Vanishing cities: What does the New Economic Geography imply about the efficiency of urbanization?

Alex Anas

Department of Economics
State University of New York at Buffalo
Amherst, New York 14260.
E-mail: aanas@adelphia.net

May 2002
Revised: November 2002

ABSTRACT

How should the size and number of cities evolve optimally as population grows? Stripped of the constraints of geography itself, the setup of the New Economic Geography implies that de-agglomeration (or de-urbanization) is efficient. The number of cities increases while the size of each decreases on the optimal path until the economy suddenly disperses to tiny towns of stand-alone firms each specializing in a unique good. The cause of this narrow result is the NEG's strong emphasis on intercity trade to satisfy the taste for more goods. For the same aggregate population, a system of smaller cities saves time lost in commuting, has a larger labor supply and makes more goods than does a system of larger cities. Falling interurban trading costs favor this de-urbanization process. Only if intraurban commuting costs fall sufficiently, can a pattern of growing city sizes be efficient with growing population. Of course, when the number of cities or the geographic space itself is limited or asymmetric, then agglomeration can arise as an artifact of the constraints imposed by geography as demonstrated by numerous NEG models. This reveals that the central agglomerative force in the NEG is space itself and not the underlying economic relations.

Key words**:** City sizes, growth, agglomeration, trade. ***JEL classification*****:** D62, F12, R12

Vanishing cities: What does the New Economic Geography imply about the efficiency of urbanization? [1]

May 2002
Revised: November, 2002

1. Introduction

The birth of the New Economic Geography (NEG) can be traced to Krugman (1980) who started adopting this unique set of functional forms soon after the Dixit-Stiglitz model made its debut in 1977. Since then, and especially in recent years, a rich variety of complex extensions of the basic model have been developed and studied. For some of the most recent contributions, see Fujita, Krugman and Mori (1999) and Fujita, Krugman, Venables (1999). But although this young field has zoomed ahead in recent years, we still do not have a clear grip on many consequences of the underlying economic toolbox employed in the NEG model. Clarity has been difficult to achieve because, in all previous applications, the fundamental economic forces that are operative in the model – the so called *centripetal* and *centrifugal* forces – are obscured and confounded by the emphasis on the asymmetries and spatial limitations that are inherent in the specific geographies that are studied.

Take a very basic and important question: what does the NEG imply about the efficiency of urbanization? A variant of the NEG model developed here will allow us to investigate this issue by seeing how city sizes and the number of cities *should* evolve under population growth, falling interurban trading cost and falling intra-urban

[1] The paper was presented at MacMaster University on September 21, 2002 at a conference honoring Yorgos Papageorgiou. I thank David Pines and Takatoshi Tabuchi for their comments and discussions on several points. Useful comments were also received from three anonymous referees of this journal and from the editor.

commuting cost. Within the NEG tradition, this question would be investigated by working out the model over a specific geographic space (much as is done in Fujita, Krugman, Mori (1999)). It turns out, however, that stripping away the geography – as we do in this paper – is insightful from a theoretical standpoint because it reveals the true nature of the underlying economic forces inherent in the NEG. Our findings imply that the constraints of specific geographies act as a centripetal force that is the cause of spatial agglomeration. Once these constraints are removed, the economics inherent in the NEG toolbox imply that agglomeration is ultimately inefficient under population growth: the centrifugal forces swamp the centripetal ones and cities vanish. This is an undesirable result in the sense that a theory of agglomeration should be able to generate agglomeration even in the absence of geography.

It is important to note at the outset that the NEG has three drawbacks that are increasingly recognized. *First*, it totally neglects normative issues. *Second*, the basic NEG setup exaggerates the importance of trade among cities by assuming that all goods are unique and that variety-hungry consumers desire all of the goods in the economy. *Third*, as pointed out by Pines (2001) the NEG explains the presence of urban agglomerations in an anachronistic manner, by relying – to a very large degree – on the trade between urban and agricultural areas, neglecting internal urban structure and land markets.[2]

In this paper, we will remove agriculture from the commonly used NEG setup and we will synthesize this modernized version with a model of a system of identical cities with land markets common in urban economics (Henderson (1988)). Using such a hybrid

[2] Actually, in his early papers, Krugman (1980, 1991) ignored agriculture but his cities without land markets were confined to a two-point limited geography. There are some recent extensions that have introduced urban land markets into Krugman's 1991 model. See, for example, Krugman and Livas-Elizondo (1996) which also deals with a two-point limited geography.

model, we will examine how the size and number of cities – in a system of identically sized cities – should evolve over time to maintain maximum welfare in an economy continually growing in aggregate population. To this end, we make two key assumptions about the number of cities and their spatial arrangement.

1) *The number of cities is endogenous*. This assumption is central to all traditional systems-of-cities models in urban economics (e.g. Henderson and Ioannides (1981)). In contrast, the NEG has emphasized models with a fixed number of sites where cities can emerge.

2) *The spatial arrangement of cities is symmetric* (illustrated in Figure 1). Imagine that there is a circle of arbitrarily large radius with a long enough perimeter to contain on it as large a number of cities as we might be interested in. Suppose also that each city is connected to the circle's center by a spoke, representing a transport route. Goods traded between cities move from the originating city through the corresponding spoke to the center of the circle and then through another spoke to the city of destination. The distance from any city to any other is the same. This symmetric geography is explicit or implicit in virtually all of the systems-of-cities models of urban economics. In the models of Henderson (1988), trading costs among cities are assumed zero and, hence, the same symmetric geography of Figure 1 holds implicitly. In models with positive trading costs between cities (e.g. Abdel-Rahman (1996) or Anas and Kai (2002)) the symmetric geography of Figure 1 is imposed explicitly.

It is necessary to use such symmetry in order to strip the basic NEG setup from the effects of specific (asymmetric) geographies. Without the confounding effect of a specific geography we are able to expose the economic forces of the NEG centered on iceberg trading cost, the taste for variety and constant markup and see whether these are inherently centripetal or centrifugal. If the NEG setup on the symmetric geography creates no agglomeration (a result we will establish in this paper) then the agglomeration that occurs in the NEG setup must be caused by the geographical constraints assumed and not by the economic relationships. This should turn our attention to finding economic forces that create agglomeration when geographic constraints are absent. Subsequently, by imposing the geography one could see what difference the geography itself makes.

[FIGURE 1 ABOUT HERE]

Where t denotes time, let $P(t)$ denote the aggregate exogenous population, $n(t)$ the endogenous optimal number of cities and $N(t)$ the endogenous optimal city size. Then, $P(t)=n(t)N(t)$. With growing $P(t)$:

$$\dot{P}(t)=\dot{n}(t)N(t)+n(t)\dot{N}(t)>0. \tag{1}$$

Table 1 shows the three patterns that are consistent with $\dot{P}(t)>0$. In the first of these patterns which I call *balanced agglomeration* the number of cities does not decrease over time as each city gets bigger or maintains its size. In the second pattern which I call *concentrated agglomeration,* cities get bigger but fewer with the possibility that there

[TABLE 1 ABOUT HERE]

would be one huge city at the limit. In the third pattern which I call *de-agglomeration* or *de-urbanization*, cities get smaller and more numerous with the possibility that there will be infinitesimally small or minimally sized cities at the limit. It is important to make a distinction between *de-agglomeration* and *dispersion*. Dispersion means that economic activity spreads out over geographic space as new agglomerations are created. This process can reduce the size of some agglomerations without totally destroying them. By *de-agglomeration* I am referring not only to spreading out but to the total decline of the agglomeration so that each settlement is minimally sized and completely specialized.

Which pattern of Table 1 is optimal will depend on the specification of a particular theoretical model. A model that generates all three urbanization patterns under different parameter values is better than a model that generates only some. Arguably, the best model should generate each pattern under alternative parameter values. A model that is locked into any one or two of these patterns may be regarded as limited in scope or suspect and we would need to understand why it produces such a limited set of results. We will show that the NEG model is locked into the de-agglomeration pattern, when population increases while other factors are constant and we will see why.

The three patterns can be explained as a tug of war between economic forces causing urbanization and those causing dispersal. For example, growth accommodated in existing cities causes location costs (rent plus transport costs) to rise creating an impetus for smaller and, hence, more cities to keep these costs from increasing too much. Hence, intra-city location costs are a *centrifugal* economic force inducing dispersal of economic activity to smaller cities. If the consumer desires all of the goods in the economy (extreme taste for variety), then if cities are smaller fewer goods are produced locally and more must be imported. The incentive to avoid importation costs acts as a *centripetal* force favoring fewer and larger cities. If this force is strong, there could be just one city at the limit. In that case, since all the goods are locally produced there is no trading cost. But, there is a third force that is centrifugal. If cities are smaller, less time is wasted in commuting and more is available for production. This means that the variety of goods in the city system is larger if cities are smaller. A central planner who is maximizing consumer welfare must balance the three forces at the margin.

We will show that under the special functional forms of the NEG, the welfare maximizing economy with growing population is on a trajectory of increasing numbers of cities and decreasing city sizes (*de-agglomeration*), suddenly collapsing to single-factory towns when aggregate population exceeds a threshold level. Hence, with population growth and without the constraining effects of geography, the NEG implies that there should be no cities eventually. To the extent that NEG models produce agglomeration, the limitations imposed by geography act as a centripetal force in the NEG models, obscuring the workings of the three forces identified in the previous paragraph.

Why should the two centrifugal forces always end up prevailing over the centripetal force in models of the NEG? The key reason for this are the variety-hungry consumers of Dixit-Stiglitz (1977). Such consumers desire all of the goods in the economy and are always hungry for more goods. Population growth makes markets larger and causes more goods to emerge. But the number of goods is largest when they are produced in the smallest cities because that saves time lost in commuting, making more labor available to production. And this centrifugal force is even stronger when interurban trading costs are low.

This last property is the opposite of that in Krugman's core-periphery model without land markets. In that model, trade being between urban areas and agricultural areas, a fall in trading cost would make it optimal to allocate urban producers to one region making a larger market for goods there while cheaply serving farmers in the periphery.[3] Hence, in that model, a fall in trading cost increases the efficiency of concentration. In our model a fall in trading cost makes it optimal to disperse producers into smaller cities, saving intra-

[3] For the core-periphery model of the NEG, see chapter 5 in Fujita, Krugman and Venables (1999).

city commuting and increasing labor supply which in turn increases the varieties available for consumption.

The de-agglomeration property of the NEG applied to a system of cities clashes with basic observations about the history of urbanization in the world economy. Although a model of identical city sizes does not apply exactly to the real world, the number of cities and the average size of cities have greatly increased over the centuries. In particular, the increase in the number of very large cities has become a dominant phenomenon in the last one hundred years or so. A prediction that the historical pattern of increasing urbanization is grossly inefficient cannot be taken seriously. So there are two possible responses to this finding. One response is to continue to defend the NEG model by observing that population growth is not the only exogenous factor and that interurban and intra-urban transport (and communication) costs have been falling due to technical progress such as motorized transportation, railroads, telephone, the Internet. To argue the efficiency of increasing agglomeration trends, one would have to argue that the historical fall in intra-urban personal transport costs has been so great as to offset the parallel and also great fall in interurban freight transportation and communication costs plus the dispersive effect of population growth. Another response is a rejection of the NEG model because it overemphasizes the importance of trade among cities while ignoring efficiencies from search costs or the formation of human capital in cities (Black and Henderson (1999)) as well as other centripetal forces causing city growth.

The paper is organized as follows. In section 2 we present a simple model of a single city in the tradition of urban economics. Then, in section 3 we set up a system of such identical cities each producing differentiated products. The cities are linked with trade among them via the special Dixit-Stiglitz setup of the NEG. Consumers in each city import all of the differentiated goods produced in the rest of the economy. In section 4, we examine how this city system should evolve when its aggregate population grows exogenously and when interurban trading costs or intra-urban commuting costs fall. We prove the basic de-agglomerative property. Section 5 discusses some numerical examples, showing that large numbers of cities before dispersal are optimal only when the taste for variety is extremely low. Section 6 concludes by assessing the alternative explanations for efficient agglomeration.

2. Internal structure of the representative city

We assume monocentric and circular cities. All production occurs at a central point (the Central Business District, or CBD). Producers do not use land. The only input is the labor supplied by the consumers living in the city. Each consumer uses one unit of land (fixed lot size) and is endowed by a unit amount of time that he allocates between labor and commuting to the CBD. The time-cost of commuting a unit distance in both directions (round trip) is s, an exogenous constant. If a consumer picks his residence location to be r miles from the CBD, then his labor supply is $H(r) = 1 - sr$. We will use N to denote the number of consumer/laborers residing in the city and $\bar{r}$ to denote the radius of the city. Then, since lot sizes are uniformly equal to one, $\bar{r} = \sqrt{\frac{N}{\pi}}$. The maximum possible radius for a city is $\frac{1}{s}$, since a consumer residing beyond that radius would spend all of his time commuting and would have no time to work. The maximum population that can be accommodated by the city is therefore $N_{\max} \equiv \frac{\pi}{s^2}$. The aggregate labor, H, supplied to the CBD is

$$H = \int_0^{\bar{r}} 2\pi r H(r) dr = N(1 - kN^{\frac{1}{2}}), \quad (2)$$

where $k \equiv \frac{2s}{3\sqrt{\pi}}$. Since k is a normalization of s, we can use either k or s to measure unit commuting cost.

The principle that determines city structure under the assumption of zero relocation costs is that identical residents achieve the same level of utility no matter where within the city they locate. Indirect utility is of the form $V(\mathbf{p}, I)$, where $\mathbf{p}$ is the price vector of the goods in the economy for the residents of the city and I is the disposable income of any resident that is available to buy those goods. For the value of utility to be invariant with location within the city, the disposable income must be the same for each resident. Disposable income is defined as income less location costs (commuting costs plus the rent of the unit sized lot). Hence, location costs must be invariant with residence location. We assume that rent at the edge of the city is zero, because there is no non-urban use for it. We assume also that time spent commuting is valued at the wage rate, w. Then, commuting cost (to the CBD and back) as a function of residential distance is $CC(r) = wsr$. Hence, the rent on land at radius $0 \le r \le \bar{r}$ is $R(r) = s(\bar{r} - r)w$. The location cost of any one resident is then

$$LC(N) = R(r) + wsr = ws\bar{r} = \frac{wsN^{\frac{1}{2}}}{\sqrt{\pi}}, \quad (3)$$

independent of r. We will assume that a local city government collects all rents and redistributes the average rent to each city resident. The total rent thus shared is

$$TR(N) = \int_0^{\bar{r}} 2\pi r R(r) dr = \frac{wsN^{\frac{3}{2}}}{3\sqrt{\pi}}. \quad (4)$$

Note that both $\frac{TR}{N}$ and LC are increasing functions of the city's population, N. However, LC rises three times faster than $\frac{TR}{N}$. The disposable income of any consumer can now be calculated as

$$I(N) = w + \frac{TR(N)}{N} - LC(N) = (1 - kN^{\frac{1}{2}})w \, . \tag{5}$$

Note that as N grows the disposable income falls because the marginal (and average) resident spends more time commuting in a larger city. The consumer has this income to purchase all the goods that will be offered in the city system. We now turn to how cities are linked through trade of the NEG type.

3. A system of cities based on the NEG

We will consider only the case of symmetric monocentric cities, each organized internally according to the description of the previous section. Suppose that there are n such cities. Let each city produce m unique goods in its CBD, each produced by a different firm.[4] Although n and m should each be integer, to simplify the analysis we will treat both as continuous and differentiable variables: $1 \le m, n < \infty$. The lower limit helps fix the notion that there must be at least one city and at least one good in the economy.

Let x_{ji} be the quantity of good j produced in city i and consumed by a consumer in any representative city. Let the utility function of such a consumer be Dixit-Stiglitz C.E.S. (1977):

$$U = \left[\sum_{i=1}^{n} \sum_{j=1}^{m} x_{ji}^{\frac{\sigma-1}{\sigma}} \right]^{\frac{\sigma}{\sigma-1}} l \tag{6}$$

[4] Whether these goods are variants (brands) in the same industry or belong to different industries is immaterial.

where $\sigma > 1$ is the elasticity of substitution between any two goods and l, lot size, is set to one by assumption, as was explained in section 2. *U* exhibits an extreme taste for variety: as the quantity of a good approaches zero, its marginal utility shoots out to infinity. Thus, a consumer prefers to consume an infinitesimal quantity of any available good (even if its price is extremely high) than do without it. This strong taste for variety disappears as $\sigma \to \infty$. As that extreme is approached, the goods begin to look like perfect substitutes (linear indifference curves) and variety no longer matters to the consumer: he would always consume only the cheapest good.

In symmetric equilibrium, we will use *q* to denote the mill price of any good. As in the NEG, intra-urban transport costs for goods are assumed to be zero. Hence, locally produced goods can be consumed without incurring any transport costs. Interurban transport costs for goods are assumed paid by the consumers and will be of the iceberg type with $0 \leq \tau \leq 1$, the fraction of the good shipped from the origin city's CBD that arrives at the destination city's CBD, to be purchased there by a consumer. Then, $\tau = 1$ corresponds to the case of zero transport costs while $\tau = 0$ is the case of infinitely high transport costs under which trade will be prohibitively costly. Suppose, for example, that $\tau = 1/2$. Then, that means that each consumer must order twice the demanded quantity because half of the quantity will melt in transport. With these assumptions, the effective unit price of a quantity imported from another city will be $\frac{q}{\tau} > q$ while the unit price of a good produced in one's own city will be *q*. Because of the symmetry we assumed in the spatial arrangement of cities (Figure 1), a consumer located in city *i* will want to consume x_i units of each good produced in city *i* and x_{-i} units of each good produced in any other city. Then, the utility of a representative resident in city i can be rewritten as follows by imposing the symmetry on (6):

$$U_i = \left[m x_i^{\frac{\sigma-1}{\sigma}} + (n-1) m x_{-i}^{\frac{\sigma-1}{\sigma}} \right]^{\frac{\sigma}{\sigma-1}} \qquad (6')$$

The budget constraint is: $m x_i q + (n-1) m x_{-i} \frac{q}{\tau} = I(N)$, where $I(N) = (1 - kN^{\frac{1}{2}})w$, from section 2, is the disposable income of a consumer after payment of location costs. From (6′), the indirect utility function of any consumer in the city system under symmetry is:

$$V = \left[m q^{1-\sigma} + (n-1) m \left(\frac{q}{\tau} \right)^{1-\sigma} \right]^{\frac{1}{\sigma-1}} I(N). \qquad (7)$$

To determine the number of product varieties, m, we will employ the assumption of the NEG models that each product is manufactured using labor only: a fixed quantity of labor, f, is required for production to start and each unit produced requires a marginal quantity, c. The manufacturers are monopolistically competitive. So, the ensuing equilibrium is that of Chamberlin (1933). As is well-known, the profit maximizing price of each differentiated product is set to satisfy a markup over marginal cost and each producer makes zero profit because of unrestricted entry by manufacturers. The markup condition is $q(1 - \frac{1}{E}) = wc$ where E is the price elasticity of the consumers' demand for any good. With the utility function (1), this price elasticity – as is well known –is approximately σ [see Dixit-Stiglitz (1977)]. Hence, the markup condition gives

$q = \frac{\sigma}{\sigma - 1} cw$ where w is the wage and wc the marginal cost.[5] Using this in the zero profit condition $qz_s - w(f + cz_s) = 0$, where z_s is the firm's output, we get $z_S = \frac{f(\sigma - 1)}{c}$. Now to determine how many varieties a city will produce, we need to look at the labor market clearing condition for a city. We saw in equation (2) that the labor supplied by a city of N consumers is $H = N(1 - kN^{\frac{1}{2}})$. The aggregate labor demanded by the city's m producers, meanwhile, is $m(f + cz_s)$ and using the z_s derived above, this is $mf\sigma$. Setting the labor demand and supply to be equal and solving, $m = \frac{N(1 - kN^{\frac{1}{2}})}{f\sigma}$.[6]

Earlier, we restricted the number of goods to be produced by a city to be 1 or larger. It is useful to derive the size of a minimal city (a single factory town) producing only one good. Call this size $N_{\min}$. Since the labor required to produce a single good is $f\sigma$, the labor market in a minimally sized city clears by $H_{\min} \equiv N_{\min}(1 - kN_{\min}^{\frac{1}{2}}) = f\sigma$. The unique solution is of the form $N_{\min} = F(f\sigma, k)$, with F increasing in $f\sigma$ and increasing in k. (As k, unit commuting cost, increases less labor is supplied by each resident on average. To meet the aggregate labor supply, $f\sigma$, more residents must be added to the factory town. More precisely, $\frac{dN_{\min}}{dk} = \frac{N_{\min}^{3/2}}{1 - (3/2)kN_{\min}^{1/2}} > 0.$ The denominator is positive since $N_{\min} < N_{\max} \equiv 4/9k^2$.) We have thus established a range for city sizes ($N_{\min} \le N \le N_{\max}$). The maximal city size $N_{\max} = \frac{\pi}{s^2} = \frac{4}{9k^2}$ is limited only by k, as we saw

---

[5] Note that the markup is independent of the number of firms sharing the market, a direct consequence of the approximation that $E \approx \sigma$. This is a standard property of all NEG models.

[6] Because of symmetry, there are effectively two markets: the labor market and the market for products. Walras's Law implies that if the labor market clears (as shown earlier) the product markets must clear as well.

in section 2, while the minimal city size, as we just saw, is limited by *f*, the economies of scale internal to a firm and by *k*. By choice of *f* we insure $N_{\min} = F(f\sigma, k) < N_{\max} = \frac{4}{9k^2}$.

Now plug $m = \frac{N(1 - kN^{\frac{1}{2}})}{f\sigma}$ and $I(N) = (1 - kN^{\frac{1}{2}})w$ into the indirect utility function (7) and set the wage rate *w* as the numeraire price, *w* = 1. After some simple algebra, the indirect utility (7) becomes:

$$V(n, N) = \left(\frac{\sigma - 1}{\sigma c}\right)\left(\frac{1}{f\sigma}\right)^{\frac{1}{\sigma-1}} \left[1 + (n-1)\tau^{\sigma-1}\right]^{\frac{1}{\sigma-1}} N^{\frac{1}{\sigma-1}} (1 - kN^{\frac{1}{2}})^{\frac{\sigma}{\sigma-1}} \quad . \tag{7'}$$

This expression serves as the measure of welfare to be maximized by the central planner with respect to *n* and *N*, taking *P*=*nN* as a constraint with *P* given. We have assumed that the planner treats equals equally by keeping consumers in different cities at the same level of utility. The planner also takes as given the land market equilibrium within each city and takes as given the markup of each producer.

4. The optimal pattern of cities

We will examine two types of optima. In the first, the local planner of a city determines the population of a city so as to parochially maximize (7′) the utility level of the representative resident of that city. In doing so, the local government does not consider that it might alter the population available to other cities. This assumption is plausible if, in fact, there are so many cities that each city accommodates only a small part of the aggregate population. Solving this problem determines what I will call the *autarkic-efficient size of a city*. This will act as a benchmark in the analysis to follow. To get this size, maximize *V* given by (7′) with respect to *N*, keeping *n*, the number of cities, constant. The result is

$N^* = \frac{9\omega^2}{9\omega^2 + 6\omega + 1} N_{\max}$, where $\omega \equiv \frac{1}{\sigma - 1}$ and $N_{\max} = \frac{4}{9k^2} = \frac{\pi}{s^2}$ is the maximum population that can be accommodated in a city as derived in section 2. The autarkic-efficient city size, N^*, depends only on σ and k (or s). Note also that it is a fraction of the maximum city size. Such an autarkic-efficient city size with $m > 1$ exists as long as $N_{\min} = F(f\sigma, k) < N^*$, easily guaranteed by the choice of a sufficiently low f, the fixed labor requirement at the level of the firm.

What fraction of maximum size is the autarkic-efficient size? Suppose that $\sigma = 10$ ($\omega = 0.11$), then the autarkic efficient size is about 6.2% of the maximum size. If $\sigma = 5$ ($\omega = 0.25$), then the autarkic size is about 18.37% of the maximum. If $\sigma = 2$ ($\omega = 1$), then the autarkic size is 56.25% of the maximum size. As these examples show, autarkic-efficient size decreases as σ increases. A higher σ means that goods variety is valued less by the consumer (the goods are more like perfect substitutes). Hence, cities can afford maximum utility at smaller sizes saving on commutation time but offering fewer local goods.

In the second type of optimum that is of direct interest to us in this paper, we assume the presence of a central planner who takes the aggregate population as given and determines the number of cities so that the utility of the representative consumer is maximized. To get the appropriate objective function, we must recognize that given aggregate population, P, the number of cities and the size of a city are related by $n = \frac{P}{N}$. Plugging this into (7′), dropping the leading constant, and making some simplifications we get:

$$\widetilde{V}(P,N) \equiv V(\frac{P}{N}, N) = \left[N(1 - \tau^{\frac{1}{\omega}}) + P\tau^{\frac{1}{\omega}} \right]^{\omega} (1 - kN^{\frac{1}{2}})^{\omega+1}. \tag{7''}$$

To determine the optimal city size and, at the same time, the number of such optimally sized cities, the planner will maximize (7′′) with respect to N in $N_{\min} \le N \le N_{\max}$ and we assume that P is sufficiently large so that there will be at least several (or many) cities at the optimum. The result is a socially optimal first-best partition of the population into n cities. The allocation is first-best, despite the fact that the planner is taking the monopolistic markup as given. This is so because, in contrast to Dixit-Stiglitz (1977), there is no other good toward which the allocation can be distorted as a result of the markup. As derived earlier, the amount of labor is $H = N(1 - kN^{\frac{1}{2}})$, and the number of varieties produced is $m = \dfrac{N(1-kN^{\frac{1}{2}})}{f\sigma}$ and both are independent of the markup.

<u>Proposition 1</u> ***(De-agglomeration with population growth)*****:** *Assume that $\sigma \ge 2$ and assume a small enough f, so that $N_{\min}$ is very close to zero and very far from $\frac{1}{4}N^*$. Keeping constant the values of σ $(\omega = \frac{1}{\sigma - 1})$, s $(k = \frac{2s}{3\sqrt{\pi}})$ and $0 < \tau < 1$, suppose that aggregate population, P, starts at some initial value P_0 and grows continuously and monotonically. Optimally-sized cities start at some size $\frac{1}{4}N^* < N_0 < N^*$ and decline toward $\frac{1}{4}N^*$ as $P \to \widetilde{P} \equiv \left(\dfrac{1-\tau^{\frac{1}{\omega}}}{\tau^{\frac{1}{\omega}}}\right)\dfrac{\omega^2}{(3\omega^2+4\omega+1)}\dfrac{1}{k^2} = \left(\dfrac{1-\tau^{\frac{1}{\omega}}}{\tau^{\frac{1}{\omega}}}\right)\dfrac{9\omega^2}{12\omega^2+16\omega+4}N_{\max}$.*

Then, at some $P = \hat{P} < \widetilde{P}$, when cities are of size $\hat{N}$, city sizes will suddenly drop to $N_{\min}$, remaining there for all larger P.

<u>Proof</u>: See Appendix. Refer to Figure 2 for an illustration of the claim of the Proposition.

[FIGURE 2 ABOUT HERE]

While the Appendix proves the claimed result for (7′), derived from the functional forms of the NEG (C.E.S./iceberg/constant markup), the result holds more generally for any reduced form indirect utility function in which the effects of n and N are separable. Thus let $V(n,N)=V_1(n)V_2(N)$ where $n=P/N$. Differentiating this with respect to N we get

$$\frac{dV(n,N)}{dN}=\frac{\partial V_1}{\partial n}\left(-\frac{P}{N^2}\right)+\frac{\partial V_2}{\partial N}. \tag{8}$$

At the optimal partition of *P* among cities, (8) must be zero. $\frac{\partial V_1}{\partial n}>0$ reflecting that utility must increase with *n* since that would increase the number of varieties. But, since *P* is constant, a higher n makes each city smaller. Hence the term $\frac{\partial n}{\partial N}=-P/N^2$. Therefore, the first term in (8) measures the negative marginal effect on utility from increasing city size, thus reducing aggregate variety. The second term is the marginal effect on city size keeping *n* constant. As we saw earlier, $V_2(N)$ is inverse U-shaped increasing as *N* rises toward the autarkically efficient size N^* and decreasing for $N>N^*$. Since, $\frac{\partial V_1}{\partial n}>0,$ (8) can be zero only if $\frac{\partial V_2}{\partial N}>0.$ Hence, the socially optimal city size must be smaller than N^* as claimed in Proposition 1. In the case of the NEG, $\frac{\partial V_1}{\partial n}=\omega\tau^{\frac{1}{\omega}}\left[1+(\frac{P}{N}-1)\tau^{\frac{1}{\omega}}\right]^{\omega-1}>0$ for any $\tau>0.$ But when $\tau=0$, then $\frac{\partial V_1}{\partial n}=0$ and in that case the socially optimal city size is N^*.

[FIGURE 3 ABOUT HERE]

Figure 3 shows the path of city sizes as a function of aggregate population. It also includes the cases of infinite ($\tau=0$) and zero ($\tau=1$) trading costs. In the former case, trade is prohibitively costly. Hence, for all *P,* cities do not trade and welfare is maximized by autarkic-efficient city sizes. As aggregate population grows, the optimal economy creates more and more autarkically-sized self sufficient cities. This is an example of a balanced agglomeration path on which identical cities are replicated ad-infinitum, a property of city systems without trade between cities (as in Henderson and Ioannides (1981)). With zero trading costs, more than minimal cities are not ever optimal because any larger concentration creates unnecessary location costs.

Proposition 1 means that the special model of the NEG when synthesized with a model of city size, always produces de-agglomeration and complete dispersal of economic activity when population grows sufficiently large. The alternative patterns of balanced or concentrated agglomeration (identified in Table 1) do not happen. Why does the NEG model produce such a narrow result? To gain more insight into that question, we will rewrite the utility function (7) as follows:

$$V(P,N) = q^{-1}\left[m + \tau^{\sigma-1}(P-N)\frac{m}{N}\right]^{\frac{1}{\sigma-1}} I(N)$$

$$= q^{-1}\left[m(N) + \tau^{\sigma-1}M(P,N)\right]^{\frac{1}{\sigma-1}} I(N) \qquad \text{(9a)}$$

where $m(N) \equiv \dfrac{N(1-kN^{\frac{1}{2}})}{f\sigma}$ and $M(P,N) \equiv (P-N)\dfrac{m}{N} = \dfrac{(P-N)(1-kN^{\frac{1}{2}})}{f\sigma}$ are the locally available varieties (previously derived in section 3) and the varieties imported from other cities, respectively. Taking logs of (9a):

$$\ln V(P,N) = -\ln q + \frac{1}{\sigma-1}\ln\left[m(N) + \tau^{\sigma-1}M(P,N)\right] + \ln I(N)\,. \qquad \text{(9b)}$$

This last equation directly shows how city size affects welfare through *three* effects. The first of these is the *local variety* (or *home market*) *effect* measured by $m(N)$. Note that as N increases, the number of local varieties increases because the derivative of $m(N)$ with respect to N is always nonnegative in the range $N_{\min} \leq N < N_{\max}$. This effect means that adding more people to a city increases the labor supply and the local demand for varieties. This causes new varieties to be supplied locally because the local market gets bigger. The second effect is the *imported varieties effect*, measured by $\tau^{\sigma-1}M(P,N)$. Note that this effect does not exist when $\tau = 0$ (infinite trading cost), since nothing is imported in that case. Note also that in the case of $\tau = 1$ (zero trading cost), there is no difference between local and imported varieties since the latter can be imported at zero cost. In that case, the two effects in (9b) add up to the total number of varieties $m(N) + M(P,N)$. Finally, note that as $\sigma \rightarrow \infty$ the combined variety effects have no weight because consumers view all goods as perfect substitutes.

The imported varieties effect is a negative one for city size: making cities bigger while keeping P constant (i.e. reducing the number of cities) reduces the number of imported varieties. The reason is that as cities get bigger average location costs in cities go up and less time is left for production. Hence, as P increases, the aggregate labor supply increases less than it would if there were more and smaller cities with lower commuting costs. This property favors dispersal as P increases because with dispersal there would be more labor supply and more imported goods. To see this directly, note that $\dfrac{\partial M(P,N)}{\partial P} = \dfrac{1-kN^{\frac{1}{2}}}{f\sigma} > 0$. Finally, the third effect is the *location cost effect*, measured by $I(N) = (1-kN^{\frac{1}{2}})$. This effect means that when cities get larger disposable income decreases because, as we saw in section 2, commuting and rent costs build up with city size. This effect also favors dispersal. How city size and the model's parameters affect each effect is shown in Table 2.

[TABLE 2 ABOUT HERE]

Since the optimal city size decreases with increasing population, *P*, it must be true that the *local variety effect* that favors larger cities is offset by the *imported variety* and *location cost effects* that individually favor smaller cities.

The next two propositions are proved in a way similar to Proposition 1 and show how de-agglomeration happens when the unit iceberg trading cost falls or unit commute cost rises through their corresponding threshold values.

Proposition 2: ***(De-agglomeration with declining trading cost)***: *Assume that $\sigma \geq 2$ and assume a small enough f, so that $N_{\min}$ is very close to zero and very far from $\frac{1}{4}N^*$. Keeping constant the values of σ $(\omega = \frac{1}{\sigma - 1})$, s $(k = \frac{2s}{3\sqrt{\pi}})$ and P, suppose that τ starts at some initial value τ_0 and increases continuously and monotonically toward $\tilde{\tau}$. Optimally-sized cities will start at some size $\frac{1}{4}N^* < N_0 < N^*$* and will *decline gradually toward $\frac{1}{4}N^*$. Then, at some $\tau = \hat{\tau} < \tilde{\tau} \equiv \left(\frac{\omega^2}{\omega^2 + (1 + 4\omega + 3\omega^2)k^2 P} \right)^{\omega}$, when cities are of size $\hat{N} > \frac{1}{4}N^*$, city sizes will suddenly drop to $N_{\min}$, remaining there for larger τ.*

Proof: It follows that of Proposition 1. ■

Why does falling trading cost encourage decentralization and dispersal? Recall our discussion of the three effects after Proposition 1. From Table 2, a lowering of trading costs strengthens the *imported varieties effect* while it has no direct effect on the *local varieties* and *location costs effects.* This favors smaller cities because for the same aggregate population a system of smaller cities affords a larger labor supply and can make more varieties than can a system of larger cities. Note that this result is the opposite of that in Krugman's core-periphery model. Although his analysis in that model is equilibrium rather than optimum analysis, the basic insight is that lower interurban trading costs make it cheaper for agricultural consumers in the periphery to import goods from the urban core. At the optimum, locating all manufacturers in the core would capture home market efficiencies. Because Krugman's core-periphery model in its standard version does not include location costs (cities are points without land) there is nothing to offset the home market efficiency gains. By contrast, in our setting, if firms were to concentrate in cities in response to lower interurban trading costs, location costs would go up and the aggregate variety of goods in the city system would fall because more time would be wasted commuting.

Proposition 3: ***(De-agglomeration with increasing commuting cost)***: *Assume a small enough f and small enough k, so that* $N_{\min} = F(f\sigma, k)$, *is very close to zero and very far from* $\frac{1}{4}N^*$. *Keeping constant the values of* σ $(\omega = \frac{1}{\sigma - 1})$, $0 < \tau < 1$, *and P, suppose that k starts at some initial value of commuting cost* k_0 *and increases continuously and monotonically toward* $\tilde{k} \equiv \left(\frac{\omega^2 (1 - \tau^{\frac{1}{\omega}})}{\tau^{\frac{1}{\omega}} (1 + 4\omega + 3\omega^2) P} \right)^{\frac{1}{2}}$. *Optimally-sized cities will start at some size* $\frac{1}{4}N^* < N_0 < N^*$ *and will decline gradually toward* $\frac{1}{4}N^*$. *Then, at some k* $= \hat{k} < \tilde{k}$ *when cities are of size* $\hat{N} > \frac{1}{4}N^*$, *city sizes will suddenly drop to* $N_{\min} = F(f\sigma, \hat{k})$, *and will eventually settle there for all larger k.*

Proof: The procedure follows that of Proposition 1. In this case, the path of the optimization being monotonic cannot be guaranteed by requiring $\sigma \geq 2$. ■

Obviously, increasing unit commuting cost strengthens the *location cost effect* which is negative for larger city size. But a higher *k* indirectly also weakens the variety effects because more time is used to commute and less is left to produce. This favors dispersal into smaller cities so that commuting time can be saved and the number of varieties increased. It should be noted that if $k = 0$ and $\tau < 1$ maximal city size is infinite and there is only one city containing all population, since that is the largest home market that can be formed and no commuting cost is incurred. If $k > 0$ and $\tau = 1$, cities are minimally sized because that maximizes the number of varieties. Finally, when $k = 0$ and $\tau = 1$, city sizes are indeterminate.

5. Numerical examples

It is worthwhile to work out a few numerical examples. The question we want to pose is related to Proposition 1. We have shown that the taste for variety that causes intensive trading of goods among cities is responsible for the result that maintaining cities (urbanization) becomes inefficient when aggregate population is high. Turning this observation around, we can ask: Just how low should the taste for variety be for cities to persist at large aggregate population? Maximum city size depends only on unit commuting cost and serves as a benchmark. Suppose that maximum city size is 100 million workers. Interurban transport costs in reality are not very high. Let us assume that $\tau = 0.9$ which means that about 11% more should be shipped than is desired for consumption. We calculate $\widetilde{P}$ and N^* for various values of σ, the elasticity of substitution among varieties. As noted earlier, the higher the value of σ is the closer the goods are to perfect substitutes and, hence, the lower is the importance of variety. A high σ delays the population level at which cities should disappear to maintain efficiency. Consider the results for $\sigma = 20$, 50 and 100. These values are way out of the range of the σ values employed in Fujita, Krugman and Venables (1999), but as we shall see at the low σ values that they employ, the taste for variety is so strong that an economy of cities falls apart even at very low levels of aggregate population. Even when σ=20, autarkic-optimal city size is only 2.4 million workers and $\widetilde{P}$ = 4.29 million. Recall that cities should break up when they are smaller than autarkic-efficient but larger than one-quarter the size of autarkic efficient. This means that at the critical point when the economy of cities suddenly disintegrates into stand-alone factory towns there will be no more than about 7 cities in existence, each with about 600,000 workers. To get more cities just before the critical point is reached we have to go to higher values of σ. When σ= 50, autarkic-optimal city size is 320,000 workers and $\widetilde{P}$ = 14.45 million so there are no more than 180 cities just before they fall apart into single-firm factory towns. And when σ= 100, autarkic-optimal city size is about 92,000 workers and $\widetilde{P}$ = 747,650,000. That gives an upper bound of about 32,500 pretty small cities just before complete de-urbanization becomes optimal. The lower the taste for variety, the higher the aggregate population required to make complete de-urbanization optimal. A high taste for variety is harmful to the long run efficiency of urbanization. And a low taste for variety implies longer lived but smaller cities. Of course, if all goods were perfect substitutes, only local goods would be consumed and there would be no trade. Cities would once again be perfectly specialized, minimizing commuting costs.

6. Concluding remarks

We showed that the special assumptions of the NEG entail a strong imported varieties effect, perhaps too strong. When stripped from the obscuring influence of a specific geography, the dispersive nature of this effect becomes clear. This strong effect causes optimal city sizes to decline and then crash as aggregate population grows. We called this process *de-agglomeration* because it signifies not only *dispersion* meaning the spreading out of economic activity over space, but it also signifies *de-urbanization* which means smaller cities and more specialized cities leading ultimately to single-factory towns.

The NEG's imported varieties effect stems from the extreme taste for variety of the Dixit-Stiglitz type. This implies that a consumer will desire all of the goods produced in the economy. That is grossly unrealistic. In reality most goods are produced in many cities and therefore many goods produced in separate cities are viewed as perfect substitutes. Trade among cities is not as indispensable as the standard NEG setup suggests. The strong emphasis on trading costs may be explained by the fact that the NEG was originated by a trade theorist (Paul Krugman). However, Krugman's collaboration with an urban economist (Masahisa Fujita) has not produced an appropriate modification of the extreme emphasis on trading costs in the NEG.

Our results also suggest that the historical tendency toward increased urban agglomerations despite great population growth may be explained by the fact that while big population increases have taken place, intercity trading and intracity commuting costs have fallen greatly also. Although this is an interesting argument, it may not rescue the NEG model. As we saw, falling intercity trading costs actually help de-agglomeration, contrary to what occurs in Krugman's agriculture based core-periphery model. Can improvements in commuting costs alone account for the increasing urban agglomeration trends? Perhaps they can. But as technical progress in intraurban transportation and road-building reach their limits, something else must drive urbanization.

To reverse the NEG result that de-agglomeration is inevitable, future models will need to emphasize other centripetal forces that can offset the dispersive forces of the NEG. There are several candidates. 1) Haddar and Pines (2002), in a model that accommodates a maximum of two locations, have shown that population will concentrate in one location if the demand for land by consumers is price and income elastic and land and product varieties are substitutes. It is not clear that this property survives to our case of an endogenously determined number of cities, but it is certainly true that the density adjustments in large cities mitigate the centrifugal effect of commuting cost, thus favoring larger cities. 2) It has been suggested that larger cities help reduce the cost of matching up firms and workers and improve the quality of such matches (Helsley and Strange (1990)). 3) Although a good theory is lacking, there is some empirical evidence that higher land use densities make workers more productive (Moomaw (1981), Ciccone and Hall (1996)). 4) Again although a good theory is lacking, there is tentative evidence that large cities are more diverse in industrial mix and that such diversity is associated with more growth (Glaeser et.al. (1992)). Any one of these ideas, and undoubtedly some others, could be used to strengthen the NEG's ability to generate agglomerations that are sustainable under population growth. Otherwise, the implications of the NEG are that urbanization is the result of falling commuting cost or of the limits of geography itself which prevents economies from rapidly spawning new cities.

References

Abdel-Rahman, H. (1996). When do cities specialize in production?, *Regional Science and Urban Economics*, 26, 1-22.

Anas, A. and K. Xiong, 2002, Intercity trade and the industrial diversification of cities, *Journal of Urban Economics*, in press.

Black, D. and Henderson, J.V. (1999) "A theory of urban growth", *Journal of Political Economy*, 107(2), 252-284.

Chamberlin, E. H. (1933) *The theory of monopolistic competition*, Cambridge, MA: Harvard University Press.

Ciccone, A. and R. Hall (1996), Productivity and the density of economic activity, *American Economic Review*, 86 (1) 54-70

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

Fujita, M. , P. Krugman and T. Mori (1999) On the evolution of hierarchical urban systems, *European Economic Review*, 43, 201-259.

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

Glaeser, E.L., H.Kallal, J.A. Scheinkman, A. Schleifer (1992) Growth in cities, *Journal of Political Economy*, 100(6), 1126-52.

Haddar, Y. and D. Pines (2002) Population growth and its distribution between cities: positive and normative aspects. Unpublished manuscript.

Helsley, R.W. and W.C. Strange, 1990, Matching and agglomeration economies in a system of cities, *Regional Science and Urban Economics,* 20(2), 189-212.

Henderson, J.V. (1988) *Urban Development* , Oxford: Oxford University Press.

Henderson, J.V. and Y. Ioannides (1981). Aspects of growth in a system of cities, *Journal of Urban Economics*, 10, 117-139.

Krugman, P.R. (1980) Scale economies, product differentiation, and the pattern of trade, *American Economic Review* 70, 950-959.

Krugman, P. (1991) Increasing returns and economic geography, *Journal of Political Economy*, 99, 483-499.

Krugman, P. and R. Livas-Elizondo (1996) Trade policy and the Third World metropolis, *Journal of Development Economics* 49 (1), 137-150.

Moomaw, R. (1981) Productivity and city size: A critique of evidence, *Quarterly Journal of Economics*, 96(4), 675-688.

Pines, D. (2001) New Economic Geography: Revolution or counter-revolution? *Journal of Economic Geography*, 139-146.

Appendix: Proof of Proposition 1

Proof: We need to examine corner as well as interior extremes in $N_{\min} \le N \le N_{\max}$. Since we make sure (by choosing f) that $N_{\min}$ (the population requirement of a single firm) is very close to zero, we will just evaluate $\widetilde{V}(P,N)$ at $N = 0$. From (7″) we can see that $\widetilde{V}(P,0) = P^{\omega}\tau$ and $\widetilde{V}(P,N_{\max}) = \left[\frac{\pi}{s^2}(1-\tau^{\frac{1}{\omega}}) + P\tau^{\frac{1}{\omega}}\right]^{\omega}\left(\frac{1}{3}\right)^{\omega+1}$.

The first order condition necessary for an interior maximum in this range is:

$$\frac{\partial\widetilde{V}(P,N)}{\partial N} = \frac{1}{2}\widetilde{V}(P,N)\left[\frac{2\omega(1-\tau^{\frac{1}{\omega}})}{N(1-\tau^{\frac{1}{\omega}}) + P\tau^{\frac{1}{\omega}}} - \frac{(1+\omega)k}{N^{\frac{1}{2}} - kN}\right] = 0. \qquad \text{(A.1)}$$

We have $\widetilde{V}(P,N) > 0$ for all $N_{\min} \le N \le N_{\max}$ and $P > 0$. Hence, at an interior extreme point, (A.1) vanishes by the bracket being zero. Cross-multiplying the expression inside the bracket, we obtain a second degree polynomial $AX^2 + BX + C = 0$, where $X \equiv N^{\frac{1}{2}}$, $A = -k(1-\tau^{\frac{1}{\omega}})(1+3\omega)$, $B = 2\omega(1-\tau^{\frac{1}{\omega}})$ and $C = -k\tau^{\frac{1}{\omega}}(1+\omega)P$. Applying the quadratic formula and simplifying the resulting expression, the two roots are:

$$N_a(P) = \left(\frac{\omega - \sqrt{\omega^2 - \frac{\tau^{\frac{1}{\omega}}}{1-\tau^{\frac{1}{\omega}}}(1+\omega)(1+3\omega)k^2 P}}{k(1+3\omega)}\right)^2 \qquad \text{(A.2a)}$$

$$N_b(P) = \left(\frac{\omega + \sqrt{\omega^2 - \frac{\tau^{\frac{1}{\omega}}}{1-\tau^{\frac{1}{\omega}}}(1+\omega)(1+3\omega)k^2 P}}{k(1+3\omega)}\right)^2. \qquad \text{(A.2b)}$$

Note that $N_a(0) = 0$ and $N_b(0) = N^*$. For an interior solution to exist the N given by (A.2a) and (A.2b) must be real. This requires that the expression under the square root sign be non-negative. This amounts to requiring that aggregate population not be above a critical threshold. From (A.2a) or (A.2b) that threshold is $\widetilde{P}$, given in the statement of the Proposition.

For any $P \le \widetilde{P}$, $N_a(P) \le N_b(P)$ with the square-root expression taken as nonnegative. It is easy to verify that $\frac{\partial N_a}{\partial P} > 0$ and $\frac{\partial N_b}{\partial P} < 0$ for all $P < \widetilde{P}$. Next, we need to determine which of the roots is the interior local maximum. From (A.1), we can show that

$$\lim_{N\to 0}\frac{d\widetilde{V}(P,N)}{dN}=-\infty \qquad \text{(A.3a)}$$

for all $P > 0$ and

$$\frac{d\widetilde{V}(P,N)}{dN}\Big|_{N=N_{\max}}=-\widetilde{V}(P,\frac{\pi}{s^2})\left[\frac{(1-\tau^{\frac{1}{\omega}})+(1+\omega)n\tau^{\frac{1}{\omega}}}{(\pi/s^2)(1-\tau^{\frac{1}{\omega}})+P\tau^{\frac{1}{\omega}}}\right]<0. \qquad \text{(A.3b)}$$

Hence, the $\widetilde{V}(P,N)$ curve has a local minimum at $N_a(P)$ and a local maximum at $N_b(P)$ in the range $N_{\min} < N < N_{\max}$ for $P < \widetilde{P}$. As $P \to \widetilde{P}$, the curve $\widetilde{V}(P,N)$ shifts up as shown in Figure 2, and when $P = \widetilde{P}$, the local minimum and maximum have converged to an inflection point. (In Figure 2 follow $P_0 \to P_1 \to P_2 \to \hat{P} \to \widetilde{P}$ and the corresponding utility curves.) The fact that the two roots converge to the same value as P reaches $\widetilde{P}$ can be verified directly from (A.2a) and (A.2b) by evaluating these at $P = \widetilde{P}$. The square root expression is zero and the unique N is $N(\widetilde{P})=\frac{1}{4}N^*=\frac{9\omega^2}{36\omega^2+24\omega+4}\left(\frac{\pi}{s^2}\right)$, one fourth of the autarkic city size. Once $P > \widetilde{P}$, the curve $\widetilde{V}(P,N)$ is negatively sloped throughout indicating a corner maximum at $N_{\min}$.

From Figure 2, the locus of optima as P increases starts at some $N_b(P_1) < N^*$ and follows the path of the interior local maximum. During this phase $\widetilde{V}(P,N_b(P)) > \widetilde{V}(P,N_{\min})$. Then, before P reaches $\widetilde{P}$, at the first instance that $\widetilde{V}(\hat{P},N_b(\hat{P})) = \widetilde{V}(\hat{P},N_{\min})$ for some $\hat{P} < \widetilde{P}$, city sizes suddenly crash to the minimum city size forming single-firm towns and the number of settlements jumping suddenly from $\frac{\hat{P}}{N_b(\hat{P})}$ to $\frac{\hat{P}}{N_{\min}}$. The maximized value of utility, meanwhile, makes no jumps through this process.

It will be important that the curve $\widetilde{V}(P,N)$ shift up in such a way that $\widetilde{V}(P,N_{\min})$ shift up faster than $\widetilde{V}(P,N_b(P))$. That will insure that when the maximum switches from the interior to the corner solution at $P = \hat{P}$, it will thereafter remain at the corner solution and not switch back to an interior maximum position again. Fortunately, this is easy to guarantee by assuming that $\sigma \geq 2$. To see this, note that

$$\frac{\partial\widetilde{V}(P,N)}{\partial P}=\omega\tau^{\frac{1}{\omega}}\frac{(1-kN^{\frac{1}{2}})^{1+\omega}}{\left[N(1-\tau^{\frac{1}{\omega}})+P\tau^{\frac{1}{\omega}}\right]^{1-\omega}}>0. \qquad \text{(A.4)}$$

The value of this derivative will decrease as N increases as long as $1-\omega \geq 0$ or, equivalently, $\sigma \geq 2$. If this condition does not hold, the optimum may vacillate between an interior and a corner position but it will eventually settle down at the corner solution as

$P \to \widetilde{P}$ and will remain at the corner solution for higher P since there are no interior local maxima for $P > \widetilde{P}$.■

Pattern	**Aggregate population growth,** $\dot{P}(t)$	**Growth in the number of cities,** $\dot{n}(t)$	**Growth in city population,** $\dot{N}(t)$	**Possible limiting state**
Balanced agglomeration	>0	≥ 0	≥ 0	*City sizes or the number of cities approach limits*
Concentrated agglomeration	>0	< 0	> 0	*One giant city*
De-agglomeration	>0	> 0	< 0	*Minimally sized cities*

TABLE 1: Possible patterns for the evolution of a system of identical cities under aggregate population growth.

Effect		N	P	f	k	τ	σ
Local varieties (centripetal)	$m(N)$	+	0	–	–	0	–
Imported varieties (centrifugal)	$\tau^{\sigma-1}M(N)$	–	+	–	–	+	–
Location cost (centrifugal)	$1-kN^{\frac{1}{2}}$	–	0	0	–	0	0

TABLE 2: The centripetal and centrifugal effects of the NEG and how they are affected by the parameters of the model.

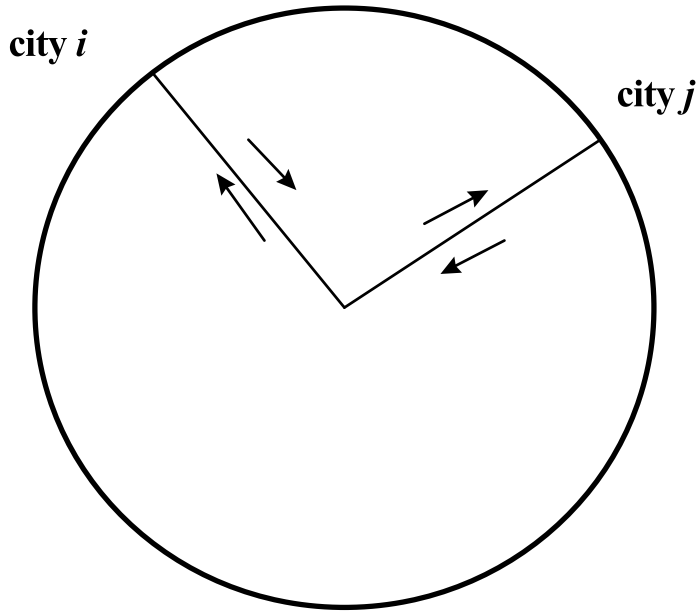

FIGURE 1: Trading routes in the symmetric system of cities located on the peripehry of a large circle

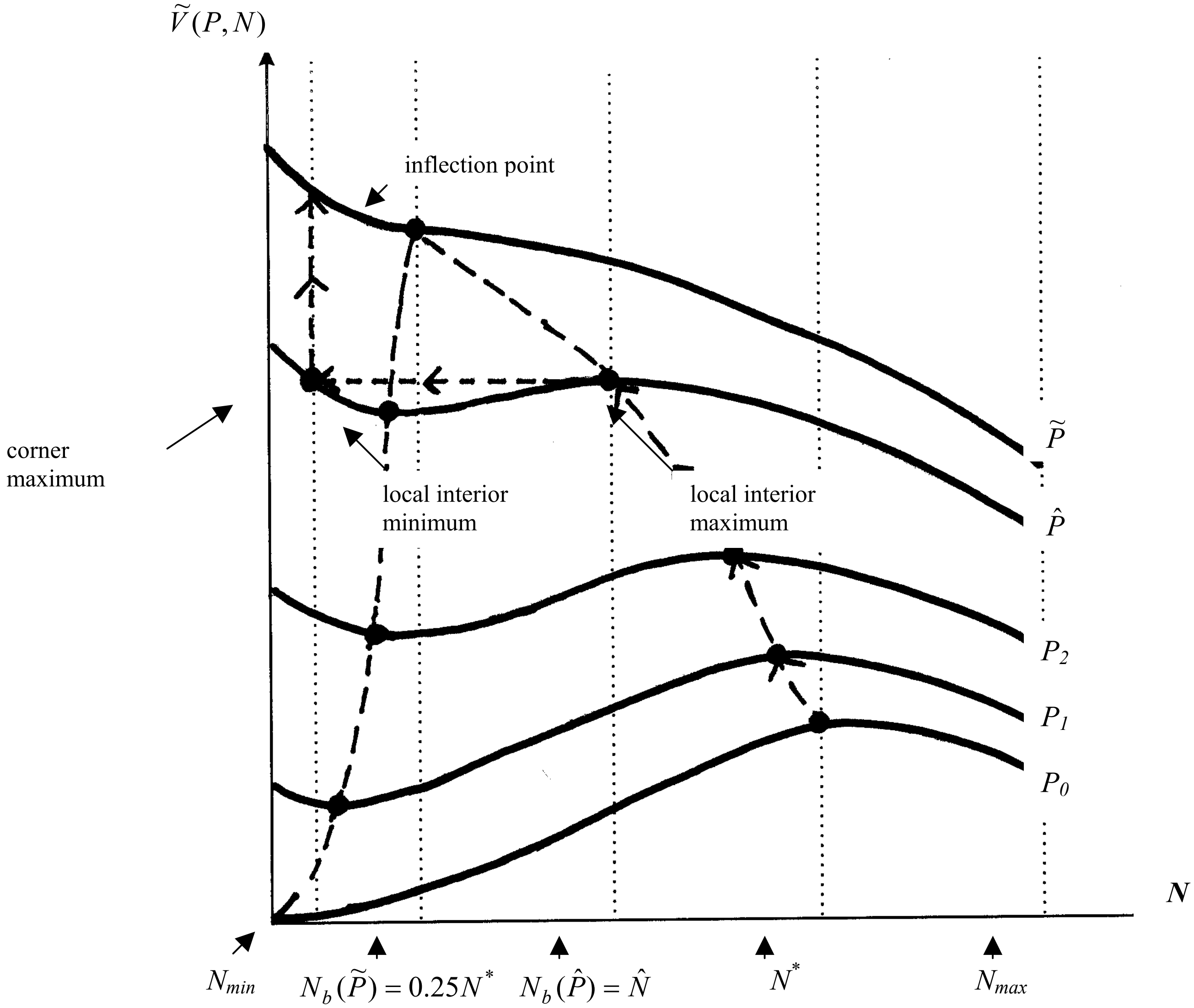

FIGURE 2: The de-agglomeration path of city size and welfare under aggregate population growth (Proposition 1)

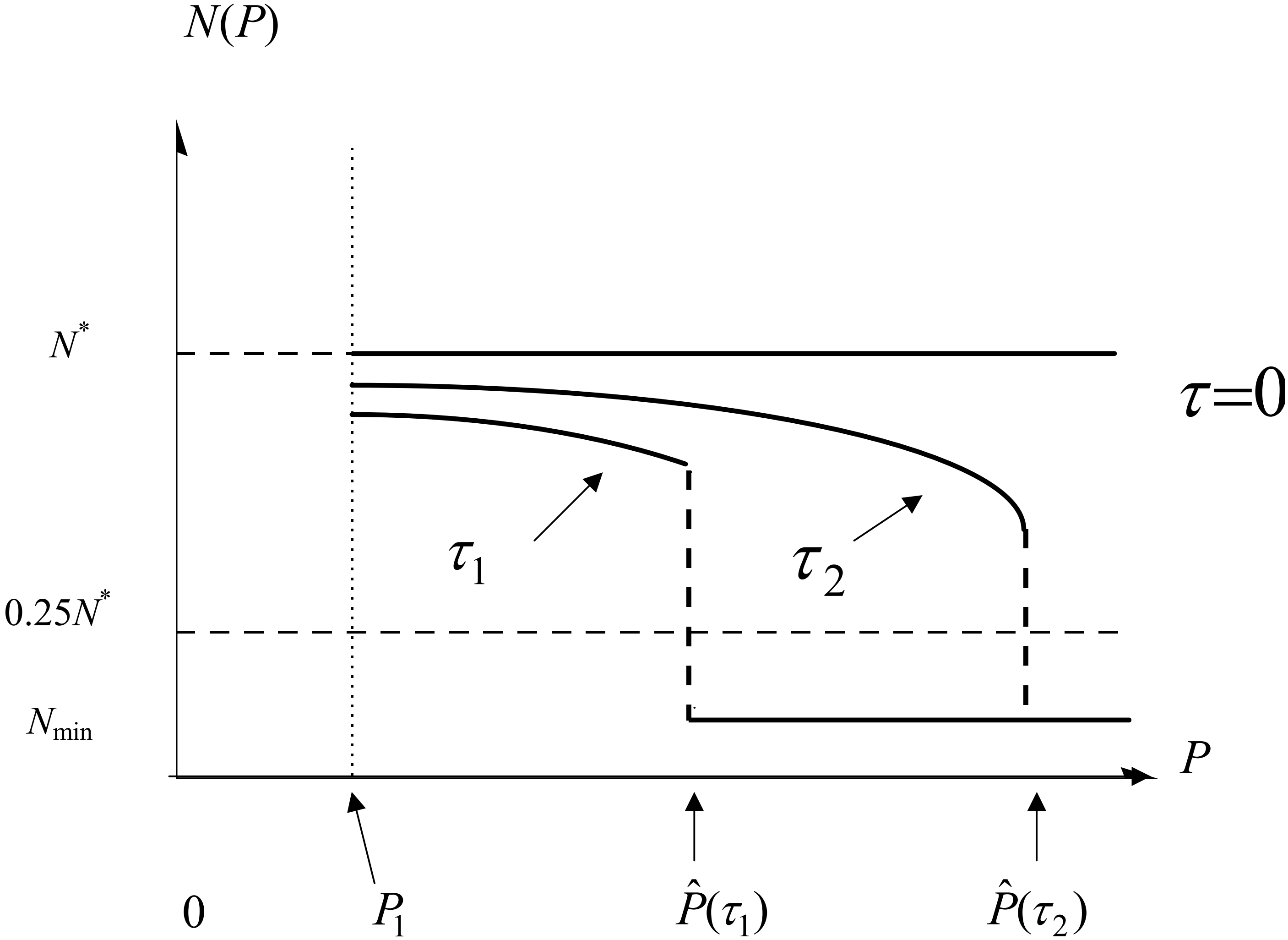

FIGURE 3: The de-agglomeration path of city size with growing aggregate population in the system of cities under different iceberg trading costs. $\tau_2 < \tau_1$.