study how urban industrial diversity and city size are jointly determined in a multi–sector economy. In this environment, there can be one, two, or just any number of sectors in a city. Adding a sector to a city has the benefit of enriching the city's workers' learning experiences by allowing the city's workers to learn from workers in that additional sector through personal contacts. The downside is that with the city's population spread among one more sector, each sector included in the city would be smaller in size. Then the city's workers could only learn with diminished effectiveness from workers in each such sector because of the diminished concentration of workers of the sector in the city. In general, I find that, following advances in communications technology, if the city becomes larger, it would also be more diversified, and vice versa.

This paper is closely related to Gasper and Glaeser (1998) in its focus on the effects of the advances in ICT on the future of city. While Gasper and Glaeser

4

restrict each agent in their model to communicate either through the "telephone" or through personal contacts, I give workers in the present model the choice of investing in both personal and impersonal learning investments. Further, while in Gasper and Glaeser, whether the city would increase in size following advances in communications technology seems to depend on the distribution of ideas among agents, my answer to the same question hinges on the nature of the learning technology.

Urban industrial diversity has been an important topic of research. Duranton and Puga (2000) and Abdel-Rahman and Anas (2004) survey this large literature. In addition, Duranton and Puga (2001) argue that young and innovative firms could find the large and diversified city, because of the availability of a wide variety of intermediate inputs, an attractive place to experiment. Pascal and McCall (1980) explores the role of path dependence in explaining why young and presumably innovative firms would choose to follow the examples of earlier and more successful entrants to locate in the large and diversified city. Berliant, Reed and Wang (2006) argue that interactions among workers would be most useful when the knowledge possessed by a pair of workers is not too alike or too different.

The rest of the paper is organized as follows. The next section studies the basic one–sector model. In section 3, I extend the analysis to study the joint determination of equilibrium urban industrial diversity and city size in the multi–sector model. Section 4 contains discussions on the role played by industrial policies as coordinating devices and how equilibrium differs from the social optimum. Section 5 concludes. Proofs that only involve routine calculations are omitted for brevity; the less straightforward proofs can be found in the appendix.

2. ANALYSIS

2.1 Preference

The economy is populated by a continuum of household–workers, each of whom has the same preference, given by

$$U = \sum_{t=0}^{\infty} \delta^t \ln c_t,$$

where $\delta < 1$ is the subjective discount factor and c_t the household's consumption in period t. Utility maximization is subject to the usual intertemporal budget constraint that equates the household's wealth to the present value of the household's consumption over time. Let r_t be the interest rate. The household's consumption plan then satisfies the familiar optimal consumption growth condition

$$\frac{c_{t+1}}{c_t} = \delta \left[1 + r_{t+1} \right] \tag{1}$$

for logarithmic preferences.

2.2 The spatial structure of the city and commuting

I assume the conventional two–dimensional circular city. Housing density is assumed fixed at 1, and the cost of housing development is normalized to 0. All workers must commute to the city center for employment. In practice, both privately–supplied inputs that include the commuter's time, the auto, the gasoline, etc. and publicly–supplied inputs that include the mass transit system, highways, traffic lights, and other related public capitals are essential inputs to urban transportation. In assuming that the worker spends a certain amount of time and/or uses up a certain number of units of her own output to commute one unit distance, the typical urban model emphasizes the roles played by privately–supplied inputs, while completely abstracts from the roles played by publicly–supplied inputs. In this paper, I choose to focus

on publicly–supplied inputs instead. Specifically, I assume that a worker who wishes to commute a distance of d must purchase d^{α} units of "urban traveling" supplied by the public sector, for some parameter $\alpha > 0$. Perhaps this setup best describes where commuting is by public transportation. But it could apply to commuting by auto as well, where the traveling uses some tolled highways, bridges, and tunnels.

In any case, if the transit authority charges some amount z for each unit of traveling service rendered, and if the border of the city extends to a distance of b from the city center, the equilibrium housing rent function becomes $h(d) = z(b^{\alpha} - d^{\alpha})$, whereas the worker's location cost, defined as the sum of housing rent and commuting expense, is equal to zb^{α}. On the supply side, assume that to provide one unit of urban traveling requires ω units of labor. In turn, to successfully hire a unit of labor, the transit authority has to pay for the value of the worker's outside option – the production of the consumption good. Thus if a worker is able to produce A units of output per unit of time spent on good production, she would be just willing to work for the transit authority for a wage rate equal to A, and that is how much the transit authority would pay the worker.

The two unusual elements in this setup :

1. monopoly supply of urban transportation by some transit authority,

2. non–linear commuting expense,

both seem to make the analysis more complicated than in their absence. But they do play important roles in guaranteeing the existence of a symmetric constant growth equilibrium under relatively simple conditions, in which the economy would not degenerate into a single–city economy, and in which the population of each city is stationary over time.

2.3 Production and learning

A worker has a time–endowment of 1 in each period of time and allocates it among learning, the production of the consumption good, and employment in the urban transportation sector. Let A_t denote the worker's productivity; i.e., output per unit of time spent on good production. Since the transit authority would pay a wage rate just equal to the worker's productivity in good production, no matter how she divides time between production and employment in the transportation sector, the worker's income is equal to

$$y_t = A_t \left(1 - x_t\right)$$

if x_t is the time spent on learning.

Suppose the worker stays in some city i. Given that the worker spends x_t units of time on learning, the worker's productivity (human capital) is assumed to evolve according to

$$A_{t+1} = A_t + k_{it} x_t, \tag{2}$$

where k_{it} denotes the productivity of time spent on learning in city i. Let τ_{it} be the lump–sum subsidy the worker receives from staying in city i and $\beta = \frac{1}{1+r}$ the discount factor. The worker chooses x_t to maximize the present value of the stream of net income:

$$V_{it}\left(A_t\right) = \max_{x_t} \left\{ A_t \left(1 - x_t\right) - z_{it} b_{it}^\alpha + \tau_{it} + \beta V_{it+1}\left(A_{t+1}\right) \right\}, \tag{3}$$

subject to (2), and taking as given the time paths for k_{it}, z_{it}, b_{it}, and τ_{it}. The first order condition for an interior optimum reads

$$-\frac{A_t}{k_{it}} + \beta \left(1 - x_{t+1} + \frac{A_{t+1}}{k_{it+1}}\right) = 0. \tag{4}$$

2.4 Agglomeration economies – learning spillovers

Consider first how much the worker can expect to gain from time spent on learning through impersonal means of communications. How much there is to learn and how well a worker can learn should only depend on the quality of the learning materials and the ease with which the materials can be accessed. If the economy's knowledge capital stock may be summarized by its average human capital – $\overline{A}_t$, the productivity of impersonal learning – by how much the worker's human capital is augmented per unit of time spent on the activity, may be most simply modeled by $\overline{A}_t\eta$, for some η that measures the efficacy of the technology of communications.

Similarly, how much there is to learn through personal contacts should only depend on what those the worker would come into contact with know – the quality of contacts (Jovanovic and Rob (1989), Berliant et al. (2006)). Moreover, the productivity of time spent on the activity would also depend on the number of contacts the worker may make within a certain time period, and the worker could make many more such contacts in a larger city (Glaeser (1999)). If the quality of contacts can be summarized by the city's average human capital – $\overline{A}_{it}$, it is simplest to assume that the productivity of time spent on personal learning is given by $\overline{A}_{it}f(n_{it})$, where n_{it} is city i's population and $f(n)$ some continuously differentiable function that satisfies $f'(n) > 0$ for $n \geq 0$. This formulation assumes that the city's average human capital is a sufficient statistic for the entire distribution of human capital in the city, and that learning productivity is multiplicatively separable in the average quality of contacts and a function that gives the frequency of contacts. Granted that such assumptions are not unduly implausible, the formulation could be a useful first step to explore the aggregative implications of how the two kinds of learning investment may interact on the determination of equilibrium city size.

In all, by how much the worker's human capital increases in the given period is

assumed to be given by

$$\Delta_{it} = \left(\left[\overline{A}_{it} f\left(n_{it}\right) x_t^P \right]^{\frac{\theta-1}{\theta}} + \left[\overline{A}_t \eta x_t^I \right]^{\frac{\theta-1}{\theta}} \right)^{\frac{\theta}{\theta-1}}, \tag{5}$$

where $\theta > 0$ is the elasticity of substitution between the two kinds of learning invest-
ment, and x_t^P and x_t^I denote, respectively, the time spent on learning through personal
interactions and impersonal means, which would add up to x_t- the aggregate time
investment on learning :

$$x_t = x_t^P + x_t^I. \tag{6}$$

Maximizing (5) subject to (6) yields

$$x_t^P = x_t \frac{\left(\overline{A}_{it} f\left(n_{it}\right) \right)^{\theta-1}}{\left(\overline{A}_{it} f\left(n_{it}\right) \right)^{\theta-1} + \left(\overline{A}_t \eta \right)^{\theta-1}}, \tag{7}$$

$$x_t^I = x_t \frac{\left(\overline{A}_t \eta \right)^{\theta-1}}{\left(\overline{A}_{it} f\left(n_{it}\right) \right)^{\theta-1} + \left(\overline{A}_t \eta \right)^{\theta-1}}, \tag{8}$$

if the second order condition that $\theta < 2$ is met, and where

$$\frac{\partial x_t^P}{\partial \eta} \underset{<}{\overset{>}{\gtrless}} 0 \Leftrightarrow \theta \underset{>}{\overset{<}{\lessgtr}} 1. \tag{9}$$

That is, the two learning technologies are complements in case $\theta < 1$ but substitutes
otherwise.

   In turn, substituting (7) and (8) into (5) yields

$$\Delta_{it} = \left(\left[\overline{A}_{it} f\left(n_{it}\right) \right]^{\frac{\varepsilon-1}{\varepsilon}} + \left[\overline{A}_t \eta \right]^{\frac{\varepsilon-1}{\varepsilon}} \right)^{\frac{\varepsilon}{\varepsilon-1}} x_t \equiv k_{it} x_t, \tag{10}$$

where

$$\varepsilon = \frac{1}{2-\theta}. \tag{11}$$

In (10), k_{it} may then be taken as the overall productivity of learning investment in

10

city i. It is convenient to normalize $f(0) = 1$ and write[5]

$$\overline{A}_t \left(1 + \eta^{\frac{\varepsilon-1}{\varepsilon}} \right)^{\frac{\varepsilon}{\varepsilon-1}} \equiv \overline{A}_t \kappa \tag{12}$$

as the productivity of learning through impersonal means of communications alone. In this case, k_{it} starts out equal to $\overline{A}_t \kappa > 0$ at $n_{it} = 0$, and is increasing in n_{it} thereafter if $\overline{A}_{it} \geq \overline{A}_t$. Finally if $\overline{A}_{it} = \overline{A}_t$, (10) specializes to

$$k_{it} = \overline{A}_t \left(f(n_{it})^{\frac{\varepsilon-1}{\varepsilon}} + \eta^{\frac{\varepsilon-1}{\varepsilon}} \right)^{\frac{\varepsilon}{\varepsilon-1}} \equiv \overline{A}_t k(n_{it}). \tag{13}$$

Because I would just be working with the reduced form learning technology in (13) hence, I should from now on state conditions on ε rather than on θ. This would not cause problems in the interpretations of the conditions at all since by (11), the condition that $\varepsilon > (<)$ 1 is strictly equivalent to the condition that $\theta > (<)$ 1. In this regard, assume that

$$k''(n) < 0 \Leftrightarrow -k(n)^{\frac{1-\varepsilon}{\varepsilon}} \eta^{\frac{\varepsilon-1}{\varepsilon}} \frac{f'(n)^2}{\varepsilon f(n)} + f''(n) < 0, \tag{14}$$

for $n \geq 0$. This condition, which helps guarantee the second order conditions for maximization are met in the analysis to follow, is weaker than assuming strict concavity on $f(n)$; it would hold even if $f''(n) \geq 0$, as long as ε is not too large.

2.5 The market for city

There is an unlimited number of homogenous sites in the economy where cities may be developed. Rural land is in excess supply and therefore a free good. Following

---

[5]Alternatively, one might have considered normalizing $f(0) = 0$. In this case and if $\varepsilon \leq 1$, $k_{it} = 0$ at $n_{it} = 0$. Underlying this normalization is then the presumption that, in case $\varepsilon \leq 1$, personal (as well as impersonal) learning is an essential input to human capital investment. It would be explained in note 10 that how the existence of equilibrium cannot be guaranteed with relatively simple conditions, and how equilibrium, given existence, can never be unique under this seemingly more natural normalization.

a long tradition in urban economics, cities are assumed to form in an economywide land market by profit–maximizing property developers (or alternatively autonomous city governments), where each city is run by a single developer, who collects rents and specifies urban population.[6] The city developer in the present model also hires labor to supply urban transportation service to the city's households.

Suppose all workers in the economy start out at the same level of productivity, equal to A_t say. Consider some city i. If the city's population in period t is n_{it}, its border would extend to $b_{it} = \sqrt{n_{it}/\pi}$. Since the sum of rent and commuting charge each household pays the city developer is equal to $z_{it}b_{it}^{\alpha}$, the developer's gross revenue is $n_{it}z_{it}b_{it}^{\alpha} = z_{it}n_{it}^{1+\alpha/2}/\pi^{\alpha/2}$. To earn this revenue, the developer supplies

$$\int_0^{b_{it}} 2\pi s^{1+\alpha} ds = \frac{2}{2+\alpha}\frac{n_{it}^{1+\alpha/2}}{\pi^{\alpha/2}}$$

units of urban transportation at a unit cost of ωA_t. The developer may also choose to subsidize each household in the city for an amount equal to τ_{it}. In sum the developer's net revenue is

$$\Pi_{it} = \max_{\{n_{it},z_{it},\tau_{it}\}}\left\{\left(z_{it} - \frac{2\omega A_t}{2+\alpha}\right)\frac{n_{it}^{1+\alpha/2}}{\pi^{\alpha/2}} - n_{it}\tau_{it}\right\}, \tag{15}$$

where the maximization is subject to the constraint that the households are at least as well off as staying elsewhere in the economy. To specify this constraint, let $V_t(A_t)$ be the present value of income a worker can earn by staying in the representative city. If in period t, the worker stays in city i, and then moves to and stays in the representative city thereafter, the present value of income is

$$\widetilde{V}_{it}(A_t) = \max_{x_t}\left\{A_t(1-x_t) - z_{it}\left(\frac{n_{it}}{\pi}\right)^{\alpha/2} + \tau_{it} + \beta V_{t+1}(A_{t+1})\right\}.$$

The maximization in (15) is then subject to

$$\widetilde{V}_{it}(A_t) \geq V_t(A_t).$$

---

[6]The tradition can be traced to Henderson (1974). See the survey of the literature in Abdel-Rahman and Anas (2004).

2.6 Symmetric constant growth equilibrium

Combining the first order conditions of (15) with respect to n_{it} and τ_{it} and the zero–profit condition for city formation yields[7]

$$xk\,(n)^{1/\varepsilon-1}\,f\,(n)^{-1/\varepsilon}\,f'\,(n) = \frac{\alpha\omega}{2+\alpha}\frac{n^{\alpha/2-1}}{\pi^{\alpha/2}}, \tag{16}$$

where, without loss of generality, I have dropped the city and time subscripts. This is the equilibrium condition for n in the symmetric constant growth equilibrium – the equilibrium that emerges when all workers in the economy happened to start out with the same level of productivity in some prior period. With production exhibiting constant returns and learning displaying constant growth, there can be no transition dynamics in the symmetric equilibrium.

The condition (16) admits a very intuitive interpretation in that it is just the first order condition of

$$\max_n \left\{ A_t\,(1-x) - \frac{2\omega A_t}{2+\alpha}\frac{n^{\alpha/2}}{\pi^{\alpha/2}} \right\} \tag{17}$$

subject to

$$A_{t+1} = A_t + A_t k\,(n)\,x, \tag{18}$$

for some given A_{t+1}, where the maximand in (17) is output per capita net of the resources used up in commuting. In this connection, the right side of (16) denotes the cost of increasing agglomeration; i.e., by how much commuting expense per capita increases per unit increase in n, normalized by A_t. The left side measures the returns to increasing agglomeration, again normalized by A_t; i.e., by how much current period per capita output increases per unit increase in n, holding constant outputs in all future periods, for which a necessary condition is that A_{t+1} in (18) stays at the given level.

---

[7]z_{it} and τ_{it} cannot be determined separately. But, given n_{it}, together they satisfy $\tau_{it} = \left(z_{it} - \frac{2\omega A_t}{2+\alpha}\right)\frac{n_{it}^{\alpha/2}}{\pi^{\alpha/2}}$ by the zero–profit condition.

The first order condition (16) is also sufficient if (17) (with x substituted out via (18)) is strictly concave. Otherwise, the agglomerative force could dominate the dispersion force at all levels of n to result in the entire economy collapsing into a single city. By (14), the left side of (16) is diminishing in n. If $\alpha > 2$, the marginal cost of agglomeration, on the right side of (16), is strictly increasing in n, and the second order condition is guaranteed to hold. To simplify analysis, I shall henceforth restrict attention to where $\alpha > 2$.[8] Then it is straightforward to verify that there exists a unique solution to (16) for any given x.[9] Further by straightforward differentiation:

Lemma 1 *An increase in η raises the marginal returns to agglomeration (the left side of* (16)*) if and only if $\varepsilon < 1$.*

To proceed, rewrite (16) as

$$x^D(n) = \frac{\alpha \omega n^{\alpha/2-1} k(n)^{1-1/\varepsilon} f(n)^{1/\varepsilon}}{(2+\alpha) \pi^{\alpha/2} f'(n)} \tag{19}$$

– a monotone increasing function that starts out equal to 0 and becomes unbounded as n becomes large. This is the inverse of a function that gives the developer's optimal choice of n as a function of x. Since by (16), the marginal returns to agglomeration are increasing in x, $x^D(n)$ should only be a strictly increasing function.

To complete the characterization of equilibrium, I return to (4), which characterizes the time path for x_t in the worker's optimum. With $k_{it} = A_t k(n)$, the equation simplifies to

$$(1-x) k(n) = r, \tag{20}$$

---

[8]Where there is linear commuting; i.e., $\alpha = 1$, the marginal cost is diminishing in n. Then the second order condition is met only if the marginal returns diminish at a faster rate. This requires either a small ε or a rapidly diminishing $f'(n)$. For example, if $f(n) = an^\phi$, a sufficient condition is that $\phi < 1/2$.

[9]The left side is positive at $n = 0$ given that $f'(0) > 0$, whereas the left side is equal to 0. Thereafter the right side increases without bound, while the left side declines.

14

where $r = \frac{1-\beta}{\beta}$. From (1), $c_{t+1}/c_t = \delta(1+r)$, and in the constant growth equilibrium, $c_{t+1}/c_t = A_{t+1}/A_t = 1 + k(n)x$. Hence

$$r = \frac{1 + k(n)x - \delta}{\delta}. \tag{21}$$

Combine (20) and (21) to obtain

$$x^W(n) = \frac{1}{1+\delta}\left[\delta - \frac{1-\delta}{k(n)}\right]. \tag{22}$$

Lemma 2 *(a) $x^W(n)$ is a strictly concave and increasing function, bounded by*

$$x^W(n) < \frac{\delta}{1+\delta}, \tag{23}$$

and positive over $n \geq 0$ if

$$\delta \geq \frac{1}{\kappa + 1}. \tag{24}$$

(b) $\partial x^W/\partial \eta > 0$.

The function $x^W(n)$ denotes the worker's optimal investment in equilibrium. Where an increase in either n or η raises learning productivity, the worker would choose a larger x to follow.

The constant growth symmetric equilibrium with positive learning investment and non–zero city size is a $n > 0$ that solves $x^D(n) = x^W(n)$. By construction, any such equilibrium is a free–mobility equilibrium since workers earn the same discounted lifetime income in any cities in the economy.

Proposition 1 *Equilibrium with positive learning investment and non–zero city size exists if (24) holds. In addition, if*

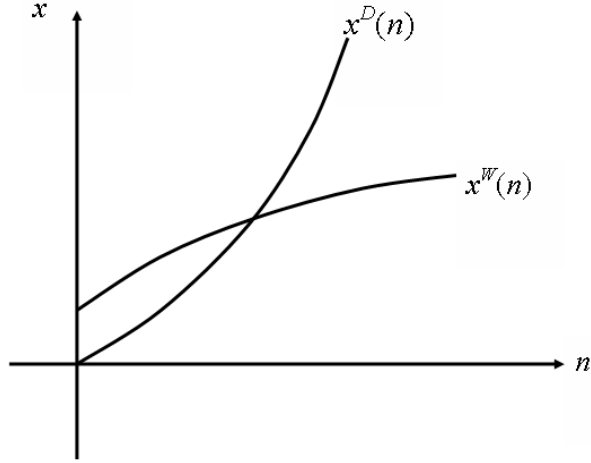$$\delta \geq \frac{2}{\kappa + 2}, \tag{25}$$

the equilibrium is unique.

15

FIG. 1. Existence and uniqueness of equilibrium

Figure 1 plots the two functions: $x^D(n)$ and $x^W(n)$. Since $x^D(0) = 0$ and is increasing without bound thereafter, while $x^W(0) > 0$ if (24) holds, and is likewise increasing but strictly concave and bounded by (23), the two functions in figure 1 must intersect at least once. Further, the appendix shows that if (25) holds, the two functions may intersect just once.[10]

2.7 City size and improvements in communications technology

Given the existence and uniqueness of equilibrium, I may now move on to study how n and x behave in equilibrium with respect to advances in the technology of impersonal learning.

---

[10]Had we chosen to normalize $f(0) = 0$, and in case $\varepsilon \leq 1$, $\kappa \equiv k(0) = 0$. In this case, (24) and (25) can never be met since $\delta < 1$. Indeed, $x^W(n)$ only becomes positive for sufficiently large n if $k(n)$ starts out equal to 0. Then the analysis in figure 1 suggests that, given existence, there must be at least two equilibria, with x^W strictly concave and bounded but x^D unbounded.

Proposition 2 *An increase in η (a) raises equilibrium city size n if $\varepsilon \leq 1$ and (b) always results in greater learning investment x.*

An increase in η exerts two kinds of influence on equilibrium city size. First, insofar as the increase in η is an increase in learning productivity, by lemma 2, the worker raises investment, which in turn raises the payoff to increasing agglomeration for the developer. Equilibrium city size tends to increase as a result. Second, for a given level of investment, by lemma 1, the marginal returns to agglomeration increase if the two kinds of learning are complements. Thus in case $\varepsilon \leq 1$, improvements in the technology of communications unambiguously results in increasing agglomeration. But in case $\varepsilon > 1$, the marginal returns to agglomeration decline at each level of investment. If this effect dominates, there could well be declining agglomeration in equilibrium. Even though the condition that $\varepsilon \leq 1$ is a sufficient condition for the conclusion of the proposition but not necessary, in various numerical analyzes, I find that when the condition fails to hold, the conclusion is indeed reversed if η is sufficiently large to begin with.

In case $\varepsilon \leq 1$, the two effects of an increase in η both serve to induce the worker to increase investment. First the increase in η raises learning productivity directly. The equilibrium increase in n that follows raises learning productivity even further. Then the worker must find it optimal to increase investment. In case $\varepsilon > 1$, the indirect effect could be negative though since the possible decline in n tends to lower learning productivity. A more careful analysis in the appendix establishes that the direct positive effect should always dominate. The reasoning is straightforward. By (22), x could fall if and only if learning productivity $k(n)$ has fallen. The assumed increase in η represents an exogenous improvement in the learning technology, which may lead to an equilibrium decline in n in case $\varepsilon > 1$. The effect of such an equilibrium response should not more than offset the effect of the initial exogenous change that has caused the decline in the first place.

17

Behind the simple conclusion that improvements in the technology of communications always result in greater overall learning investment are important differences in how the increase in investment comes about between the two regimes of whether ε is below or above 1.

Proposition 3 *(a) In case $\varepsilon \leq 1$, personal learning investment x^P is increasing in η if*

$$\frac{f'(n)^2}{f(n)} - f''(n) \geq 0. \tag{26}$$

(b) In case $\varepsilon \geq 1$, impersonal learning investment x^I is increasing in η. Further in case $\varepsilon > 1$, personal learning investment x^P is decreasing in η if in equilibrium $dn/d\eta < 0$.

When the two kinds of learning investment are not substitutes ($\varepsilon \leq 1$), there would be a larger city to follow the increase in η, in which case the worker should only find it optimal to raise x^P.[11] When the two kinds of investment are relatively substitutable ($\varepsilon \geq 1$), after the same exogenous change, perhaps at the expense of x^P, the worker would first for sure find it optimal to increase x^I. If there is also a smaller city in equilibrium, the increase in x^I is indeed partially achieved by reducing time on x^P.

One case not covered by the proposition is how x^I should behave in case $\varepsilon < 1$. There should be two effects. The improvement in the technology of impersonal learning tends to induce the worker to increase x^I. In the mean time, given the two kinds of investment are complements, the worker also wishes to raise x^P. This increase in x^P could possibly be at the expense of x^I. Figure 2 illustrates via a numerical example how the direct positive effect dominates at small η but the indirect negative effect dominates at larger η.[12]

---

[11]The condition for this result to hold (26) is sufficient but not necessary, and in any case is a rather weak restriction. Because $k(n)^{1/\varepsilon-1} f(n)^{1/\varepsilon-1} < 1$ but $1/\varepsilon \geq 1$ in case $\varepsilon \leq 1$, the condition is neither stronger nor weaker than (14).

[12]This example assumes $f(s) = 1 + s^\phi$, $\phi = 0.6$, $\omega = 0.1$, $\delta = 0.98$, $\alpha = 2.2$, and $\varepsilon = 0.8$.
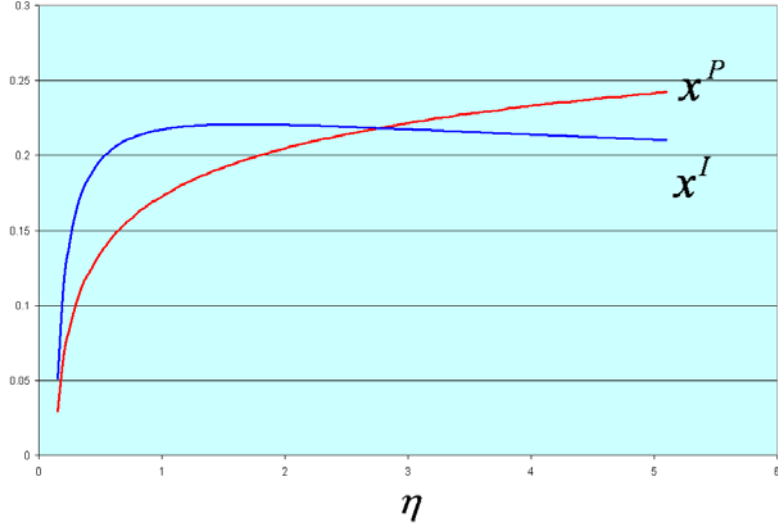
FIG. 2. Personal vs Impersonal Learning Investments, $\varepsilon < 1$.

2.8 Long–run growth

Since both x and $k\left(n\right)$ are increasing in η, the growth rate of output :

$$\frac{A_{t+1}}{A_t} = 1 + k\left(n\right)x,$$

should only rise in tandem. How that additional growth is achieved, however, differs between the two regimes of whether ε is below or above 1. In the first case, there is greater agglomeration of production, and workers tend to spend more time on personal and possibly less on impersonal learning. In the second case, the city could shrink in size, whereby workers would raise impersonal at the expense of personal learning investments.

2.9 Alternative assumptions on commuting

In closing, it would be instructive to examine how the analysis would be affected under the more usual assumption in urban models that the inputs to commuting are

19

supplied by the same agent who demands the commuting.

Consider first the most often adopted assumption that it costs an agent a certain number of units of her own output to travel one unit distance. In this case, the congestion cost of agglomeration would just stay constant over time. In a dynamic urban model, where the returns to agglomeration grow at the rate of productivity increase, the city would grow in size in parallel to the productivity growth. In this environment, the analysis of how equilibrium city size behaves with respect to changes in the technology of communications would be considerably more involved than the same analysis in which equilibrium city size stays constant over time.

An apparently benign alternative is to assume that commuting cost is in terms of time lost to production, in which case the congestion cost of agglomeration could well grow with the returns to agglomeration. The problem with this assumption is that, unless the monopoly developer assumption is literally true, and that the city planner, as a landlord, can observe the productivity of each household–worker and commit to take–it–or–leave–it offers, a symmetric equilibrium under which each worker's human capital grows at the same rate cannot be sustained. The arguments run like the following: Suppose all but one worker have chosen to attain the same productivity level, equal to say $\overline{A}_{t+1}$, in the upcoming period. Assuming, without loss of generality, that it takes 1 unit of time to commute 1 unit distance, the equilibrium rent function in the period would be given by

$$h_{t+1}\left(d\right) = \overline{A}_{t+1}\left(b_{t+1} - d\right). \tag{27}$$

If the worker in question shall a attain a productivity level $A_{t+1} < \overline{A}_{t+1}$, it would be optimal for her to reside just at the city's border, paying zero rent and incurring $b_{t+1}A_{t+1}$ as the opportunity cost of commuting. If the worker instead chooses to attain $A_{t+1} = \overline{A}_{t+1}$, she would be indifferent among any locations $d \leq b_{t+1}$ as she pays the same location cost equal to $b_{t+1}A_{t+1}$ anywhere in the city. For $A_{t+1} \leq$

$\overline{A}_{t+1}$ then, the worker's location cost is given by $b_{t+1}A_{t+1}$. But should the worker decide to attain some $A_{t+1} > \overline{A}_{t+1}$, with the rent function given by (27), the worker would find it optimal to reside just at $d = 0$, incurring a location cost equal to $b_{t+1}\overline{A}_{t+1} < b_{t+1}A_{t+1}$. Thus, there would be a kink in the worker's payoff function at $A_{t+1} = \overline{A}_{t+1}$, giving rise to a discontinuity in the first order condition, from which it follows that the worker's optimal A_{t+1} can never occur at $\overline{A}_{t+1}$.[13] A symmetric equilibrium fails to exist as a result. Further, in the social optimum, workers locating differently should indeed invest differently: The time that remains after commuting for workers cannot be identical when they locate differently. Then the returns on human capital investment would also differ among workers. While equilibrium and optimum asymmetries are interesting in their own rights, they are beyond the scope of this paper. And again the analysis of how equilibrium city size behaves with respect to changes in communications technology would be considerably more involved in such an environment.

---

[13]If the monopoly developer assumption is literally true, and that the developer can observe each worker's productivity in the *housing market* and commit to take–it–or–leave–it offers, it would charge a worker whose productivity is equal to A_{t+1} a rent of $A_{t+1}(b_{t+1} - d)$ for location d. The worker's location cost is then equal to $A_{t+1}b_{t+1}$ at any levels of A_{t+1}. But such assumptions are probably not plausible. True, an essential element in the publicly–supplied commuting setup I adopt in this paper is that each worker's productivity is observable to the transit authority in the *labor market*. Otherwise, a worker's wage rate in the urban transportation sector needs not be identical to the worker's productivity in the good sector. Still it seems much more agreeable to assume that a worker's productivity is observable in the labor market than to assume that it is observable in the housing market.

3. CROSS–SECTOR SPILLOVERS AND INDUSTRIAL DIVERSITY

3.1 Learning in a multi–sector economy

In this section, I extend the analysis to study how urban industrial diversity is determined together with city size in a multi–sector economy. To begin, assume that now the consumption good can be produced via a continuum of technology, indexed by $j \in [0,1]$, where each technology is identified as a distinct sector. Each worker's human capital is assumed to be specific to one such technology only.

In this economy, each city may host one, two, or just any number of sectors. In the absence of cross–sector spillovers though, in equilibrium, there would not be multi–sector cities. Adding a sector to a city raises the city's population and thereby the congestion cost of agglomeration. With no benefits to counteract the increased congestion cost, the city is unambiguously a less desirable place to learn and to produce.[14] In the presence of cross–sector spillovers in learning, however, whereby a sector j worker may learn from workers in other sectors, in addition to learning from her peers in the same sector, there could well be room for a multi–sector city equilibrium.

Specifically, let k_{ijt} represent the sector j knowledge capital that city i workers can access in human capital investment. It is simplest to consider a symmetric configuration in which the productivity of learning investment in city i is given by a CES aggregate of the sectoral and city specific knowledge capital stocks :

$$\overline{k}_{it} = \left(\int_0^1 k_{ijt}^{\frac{\sigma-1}{\sigma}} \, dj \right)^{\frac{\sigma}{\sigma-1}}, \tag{28}$$

where $\sigma > 0$ is the elasticity of substitution between the knowledge capitals drawn from any two sectors. In parallel with the setup in the last section, assume that

$$k_{ijt} = \left(\left[\overline{A}_{ijt} f\left(s_{ijt}\right) \right]^{\frac{\varepsilon-1}{\varepsilon}} + \left[\overline{A}_{jt} \eta \right]^{\frac{\varepsilon-1}{\varepsilon}} \right)^{\frac{\varepsilon}{\varepsilon-1}}, \tag{29}$$

---

[14]See Duranton and Puga (2004) and Abdel-Rahman and Anas (2004) for more discussion.

where $\overline{A}_{ijt}$ and $\overline{A}_{jt}$ represent, respectively, the average human capital of sector j workers in city i and the economywide average human capital of sector j workers, and s_{ijt} the mass of sector j workers residing in city i. Then the city's population is

$$n_{it} = \int_0^1 s_{ijt} dj.$$

Implicit in (28) and (29) is the same dichotomy between learning through impersonal means of communications and through personal interactions. That is, if a sector, say j, is not present in city i, the city's workers may only learn from sector j workers elsewhere in the economy through impersonal means of communications. On the other hand, if there are sufficiently many sector j workers residing in city i, the learning investment may also be facilitated by the physical interactions with these workers.

In case $\overline{A}_{ijt} = \overline{A}_{jt} = \overline{A}_t$, and further if each sector present in the city is of the same size, given by $s_{ijt} = s_{it}$, (28) specializes to

$$\overline{k}_{it} = \overline{A}_t \left(\frac{n_{it}}{s_{it}} k\left(s_{it}\right)^{\frac{\sigma-1}{\sigma}} + \left(1 - \frac{n_{it}}{s_{it}}\right) \kappa^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}} \equiv \overline{A}_t \overline{k}\left(n_{it}, s_{it}\right), \tag{30}$$

where $n_{it}/s_{it} \leq 1$ is the range of sectors the city hosts. The equation of motion in (2) is henceforth replaced by

$$A_{t+1} = A_t + \overline{k}_{it} x_t. \tag{31}$$

Previously, I assumed that the learning related agglomeration economies would begin to take effect with the very first worker in the city; i.e., $f'(n) > 0$ for $n \geq 0$. It turns out that with the same assumption in the present environment, the city developer would always find it optimal to include the full range of sectors in the city under the respective CES assumptions on k_{ijt} and $\overline{k}_{it}$ in (29) and (30). Then there would not be a non–trivial analysis of urban industrial diversity. The simplest modeling strategy to get around the problem is to assume that $f'(s) > 0$ only for $s \geq \mu$, for some threshold $\mu > 0$, whereas $f(s)$ remains equal to 1 for $s < \mu$.

23

One possible justification for this admittedly unusual assumption is that learning from workers of a particular sector facilitated by personal contacts is probably not possible unless there are sufficiently many of the sector's workers present, for otherwise the interactions could just be too infrequent to give rise to any meaningful learning opportunities.[15],[16]

3.2 Optimal time division

The specification of the learning technology in (30) and (31) can be derived as the outcome of the worker's optimal division of time among learning from workers in each sector in the economy, in much the same way (2) and (10) fall out from the worker's optimal time division between the two kinds of learning investment in section 2.

Specifically, let x_{jt} denote the time the worker spends on learning from sector j workers. If the aggregate time investment is x_t,

$$\int_0^1 x_{jt} dj = x_t. \tag{32}$$

Assume that in period t the worker's human capital increases by an amount given by

$$\Delta_{it} = \left(\int_0^1 [k_{ijt} x_{jt}]^{\frac{\rho-1}{\rho}} dj \right)^{\frac{\rho}{\rho-1}}. \tag{33}$$

Taking the learning environment as given, the worker chooses x_{jt}, $j \in [0,1]$, to maximize (33) subject to (32). If $\overline{A}_{ijt} = \overline{A}_{jt} = \overline{A}_t$, and $s_{ijt} = s_{it}$ for each sector present in the city, the worker would find it optimal to set $x_{jt} = x_t^S$ for all j in which

---

[15] If for example learning takes place through pairwise meetings of workers, the matching of workers could be subject to some kind of increasing returns.

[16] Alternatively, one may assume that a certain local public good is an essential input to production in the city, and that the quantity required is increasing in the range of sectors present in the city. Then the city developer may find it optimal not to include the full range of sectors in the city, notwithstanding assuming $f'(s) > 0$ for $s \geq 0$. Such an assumption is arguably less plausible than assuming the kind of increasing returns explained above.

$s_{ijt} = s_{it} > \mu$ and $x_{jt} = x_t^B$ for all j in which $s_{ijt} = 0$, for some x_t^S and x_t^B. Then (33) reads

$$\Delta_{it} = \overline{A}_t \left(\frac{n_{it}}{s_{it}} \left[k\left(s_{it}\right) x_t^S \right]^{\frac{\rho-1}{\rho}} + \left(1 - \frac{n_{it}}{s_{it}}\right) \left[\kappa x_t^B\right]^{\frac{\rho-1}{\rho}} \right)^{\frac{\rho}{\rho-1}}, \tag{34}$$

while (32) becomes

$$\frac{n_{it}}{s_{it}} x_t^S + \left(1 - \frac{n_{it}}{s_{it}}\right) x_t^B = x_t. \tag{35}$$

Maximizing (34) with respect to x_t^S and x_t^B subject to (35) yields

$$x_t^S = \frac{x_t k\left(s_{it}\right)^{\rho-1}}{\frac{n_{it}}{s_{it}} k\left(s_{it}\right)^{\rho-1} + \left(1 - \frac{n_{it}}{s_{it}}\right)\kappa^{\rho-1}}, \tag{36}$$

$$x_t^B = \frac{x_t \kappa^{\rho-1}}{\frac{n_{it}}{s_{it}} k\left(s_{it}\right)^{\rho-1} + \left(1 - \frac{n_{it}}{s_{it}}\right)\kappa^{\rho-1}}. \tag{37}$$

Then (30) and (31) fall out from substituting (36) and (37) into (34), where $\sigma = \frac{1}{2-\rho}$.

3.3 Equilibrium sector size

The city developer's maximization remains as given by (15) in section 2, except that the developer in the multi–sector economy also faces the choice of the range of sectors to be included in the city. Given n_{it}, choosing the range of sectors to be included in the city is equivalent to choosing the size of each sector in the city, of which the first order condition reads

$$-\frac{\sigma}{\sigma-1}\left(k\left(s\right)^{\frac{\sigma-1}{\sigma}} - \kappa^{\frac{\sigma-1}{\sigma}}\right) + sk\left(s\right)^{1/\varepsilon-1/\sigma} f\left(s\right)^{-1/\varepsilon} f'\left(s\right) = 0, \tag{38}$$

where I have, without loss of generality, dropped the city and the time subscripts. Now (38) is simply the first order condition of maximizing $\overline{k}\left(s,n\right)$ with respect to s, where the left side is indeed diminishing in s given (14). Further, if

$$\lim_{s\to\infty} sk\left(s\right)^{1/\varepsilon-1/\sigma} f\left(s\right)^{-1/\varepsilon} f'\left(s\right) = 0, \tag{39}$$

there exists a unique solution for $s > \mu$ to the equation.[17]

Given city size, as s increases, the city may only host fewer sectors. With the fallen industrial diversity, the city's workers may only be able to learn through impersonal means from workers in sectors no longer present in the city. But as s increases, there would also be increasing concentration of workers from the remaining sectors in the city, allowing the city's workers to learn from them with greater intensity and effectiveness. In equilibrium, the marginal cost (the first term of (38)) and the marginal returns (the second term) are equated, giving rise to (38). The natural question to follow up is how the cost and returns to increasing sector size may be affected by an increase in η.

Proposition 4 *In case $\varepsilon < 1$, s is increasing in η if and only if $\varepsilon < \sigma$. In case $\varepsilon > 1$, s is increasing in η if and only if $\varepsilon > \sigma$.*

It turns out that an increase in η would raise the returns, as well as the cost, to increasing sector size if and only if $\varepsilon < \sigma$. In this event, if $\varepsilon < 1$ as well, the increase in marginal returns would exceed the increase in marginal cost, yielding a larger s in equilibrium. Conversely if $\varepsilon > \sigma$, and further if $\varepsilon > 1$, the decline in marginal returns would not be as large as the decline in marginal cost to also result in a larger s.

3.4 Equilibrium city size

Where (38) is an equation in s only, equilibrium sector size is completely isomorphic to city size. This special feature of the model helps simplify the analysis considerably

---

[17]With $f(\mu) = 1$, $k(\mu) = \kappa$. At $s = \mu$ then, the left side of (38) starts out equal to $\mu \kappa^{1/\varepsilon - 1/\sigma} f'(\mu) > 0$, and if (39) holds, it ends up falling below 0 as s becomes large. Had we continued to assume that $f(0) = 1$, so that $k(0) = \kappa$, the left side would start out equal to 0 at $s = 0$ and becomes negative thereafter. In this case, $\overline{k}(n,s)$ is maximized at the corner of $n/s = 1$; i.e., the city developer always finds it optimal to include the full range of sectors in the city.

since in the determination of city size, sector size can be treated as given. Thus given s, the city may have a population $n \in [0, s]$. As n increases from 0 to s, the range of sectors included in the city increases from 0 to 1, while $\overline{k}(n, s)$ rises from κ to $k(s)$. The equilibrium condition for n, as in the analysis in section 2, can be derived by combining the first order conditions of (15) with respect to n_{it} and τ_{it} and the zero–profit condition for city formation :

$$x\overline{k}(n, s)^{\frac{1-\sigma}{\sigma}} \frac{\sigma}{\sigma - 1} \frac{k(s)^{\frac{\sigma-1}{\sigma}} - \kappa^{\frac{\sigma-1}{\sigma}}}{s} = \frac{\omega\alpha}{2 + \alpha} \frac{n^{\alpha/2-1}}{\pi^{\alpha/2}}. \tag{40}$$

The left side of (40), denoting the marginal returns to agglomeration, is decreasing if and only if $\sigma > 1$ – an assumption that I should maintain for the following. The appendix shows that

Lemma 3 *(a) The marginal returns to agglomeration are increasing in η if $\varepsilon < 1 < \sigma$. (b) The marginal returns would decline if $\varepsilon > \sigma > 1$.*

Next, rewrite (40) as

$$x^D(n, s) = \frac{\omega\alpha}{2 + \alpha} \frac{n^{\alpha/2-1}}{\pi^{\alpha/2}} \frac{\overline{k}(n, s)^{\frac{\sigma-1}{\sigma}} s}{\left(k(s)^{\frac{\sigma-1}{\sigma}} - \kappa^{\frac{\sigma-1}{\sigma}}\right)} \frac{\sigma - 1}{\sigma}, \tag{41}$$

– a strictly increasing, and unbounded function of n that starts out equal to 0 at $n = 0$, the counterpart of (19) for the single–sector economy.

Equation (22) that relates how the worker's optimal x depends on n remains valid, except that k is replaced by $\overline{k}$:

$$x^W(n, s) \equiv \frac{1}{1 + \delta} \left[\delta - \frac{1 - \delta}{\overline{k}(n, s)}\right], \tag{42}$$

while the claims of lemma 2 continue to apply with some minor modification.

Lemma 4 $x^W(n)$ *is strictly increasing, strictly concave if $\sigma > 1/2$, bounded by (23), and positive over $n \geq 0$ if (24) holds.*

27

The two functions (41) and (42) are the counterparts of (19) and (22), respectively, for our previous analysis of the single-sector economy. In the present analysis, an interior symmetric constant growth equilibrium can be defined as a $n \in (0, s)$ that solves $x^D(n, s) = x^W(n, s)$, where s is the solution to (38). Such an equilibrium exists if $x^W(0, s) > x^D(0, s)$ and $x^D(s, s) > x^W(s, s)$. The first condition can easily be shown to hold given (24). Whether the second condition holds depends on the value of s at which x^D and x^W are evaluated. I should just proceed by assuming that the condition holds rather than to dwell on the details, in which case :

Proposition 5 *Given the existence of an interior equilibrium in the multi–sector economy, the equilibrium is unique if* (25) *holds.*

3.5 City size, urban industrial diversity, and improvements in communications technology

To begin analyzing how improvements in communications technology may affect city and sector sizes and urban industrial diversity, it is useful first to establish that :

Proposition 6 *An increase in the productivity of learning through impersonal means of communications η raises equilibrium city size n if $\varepsilon \leq 1 < \sigma$.*

The claim of proposition 6 is almost identical to the claim of proposition 2(a). The novelty in the extension to studying a multi–sector city equilibrium is that it allows an analysis of how urban industrial diversity; i.e., the range of sectors hosted in each city n/s, behaves with respect to changes in the learning environment. Table 1 summarizes the claims of propositions 4 and 6. Because the analysis is restricted to where $\sigma > 1$, there are just three mutually exclusive possibilities to consider.

Table 1 : Comparative steady state of η

		sector size $\partial s/\partial \eta$	city size $\partial n/\partial \eta$	urban industrial diversity $\partial (n/s) /\partial \eta$
1.	$\varepsilon \leq 1$	+	+	?
2a.	$1 < \sigma < \varepsilon$	+	?	?
2b.	$1 < \varepsilon < \sigma$	$-$	?	?

$\varepsilon \leq 1$. When the two learning technologies are complements, as a result of the increase in η, by propositions 4 and 6, respectively, the city and each of the sector it hosts get larger at the same time. Then urban industrial diversity as measured by n/s may increase or decline. Even though no firm analytical conclusions may be derived, there is good reason to believe that n/s should change in the same general direction as how n and s would change. If it pays to increase city size and the size of each sector in the city because of the increase in the marginal returns to agglomeration, it should pay to having more sectors located in the city too, for this serves the same purpose of taking advantage of the increase in the marginal returns to agglomeration. Figure 3 illustrates this tendency via a numerical example.[18]

$1 < \sigma < \varepsilon$. By proposition 4, if η increases, s increases too in equilibrium. Although proposition 6 does not cover how n would be affected, lemma 4 does establish that the marginal returns to agglomeration would fall in this case. In various numerical experiments, I find that, not surprisingly, for sufficiently large η, following the decline in the marginal returns to agglomeration, n must fall in equilibrium. Granted that there is a larger s in the mean time, the ratio n/s declines as a result.

---

[18]This and the next example assume $f(s) = 1 + (s - \mu)^{\phi}$, where $\phi = 0.6$ and $\mu = 1$. For this example, $\alpha = 3.5$, $\varepsilon = 0.65$, $\sigma = 1.3$, $\omega = 0.16$, and $\delta = 0.98$.
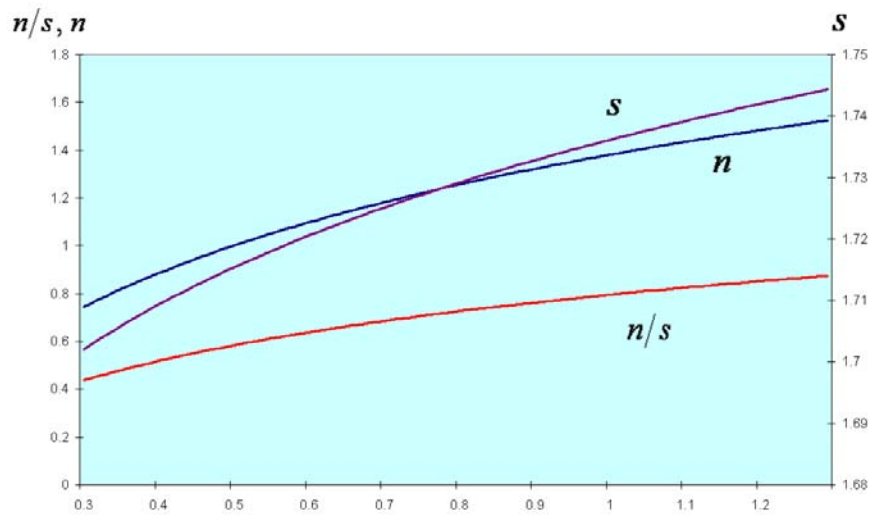
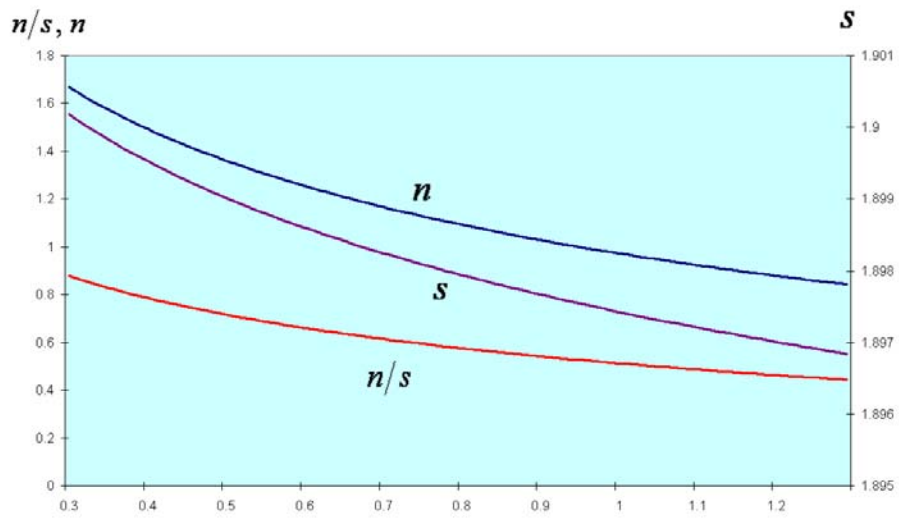Fig. 3. Urban industrial diversity; $\varepsilon \leq 1$

Fig. 4. Urban industrial diversity; $\sigma > \varepsilon > 1$

$1 < \varepsilon < \sigma$. By proposition 4, if η increases, there would be a smaller s in equilibrium. But this case is not covered by either lemma 4 nor proposition 6. To check how the marginal returns to agglomeration and indeed equilibrium city size would be affected by an increase in η, I resort to a numerical example. Figure 4 illustrates one typical instance of how n and n/s would be affected by the increase in η.[19] Where the two kinds of learning investment are substitutes, not surprisingly, these two variables are found to follow similar downward trends.

Summing Up When city size and urban industrial diversity are jointly determined, the two variables tend to move in the same direction with respect to advances in communications technology. In case the two kinds of learning investment are complements, city size, sector size, and urban industrial diversity tend to increase together. In case the two kinds of learning investment are substitutes, city size and urban industrial diversity tend to fall together.

4. POLICY IMPLICATIONS

4.1 Industrial policy

In the multi–sector economy, implicit in the city developer's maximization is also a choice of a particular set of sectors to be included in the city. Which set is unimportant under the symmetry assumption on the learning technology given in (28). Still, for the city to offer its residents the planned learning environment, its population has to be comprised of exactly s workers from each of the chosen sectors. On the contrary, given symmetry, workers are indifferent among which cities to stay in equilibrium. The problem that arises then is that an incentive compatible sorting mechanism to coordinate workers to move to their rightful locations appears lacking.

---

[19]In this example, $\alpha = 2.2$, $\varepsilon = 1.2$, $\sigma = 1.4$, $\omega = 0.1$, and $\delta = 0.98$.

The city developer has every interest to maximize $\overline{k}(n,s)$ so as to be able to levy the highest commuting charge and collect the maximum housing rent, while paying the least subsidy. To this end, once a set of sectors is chosen, the developer could choose to offer the subsidy only to workers belonging to the selected sectors. This helps exclude workers not in those sectors from choosing to reside in the city. Besides, the number of workers in each selected sector who would qualify for the subsidy should be limited to just the planned sector size. In effect, the developer may control the precise sectoral composition of the city through actively targeting subsidies to particular workers. Such practices are widespread among city and regional governments in many places in the world, and with cross–sector spillovers, they could serve an important coordinating function.

4.2 The social optimum

How may equilibrium differ from social optimality? Consider the model in section 2 for simplicity.[20] Granted that in equilibrium n maximizes per capita output net of the resources used up in commuting, subject to attaining a certain level of productivity growth, equilibrium city size should only coincide with socially optimum city size for a given level of investment. This is a familiar result : In static models of an urban system, a competitive market for cities would help deliver the first best allocation.[21] But the competitive market for cities could not be relied upon to help induce efficient investment on the part of workers in the presence of external effects.

First in the decentralized equilibrium, the worker takes as given the evolution of

---

[20]The statement of the planner's problem and its solution can be found in an appendix available for download in http://www.econ.hku.hk/~tsechung. The discussion to follow is equally applicable to the model in section 3.

[21]Henderson (1974), Flatters, Henderson and Mieszkowski (1974), Hamilton (1975), Arnott and Stigilitz (1979), and Henderson and Becker (2000), among others.

k_{it} in making investment decisions. In equilibrium, $k_{it} = \overline{A}_t k(n_{it})$. The planner takes this relation into account. When workers fail to internalize the externality that greater investment today helps *contribute* to raising the productivity of learning investment for all in the periods to follow, they tend to underinvest. Because the marginal returns to agglomeration are increasing in the level of investment, the city tends to be underpopulated as a result.

A similar externality is at work in the model of Black and Henderson (1999), where a worker's productivity depends on the average productivity of workers in the same city. But the sense in which the city in Black and Henderson is underpopulated is not quite the same as the sense in which the city in the present model could be underpopulated. In Black and Henderson, where the agglomeration economies are static production economies, equilibrium city size does not depend on the level of investment but only on the level of productivity. Thus given contemporaneous productivity, equilibrium city size indeed coincides with socially optimum city size. The cities in their model are underpopulated in the sense that the levels of productivity are suboptimal.

Unlike the model in Black and Henderson, where equilibrium investment is unambiguously suboptimal, in the present model, equilibrium investment could be excessive, and cities could become overpopulated because of the presence of a second externality. Recall that it costs the city developer $\omega \overline{A}_t$ units of output to supply one unit of urban traveling. To recover the expenditure, the developer charges each household a location cost of $z_{it} b_{it}^\alpha$ while subsidizing each household for the amount τ_{it} in the zero–profit equilibrium. In making investment decisions, the worker takes as given the evolutions of z_{it} and τ_{it} since to an individual household–worker, by how much they grow is a function of economywide productivity increase. The planner takes this relation into account. When workers fail to internalize the externality that greater investment today helps contribute to costlier commuting for all in the periods to follow, they tend to overinvest. Such externalities need not be peculiar to assuming

33

a public urban transportation sector. In general, such a negative externality would arise whenever labor is an essential input to the production of a local public good, for which labor productivity is stagnant – the case of the Baumol's disease.[22]

Could these investment externalities be internalized by the city developer via some particular local tax/subsidy, and does the city developer has incentives to do just that? Black and Henderson discuss in details two possibilities :

1. the city developer sets up a human capital requirement for entry into the city,

2. the city developer subsidizes the investment in human capital,

and conclude that in theory the answers are affirmative, but their implementations are problematical, to say the least.[23] They then argue it follows that the national government has important roles to play to subsidize human capital investment. In the present model, such a policy recommendation is not quite warranted before we are able to resolve whether equilibrium investment is indeed suboptimal or excessive.

Now suppose we are able to answer in certain that investment is suboptimal, perhaps from the quantitative analysis of the model under a set of plausible parameters. The first best intervention is of course for the national government to directly subsidize investment. Alternatively, the national government may choose to subsidize the provision of urban transportation in particular and in the supply of local public good in general. This would not only help move equilibrium city size, but also equilibrium

---

[22]Baumol (1967).

[23]The problem with the first option is that it is not clear how an entry requirement in human capital can be enforced in a free society. Besides, they point out that an individual developer has incentives to cheat to let in (or in the present model, to deny entry in case equilibrium investment turns out to be excessive) a marginal household who has not met the entry requirement. The problem with the second option is that with free mobility across cities, developers have incentives to "steal" households away from cities that have subsidied their human capital investment. In the present model and in case equilibrium investment is exessive, developers have incentives to provide tax asylums to households who are discouraged to invest in other locales.

investment, closer to optimum, given that in the decentralized equilibrium, workers would choose to invest a greater amount in the larger city. True such policy interventions could not be relied upon to attain the first best optimum : Even if each city is of the optimum size, the investment externalities remain. They could nevertheless help move equilibrium investment closer to optimum investment.

5. CONCLUDING REMARKS

In academia, we tend to believe that face–to–face and impersonal communications should always be complements. The better we read up on the relevant literature, the more fruitful the discussions with colleagues would become.[24] The finding by Kolko (2000) that internet domains tend to be disproportionately located in large cities perhaps could also be taken to imply that personal contacts and impersonal communications are complements.

The degree of complementarity between the two kinds of communications needs not be given and remains at the same level over time though. For example, one way in which the two kinds of communications could be complements is that often we learn from others about where information on a particular subject can be found. Such personal contacts could become increasingly valuable for two reasons :

1. As the stock of knowledge grows over time and becomes increasingly specialized, we depend increasingly on others as sources of references.

2. When it becomes easier to access publicly available information because of improvements in communications technology, a certain personal contact could lead to uncovering a much greater amount of information in the aftermath.

With the advent of internet search engines and related technologies, such complementaries could weaken, however, when technologies gradually come to replace people as sources of references. An investigation into the microeconomic underpin-

---

[24]I owe this observation to an editor of the journal.

nings of how the interaction between the two kinds of learning investment determines the degree of substitutability between them could advance our understanding of this important and interesting question.

APPENDIX

Proof of proposition 1

To establish uniqueness, combine (16) and (22) :

$$\left(\delta - (1-\delta) k(n)^{-1}\right) k(n)^{1/\varepsilon-1} f(n)^{-1/\varepsilon} f'(n) - \frac{\alpha(1+\delta)\omega}{(2+\alpha)\pi^{\alpha/2}} n^{\alpha/2-1} = 0, \qquad (43)$$

to obtain an equation in n only. Equilibrium is unique if this equation has at most one solution. Differentiating,

$$D = k(n)^{1/\varepsilon-1} f(n)^{-1/\varepsilon} \Phi + \frac{(2-\alpha)(1+\delta)\alpha\omega}{2(2+\alpha)\pi^{\alpha/2}} n^{\alpha/2-2}, \qquad (44)$$

where

$$\Phi \equiv \left(\delta - (1-\delta) k(n)^{-1}\right) \left(\left(\frac{1}{\varepsilon}-1\right) k(n)^{1/\varepsilon-1} f(n)^{-1/\varepsilon} f'(n)^2 - \frac{1}{\varepsilon}\frac{f'(n)^2}{f(n)} + f''(n)\right)$$
$$+ (1-\delta) k(n)^{1/\varepsilon-2} f(n)^{-1/\varepsilon} f'(n)^2.$$

The last term of (44) is negative if $\alpha > 2$; then it suffices to show that Φ is likewise negative. Now if (25) holds, since $k(n) \geq \kappa$,

$$\delta > \frac{2}{k(n)+2},$$

from which it follows that

$$\delta - (1-\delta) k(n)^{-1} > (1-\delta) k(n)^{-1}. \qquad (45)$$

Then

$$\Phi < \left(\delta - (1-\delta) k(n)^{-1}\right) \left[\frac{1}{\varepsilon} k(n)^{1/\varepsilon-1} f(n)^{-1/\varepsilon} f'(n)^2 - \frac{1}{\varepsilon}\frac{f'(n)^2}{f(n)} + f''(n)\right] < 0,$$

where the second inequality is by (14).

Proof of proposition 2

To establish (a), totally differentiate (43) to obtain

$$\frac{dn}{d\eta} = \frac{k\left(n\right)^{2/\varepsilon-2} f\left(n\right)^{-1/\varepsilon} \eta^{-1/\varepsilon} f'\left(n\right)}{-D} \left([1-\delta]\, k\left(n\right)^{-1} + \left(\delta - [1-\delta]\, k\left(n\right)^{-1}\right)\left(\frac{1}{\varepsilon}-1\right) \right).$$
(46)

First $D < 0$ if (25) holds, while the bracketed term is guaranteed to be positive if $\varepsilon < 1$. To establish (b), by (22),

$$\frac{dx}{d\eta} = \frac{1-\delta}{1+\delta} k\left(n\right)^{-2} \frac{dk\left(n\right)}{d\eta},$$
(47)

while

$$
\begin{aligned}
\frac{dk\left(n\right)}{d\eta} &= \frac{\partial k\left(n\right)}{\partial \eta} + \frac{\partial k\left(n\right)}{\partial n}\frac{\partial n}{\partial \eta} \\
&= \frac{k\left(n\right)^{1/\varepsilon}}{\eta^{1/\varepsilon}}\left(1 + \eta^{1/\varepsilon} f\left(n\right)^{-1/\varepsilon} f'\left(n\right)\frac{\partial n}{\partial \eta}\right) \\
&= \frac{k\left(n\right)^{1/\varepsilon}}{\eta^{1/\varepsilon}}\left(1 - \frac{(1-\delta)\, k\left(n\right)^{-1} + \left(\delta - [1-\delta]\, k\left(n\right)^{-1}\right)\left(\frac{1}{\varepsilon}-1\right)}{f\left(n\right)^{2/\varepsilon} f'\left(n\right)^{-2} k\left(n\right)^{2-2/\varepsilon} D}\right),
\end{aligned}
$$
(48)

where the last equality is by (46). Since $D < 0$, the sign of $dk/d\eta$ is the same as the sign of

$$
\begin{aligned}
\Psi &= (1-\delta)\, k\left(n\right)^{-1} + \left(\delta - [1-\delta]\, k\left(n\right)^{-1}\right)\left(\frac{1}{\varepsilon}-1\right) - f\left(n\right)^{2/\varepsilon} f'\left(n\right)^{-2} k\left(n\right)^{2-2/\varepsilon} D \\
&= \frac{(\alpha-2)\left(1+\delta\right)\alpha\omega}{2\left(2+\alpha\right)\pi^{\alpha/2}} n^{\alpha/2-2}\frac{f\left(n\right)^{2/\varepsilon}}{f'\left(n\right)^2} k\left(n\right)^{2-2/\varepsilon} - \left(\delta - [1-\delta]\, k\left(n\right)^{-1}\right)\times \\
&\qquad \left(k\left(n\right)^{1-1/\varepsilon} f\left(n\right)^{1/\varepsilon} f'\left(n\right)^2 f''\left(n\right) - \frac{1}{\varepsilon} k\left(n\right)^{1-1/\varepsilon} f\left(n\right)^{1/\varepsilon-1}\right).
\end{aligned}
$$

The last bracketed term is negative by (14) and thus $\Psi > 0$.

Proof of proposition 3

By (11), (7), and that $\overline{A}_{it} = \overline{A}_t$,

$$\frac{dx^P}{d\eta} = f\left(n\right)^{1-1/\varepsilon} k\left(n\right)^{1/\varepsilon-1}\left(\frac{dx}{d\eta} + x\left(\frac{1}{\varepsilon}-1\right)\times\right.$$

$$\left[k\left(n\right)^{1/\varepsilon-1}\eta^{-1/\varepsilon} + \left(k\left(n\right)^{1/\varepsilon-1} f\left(n\right)^{-1/\varepsilon} - f\left(n\right)^{-1}\right) f'\left(n\right)\frac{dn}{d\eta}\right]\right)$$

$$= \left(\frac{f'\left(n\right)^2}{f\left(n\right)} - f''\left(n\right) + \frac{\alpha-2}{2}n^{-1}\right)\frac{f\left(n\right)^{1-1/\varepsilon}}{f'\left(n\right)k\left(n\right)}x\frac{dn}{d\eta},$$

where the second inequality is by (46), (47), and (16). In case $\varepsilon \leq 1$, $dn/d\eta > 0$. Then $dx^P/d\eta$ is guaranteed to be positive if (26) holds. Suppose $\varepsilon > 1$ and further suppose $dn/d\eta < 0$. Then $dx^P/d\eta$ is guaranteed to be negative if the terms inside the bracket sum to a positive expression, which follows from (14). Next by (8), (11), and that $\overline{A}_{it} = \overline{A}_t$,

$$\frac{dx^I}{d\eta} = \eta^{1-1/\varepsilon}k\left(n\right)^{1/\varepsilon-1}\left(\frac{dx}{d\eta} + x\left(\frac{1}{\varepsilon}-1\right)\left[k\left(n\right)^{1/\varepsilon-1} f\left(n\right)^{-1/\varepsilon} f'\left(n\right)\frac{dn}{d\eta} - \eta^{-1}\right]\right)$$

$$= \frac{-D\eta\left(\frac{dn}{d\eta}\right)^2 + (1-\delta)k\left(n\right)^{2/\varepsilon-3}\eta^{1-2/\varepsilon}}{1+\delta} + xk\left(n\right)^{1/\varepsilon-1}\eta^{-1/\varepsilon}\left(1 - \frac{1}{\varepsilon}\right),$$

where the second equality is once again by (47) and (46). Then $dx^I/d\eta$ is guaranteed to be positive if $\varepsilon \geq 1$.

Proof of proposition 4

Totally differentiating (38) yields

$$\frac{\partial s}{\partial\eta} = \frac{f\left(s\right)^{1/\varepsilon}\left(1 - \left(\frac{\kappa}{k(s)}\right)^{1/\varepsilon-1/\sigma} - sf\left(s\right)^{-1/\varepsilon} f'\left(s\right)\left(\frac{1}{\varepsilon}-\frac{1}{\sigma}\right) k\left(s\right)^{1/\varepsilon-1}\right)}{s\eta^{1/\varepsilon}\left(\left(\frac{1}{\varepsilon}-\frac{1}{\sigma}\right) k\left(s\right)^{1/\varepsilon-1} f\left(s\right)^{-1/\varepsilon} f'\left(s\right)^2 - \frac{1}{\varepsilon} f\left(s\right)^{-1} f'\left(s\right)^2 + f''\left(s\right)\right)}. \quad (49)$$

The denominator is negative by (14). Multiply the numerator by $f\left(s\right)^{-1/\varepsilon} k\left(s\right)^{1-1/\sigma}$. The sign of $\partial s/\partial\eta$ is thus the same as the sign of

$$\Omega \equiv -k\left(s\right)^{\frac{\sigma-1}{\sigma}} + \kappa^{\frac{\sigma-1}{\sigma}}\left(\frac{k\left(s\right)}{\kappa}\right)^{\frac{\varepsilon-1}{\varepsilon}} + sf\left(s\right)^{-1/\varepsilon} f'\left(s\right)\left(\frac{1}{\varepsilon} - \frac{1}{\sigma}\right) k\left(s\right)^{1/\varepsilon-1/\sigma}$$

$$= -k\left(s\right)^{\frac{\sigma-1}{\sigma}} + \kappa^{\frac{\sigma-1}{\sigma}}\left(\frac{k\left(s\right)}{\kappa}\right)^{\frac{\varepsilon-1}{\varepsilon}} + \frac{\sigma-\varepsilon}{\varepsilon\left(\sigma-1\right)}\left(k\left(s\right)^{\frac{\sigma-1}{\sigma}} - \kappa^{\frac{\sigma-1}{\sigma}}\right) \quad \text{(by } (38))$$

$$= \kappa^{1/\varepsilon-1/\sigma}\left(k\left(s\right)^{\frac{\varepsilon-1}{\varepsilon}} - \kappa^{\frac{\varepsilon-1}{\varepsilon}}\right) - \frac{\sigma}{\sigma-1}\frac{\varepsilon-1}{\varepsilon}\left(k\left(s\right)^{\frac{\sigma-1}{\sigma}} - \kappa^{\frac{\sigma-1}{\sigma}}\right)$$

$$= \kappa^{1/\varepsilon-1/\sigma}\left(f\left(s\right)^{\frac{\varepsilon-1}{\varepsilon}} - 1\right) - \frac{\sigma}{\sigma-1}\frac{\varepsilon-1}{\varepsilon}\left(k\left(s\right)^{\frac{\sigma-1}{\sigma}} - \kappa^{\frac{\sigma-1}{\sigma}}\right) \quad \text{(by } (12) \text{ and } (13)),$$

39

which vanishes at $s = \mu$ since $f(\mu) = 1$ and $k(\mu) = \kappa$. Differentiating with respect to s,

$$\frac{\partial \Omega}{\partial s} = \frac{\varepsilon - 1}{\varepsilon} f(s)^{-1/\varepsilon} f'(s) \left(\kappa^{1/\varepsilon - 1/\sigma} - k(s)^{1/\varepsilon - 1/\sigma} \right).$$

Since $\kappa < k(s)$ for $s > \mu$, if $\varepsilon > 1$, $\partial \Omega / \partial s > 0$ iff $\varepsilon > \sigma$. Conversely, if $\varepsilon < 1$, $\partial \Omega / \partial s > 0$ iff $\varepsilon < \sigma$.

Proof of lemma 3

First substitute (38) into (40) :

$$x \overline{k}(n,s)^{\frac{1-\sigma}{\sigma}} k(s)^{1/\varepsilon - 1/\sigma} f(s)^{-1/\varepsilon} f'(s) = \frac{2\omega}{2 + \alpha} \frac{\alpha}{2} \frac{n^{\alpha/2 - 1}}{\pi^{\alpha/2}}. \tag{50}$$

Totally differentiating the left–side with respect to η, while recognizing $\partial \overline{k} / \partial s = 0$ holds at where s satisfies (38), and then multiplying through by

$$x^{-1} \overline{k}(n,s)^{1-1/\sigma} k(s)^{1/\sigma - 1/\varepsilon} f(s)^{1/\varepsilon}$$

yields

$$\begin{aligned}
\Theta &\equiv \frac{1 - \sigma}{\sigma} \overline{k}(n,s)^{\frac{1-\sigma}{\sigma}} f'(s) \left(\frac{n}{s} k(s)^{1/\varepsilon - 1/\sigma} \eta^{-1/\varepsilon} + \left(1 - \frac{n}{s}\right) \kappa^{1/\varepsilon - 1/\sigma} \eta^{-1/\varepsilon} \right) + \\
&\quad \left(\frac{1}{\varepsilon} - \frac{1}{\sigma} \right) k(s)^{\frac{1-\varepsilon}{\varepsilon}} \eta^{-1/\varepsilon} + \frac{\partial s}{\partial \eta} \left\{ \left(\frac{1}{\varepsilon} - \frac{1}{\sigma} \right) k(s)^{\frac{1-\varepsilon}{\varepsilon}} f(s)^{-1/\varepsilon} f'(s)^2 \right. \\
&\quad \left. - \frac{1}{\varepsilon} f(s)^{-1} f'(s)^2 + f''(s) \right\},
\end{aligned} \tag{51}$$

where $\partial s / \partial \eta$ is given by (49). To proceed, substitute (30) and (49) into (51),

$$\begin{aligned}
\Theta &= \left(\frac{f(s)}{\eta} \right)^{1/\varepsilon} s^{-1} \left(1 - \left[\frac{\kappa}{k(s)} \right]^{1-1/\sigma} \right) \times \\
&\quad \left(\frac{1 - [\kappa/k(s)]^{1/\varepsilon - 1/\sigma}}{1 - [\kappa/k(s)]^{1-1/\sigma}} - \frac{\frac{n}{s} k(s)^{\frac{\sigma-1}{\sigma}} + \left(1 - \frac{n}{s}\right) \kappa^{\frac{\sigma-1}{\sigma}} [k(s)/\kappa]^{\frac{\varepsilon-1}{\varepsilon}}}{\frac{n}{s} k(s)^{\frac{\sigma-1}{\sigma}} + \left(1 - \frac{n}{s}\right) \kappa^{\frac{\sigma-1}{\sigma}}} \right).
\end{aligned}$$

The above is negative if $\varepsilon > \sigma > 1$. This proves (b). If $\varepsilon < 1 < \sigma$,

$$\Theta > \left(\frac{f(s)}{\eta} \right)^{1/\varepsilon} s^{-1} \left(\left[\frac{\kappa}{k(s)} \right]^{1-1/\sigma} - \left[\frac{\kappa}{k(s)} \right]^{1/\varepsilon - 1/\sigma} \right) > 0.$$

This proves (a).

40

REFERENCES

[1] Abdel-Rahman, Hesham M. and Anas, Alex, 2004, "Theories of systems of cities," in J. V. Henderson and J. F. Thisse (eds), *Handbook of Regional and Urban Economics,* Vol. 4, Elsevier, Amsterdam.

[2] Arnott, Richard J. and Stiglitz, Joseph E., 1979, "Aggregate Land Rents, Expenditure on Public Goods, and Optimal City Size," *Quarterly Journal of Economics* 92, 471–500.

[3] Audretsch, David B. and Feldman, Maryann P., 1996, "R&D Spillovers and the Geography of Innovation and Production," *American Economic Review* 86, 630–640.

[4] Baumol, William J., 1967, "Macroeconomics of Unbalanced Growth : The Anatomy of Urban Crisis," *American Economic Review* 57, 415–426.

[5] Black, Duncan and Henderson, Vernon, 1999, "A Theory of Urban Growth," *Journal of Political Economy* 107, 252–284.

[6] Berliant, Marcus; Reed, Robert R. III, and Wang, Ping, 2006, "Knowledge Exchange, Matching, and Agglomeration," *Journal of Urban Economics* 60, 69-95.

[7] Berliant, Marcus and Wang, Ping, 2004, "Dynamic Urban Models: Agglomeration and Growth" in *Urban dynamics and growth: Advances in urban economics*, 533-81, Elsevier, Amsterdam.

[8] Chinitz, Benjamin, 1999, "The Influence of Communications and Data Processing Technology on Urban form," *Research in Urban Economics* 12, 151–161.

[9] Duranton, Gilles and Puga, Diego, 2000, "Diversity and Specialization in Cities: Why, Where and When does it Matter?", *Urban Studies* 37, 533-555.

[10] —, 2001, "Nursery Cities : Urban Diversity, Process Innovation, and the Life Cycle of Products," *American Economics Review* 91, 1454–1477.

[11] —, 2004, "Micro-foundations of urban agglomeration economies," in J.V. Henderson and J.F. Thisse (eds), *Handbook of Regional and Urban Economics,* Vol. 4, Elsevier, Amsterdam.

[12] Eaton, Jonathan and Eckstein, Zvi, 1997, "Cities and growth: Theory and evidence from France and Japan," *Regional Science and Urban Economics* 27, 443–474.

[13] Flatters, Frank; Henderson, J. Vernon; and Mieszkowsk, Peter, 1974, "Public Goods, Efficiency, and Regional Fiscal Equalization," *Journal of Public Economics* 3, 99–112.

[14] Fujita, Masahisa and Thisse, Jacques–Francois, 2002, *Economics of Agglomeration, Cities, Industrial Location, and Regional Growth*, Cambridge University Press, Cambridge.

[15] Gasper, Jess and Glaeser, Edward L., 1998, "Information Technology and the Future of Cities," *Journal of Urban Economics* 43, 136–156.

[16] Glaeser, Edward L. and Kohlhase, Janet E., 2004, "Cities, Regions and the Decline of Transport Costs," *Papers in Regional Science* 83, 197–228.

[17] Glaeser, Edward L. and Maré, David C., 2001, "Cities and Skills," *Journal of Labor Economics* 19, 316–342.

[18] Hamilton, Bruce W., 1975, "Zonging and property taxation in a system of local governments," *Urban Studies* 12, 205-211.

[19] Henderson, J. V., 1974, "The sizes and types of cities," *American Economic Review* 64, 640-656.

[20] Henderson, Vernon and Becker, Randy, 2000,"Political Economy of City Sizes and Formation," *Journal of Urban Economics* 48, 453–484.

[21] Jaffe, Adam B.; Trajtenberg, Manuel, and Henderson, Rebecca, 1993, "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations," *Quaterly Journal of Economics* 108, 577-598.

[22] Jovanovic, Boyan and Rob, Rafael, 1989, "The Growth and Diffusion of Knowledge," *Review of Economic Studies* 56, 569–582.

[23] Kolko, Jed, 2000, "The Death of Cities? The Death of Distance? Evidence from the Geography of Commercial Internet Usage," in Ingo Vogelsang and Bengamin M. Compaine (eds), *The Internet Upheaval; Raising Questions, Seeking Answers in Communications Policy*, MIT Press, Cambridge, Mass.

[24] Leamer, Edward E. and Storper, Michael, 2001, "The Economic Geography of the Internet Age," *Journal of International Business Studies* 32, 641–665.

[25] Lucas, Robert E. Jr., 1988, "On the Mechanics of Economic Development," *Journal of Monetary Economics* 22, 3–42.

[26] Marshall, Alfred, 1890, *Principles of Economics,* Macmillan, London.

[27] Pascal, Anthony H. and McCall, John J., 1980, "Agglomeration Economies, Search Costs, and Industrial Location, *Journal of Urban Economics* 8, 383-388.

[28] Rossi–Hansberg, Esteban and Wright, Mark L. J., 2007, "Urban Structure and Growth," *Review of Economic Studies* 74, 597 - 624.

[29] Swann, G.M. Peter, 1999, "The Internet and the Distribution of Economic Activity," in S. Macdonald and J. Nightingale (eds), *Information and Organization,* Elsevier Science B.V., Amsterdam.

[30] Thompson, Peter, 2006, "Patent Citations and the Geography of Knowledge Spillovers: Evidence from Inventor- and Examiner-added Citations," *Review of Economics and Statistics* 88, 383–388.

APPENDIX NOT TO BE CONSIDERED FOR PUBLICATION

The social optimum

In the social optimum, given A_0, the planner maximizes the utility of the representative household :

$$U = \sum_{t=0}^{\infty} \delta^t \ln c_t,$$

subject to

$$c_t = A_t (1 - x_t) - \frac{2\omega A_t}{2 + \alpha} \frac{n_t^{\alpha/2}}{\pi^{\alpha/2}}, \tag{52}$$

$$A_{t+1} = A_t + A_t k(n_t) x_t. \tag{53}$$

by choosing (x_t, n_t). The value function of maximizing U reads

$$U(A_t) = \max_{A_{t+1}, n_t} \left\{ \ln \left[A_t \left(1 - \frac{A_{t+1} - A_t}{A_t k(n_t)} \right) - \frac{2\omega A_t}{2 + \alpha} \frac{n_t^{\alpha/2}}{\pi^{\alpha/2}} \right] + \delta U(A_{t+1}) \right\}. \tag{54}$$

The first order conditions are

$$- \left[1 - x_t - \frac{2\omega A_t}{2 + \alpha} \frac{n_t^{\alpha/2}}{\pi^{\alpha/2}} \right]^{-1} \frac{1}{A_t k(n_t)} + \delta U'(A_{t+1}) = 0, \tag{55}$$

$$x_t k(n_t)^{1/\varepsilon - 1} f(n_t)^{-1/\varepsilon} f'(n_t) - \frac{2\omega}{2 + \alpha} \frac{\alpha}{2} \frac{n_t^{\alpha/2 - 1}}{\pi^{\alpha/2}} = 0. \tag{56}$$

Differentiating (54) with respect to A_t yields

$$U'(A_t) = \frac{1}{A_t} \left(1 + \left[1 - x_t - \frac{2\omega}{2 + \alpha} \frac{n^{\alpha/2}}{\pi^{\alpha/2}} \right]^{-1} \left(x_t + \frac{1}{k(n_t)} \right) \right).$$

and substitute the result into (55) yields

$$- \left[1 - x_t - \frac{2\omega A_t}{2 + \alpha} \frac{n_t^{\alpha/2}}{\pi^{\alpha/2}} \right]^{-1} + \delta \frac{k(n_t)}{1 + k(n_t) x_t} \left(1 + \left[1 - x_{t+1} - \frac{2\omega}{2 + \alpha} \frac{n_{t+1}^{\alpha/2}}{\pi^{\alpha/2}} \right]^{-1} \times \right.$$

$$\left. \left(x_{t+1} + \frac{1}{k(n_{t+1})} \right) \right) = 0, \tag{57}$$

45

The optimum solution is a (x_t, n_t) pair that satisfies (56) and (57). Because the state variable A_t does not appear in either equation, the system (56) and (57) is solved by a time–stationary pair (x, n). In this case, (57) simplifies to

$$x^* (n_t) = \delta \left(1 - \frac{2\omega}{2 + \alpha} \frac{n_t^{\alpha/2}}{\pi^{\alpha/2}} \right) + (1 - \delta) k (n_t)^{-1} . \tag{58}$$

The first optimality condition (56) is identical to the equilibrium condition (16). Thus given x_t, equilibrium city size coincides with socially optimum city size. But optimum investment $x^* (n)$ differs from equilibrium investment $x^W (n)$, given by (22) in two important respects.

1. In the free market equilibrium, the worker takes as given the evolution of k_{it} in making investment decisions. In the symmetric constant growth equilibrium, $k_{it} = \overline{A}_t k (n_{it})$. The planner takes this relation into account. It can be shown that if (2) in the worker's maximization is replaced by (53), in equilibrium,

$$x^D (n) = \delta + (1 - \delta) k (n)^{-1} ,$$

   would obtain in place of (22). That is, workers tend to underinvest when they fail to internalize the externality that greater investment today helps contribute to raising the productivity of learning investment for all in the periods to follow.

2. It costs the city developer $\omega \overline{A}_t$ units of output to supply one unit of urban traveling. To recover the expenditure, the developer charges each household a location cost of $z_{it} b_{it}^\alpha$ while subsidizing each household for the amount τ_{it} in the zero–profit equilibrium. In making investment decisions, the worker takes as given the evolutions of z_{it} and τ_{it} since to an individual household–worker, by how much they grow is a function of economywide productivity increase. The planner takes this relation into account. It can be shown that if $-z_{it} b_{it}^\alpha + \tau_{it}$ in

46

(3) is replaced by $-\frac{2\omega A_t}{2+\alpha}\frac{n_{it}^{\alpha/2}}{\pi^{\alpha/2}}$ via the zero–profit condition for city formation, in equilibrium,

$$x^D(n) = \frac{\delta\left(1 - \frac{2\omega}{2+\alpha}\frac{n^{\alpha/2}}{\pi^{\alpha/2}}\right) + (1-\delta)\,k\,(n)^{-1}}{1+\delta}$$

would obtain in place of (22). That is, workers tend to overinvest when they fail to internalize the externality that greater investment today helps contribute to costlier commuting for all in the periods to follow.

Derivation of (4)

By (2), (3) reads

$$V_{it}(A_t) = \max_{A_{t+1}}\left\{A_t\left(1 - \frac{A_{t+1} - A_t}{k_{it}}\right) - z_{it}b_t^\alpha + \tau_{it} + \beta V_{it+1}(A_{t+1})\right\}.$$

Taking first order condition :

$$-\frac{A_t}{k_{it}} + \beta V'_{it+1}(A_{t+1}) = 0. \tag{59}$$

Differentiate $V_{it}(A_t)$ with respect to A_t :

$$V'_{it}(A_t) = 1 - \frac{A_{t+1} - A_t}{k_{it}} + \frac{A_t}{k_{it}},$$

and substitute the result into (59) yields (4).

Derivation of (16)

The Lagrangian of (15) reads

$$\mathcal{L} = \max_{\{n_{it}, z_{it}, \tau_{it}, \lambda\}}\left\{\left(z_{it} - \frac{2\omega A_t}{2+\alpha}\right)\frac{n_{it}^{1+\alpha/2}}{\pi^{\alpha/2}} - n_{it}\tau_{it} + \lambda\left(\tilde{V}_{it}(A_t) - V_t(A_t)\right)\right\}. \tag{60}$$

Taking first order conditions with respect to τ_{it} and n_{it},

$$-n_{it} + \lambda = 0, \tag{61}$$

$$\left(z_{it} - \frac{2\omega A_t}{2+\alpha}\right)\frac{2+\alpha}{2}\frac{n_{it}^{\alpha/2}}{\pi^{\alpha/2}} - \tau_{it} + \lambda\left(\frac{A_{t+1}-A_t}{k\left(n_{it}\right)^2}\frac{\partial k\left(n_{it}\right)}{\partial n_{it}} - z_{it}\frac{\alpha}{2}\frac{n_{it}^{\alpha/2-1}}{\pi^{\alpha/2}}\right) = 0. \quad (62)$$

The zero profit condition is

$$\left(z_{it} - \frac{2\omega A_t}{2+\alpha}\right)\frac{n_{it}^{1+\alpha/2}}{\pi^{\alpha/2}} - n_{it}\tau_{it} = 0 \Rightarrow \tau_{it} = \left(z_{it} - \frac{2\omega A_t}{2+\alpha}\right)\frac{n_{it}^{\alpha/2}}{\pi^{\alpha/2}}. \quad (63)$$

Combining $(61) - (63)$, while evaluating

$$\frac{\partial k\left(n_{it}\right)}{\partial n_{it}} = k\left(n_{it}\right)^{1/\varepsilon} f\left(n_{it}\right)^{-1/\varepsilon} f'\left(n_{it}\right), \quad (64)$$

using (10) yields (16). Finally the first order condition of (60) with respect to z_{it} :

$$\frac{n_{it}^{1+\alpha/2}}{\pi^{\alpha/2}} + \lambda\left(\frac{n_{it}}{\pi}\right)^{\alpha/2} = 0$$

is satisfied at any z_{it} with λ given by (61). In this case, given n_{it}, τ_{it} and z_{it} jointly satisfy (63).

Proof of lemma 1

Differentiating the left–side of (16) with respect to η yields

$$x\frac{1-\varepsilon}{\varepsilon}k\left(n\right)^{2/\varepsilon-2}\eta^{-1/\varepsilon}f\left(n\right)^{-1/\varepsilon}f'\left(n\right) \gtrless 0 \quad \text{iff } \varepsilon \lessgtr 1.$$

Proof of lemma 2

Differentiate (22) with respect to n:

$$\frac{\partial x^W\left(n\right)}{\partial n} = \frac{1-\delta}{1+\delta}k\left(n\right)^{1/\varepsilon-2}f\left(n\right)^{-1/\varepsilon}f'\left(n\right) > 0. \quad (65)$$

Further differentiating with respect to n yields a negative expression by (14). Thus, (22) is a strictly increasing and strictly concave function of n. Differentiating the left side of (22) with respect to η yields

$$\frac{\partial x^W\left(n\right)}{\partial \eta} = \frac{1-\delta}{1+\delta}k\left(n\right)^{1/\varepsilon-2}\eta^{-1/\varepsilon} > 0.$$

Derivations of (38) **and** (40)

The Lagrangian for the profit maximization remains given by (60), except that k_{it} is replaced by $\overline{k}_{it}$. In this case, maximizing (60) with respect to s_{it} is simply maximizing $\overline{k}(n_{it}, s_{it})$ with respect to s_{it}. Evaluating the partial derivatives using (30) yields (38). As in the derivation of (16), (40) is obtained by combining the first order conditions of (60) with respect to n_{it} and τ_{it} and the zero-profit condition (63), except that $\partial k(n_{it})/\partial n_{it}$ in (62) is replaced by $\partial \overline{k}(n_{it}, s_{it})/\partial n_{it}$.

Proof of lemma 4

Differentiate (42) with respect to n :

$$\frac{\partial x^W}{\partial n} = \frac{1-\delta}{1+\delta}\overline{k}(n,s)^{1/\sigma-2}\frac{\sigma}{\sigma-1}\frac{k(s)^{\frac{\sigma-1}{\sigma}} - \kappa^{\frac{\sigma-1}{\sigma}}}{s}$$

Differentiate further :

$$\frac{\partial^2 x^W}{\partial n^2} = \frac{1-\delta}{1+\delta}\left(\frac{1}{\sigma} - 2\right)\overline{k}(n,s)^{2/\sigma-3}\left(\frac{\sigma}{\sigma-1}\frac{k(s)^{\frac{\sigma-1}{\sigma}} - \kappa^{\frac{\sigma-1}{\sigma}}}{s}\right)^2 < 0,$$

where $\sigma > 1/2$.

Proof of proposition 5

Substitute (42) into (40) to obtain an equation in n only :

$$\frac{\delta - (1-\delta)\overline{k}(n,s)^{-1}}{1+\delta}\overline{k}(n,s)^{\frac{1-\sigma}{\sigma}}\frac{\sigma}{\sigma-1}\frac{k(s)^{\frac{\sigma-1}{\sigma}} - \kappa^{\frac{\sigma-1}{\sigma}}}{s} = \frac{2\omega}{2+\alpha}\frac{\alpha}{2}\frac{n^{\alpha/2-1}}{\pi^{\alpha/2}}. \qquad (66)$$

Since the right side is increasing in n, the equation has at most one solution if the left side is decreasing in n. If (25) holds, the condition follows, as can be established by straightforward differentiation.

Proof of proposition 6

Given $\varepsilon \leq 1 < \sigma$, by lemma 3, the term

$$\overline{k}(n,s)^{\frac{1-\sigma}{\sigma}} \frac{\sigma}{\sigma-1} \frac{k(s)^{\frac{\sigma-1}{\sigma}} - \kappa^{\frac{\sigma-1}{\sigma}}}{s}$$

in (66), denoting the marginal returns to agglomeration normalized by x, is increasing in η. Further,

$$x^W(n,s) = \frac{\delta - (1-\delta)\overline{k}(n,s)^{-1}}{1+\delta}$$

in (66) is likewise increasing in η since by lemma 4 $\partial \overline{k}/\partial n > 0$, and then $\partial \overline{k}/\partial s = 0$ at where s satisfies (38). Granted that the whole of the left side of (66) is decreasing in n, the proposition follows.