SYMPOSIUM ON ECHOS OF HENRY GEORGE'S ECONOMICS IN MODERN ANALYSIS

Does the Henry George Theorem Provide a Practical Guide to Optimal City Size?

By RICHARD ARNOTT*

ABSTRACT. The spatial distribution of economic activity is determined by a balancing of increasing and decreasing returns to scale activities. The Henry George Theorem states roughly that, if economic activity is efficiently organized over a "large" space, aggregate land rents equal the aggregate losses from the decreasing returns to scale activities. Kanemoto, Ohkawara, and Suzuki have tentatively applied the Henry George Theorem to investigate whether Tokyo has too large a population. This paper has two aims. The first is to explore the Theorem and its generality; the second is to examine whether it provides a promising conceptual foundation for estimating whether particular cities are over- or underpopulated.

I

Introduction

THE *BASIC* HENRY GEORGE THEOREM states that, *with identical individuals, in a city of optimal population size, differential land rents* (the aggregate over the city of urban land rent less the opportunity cost of land in nonurban use) *equal expenditure on pure local public goods*. The Theorem is so named because it characterizes a situation in which only Henry George's "single tax"—a confiscatory tax on land

*The author can be reached at richard.arnott@bc.edu and serves as a Professor of Economics at Boston College. His principal areas of research are urban and public economic theory and his recent research has been on the economics of property taxation and of downtown parking. He currently edits *Regional Science and Urban Economics*.
This paper was presented at the Southern Economic Association Meetings, 2002. The author would like to thank the discussants Randall Holcombe and Nicolaus Tideman for their helpful comments.

The American Journal of Economics and Sociology, Vol. 63, No. 5 (November, 2004).

rents—is needed to finance urban public expenditures. In the models for which the Theorem holds, there are two opposing effects that interact to determine optimal city size. As population size increases, the fixed cost of the pure local public goods can be shared between a larger number of residents; this is the single source of spatially localized increasing returns to scale that encourages the agglomeration of economic activity. But as population size increases, marginal travel costs and hence the marginal cost of "producing" lots increases; this is the single source of spatially localized decreasing returns to scale that encourages the spatial dispersion of economic activity. With optimal population size, at the margin the increasing returns to scale due to the local public goods just balance the decreasing returns to scale due to land scarcity. At the corresponding locally constant returns to scale allocation, the average cost of providing residents with an exogenous level of utility is minimized.

With heterogenous individuals who may differ according to tastes or production characteristics, optimal population size is not well defined. In these circumstances, the Henry George Theorem states that in *any Pareto optimal allocation* differential land rents equal expenditures on pure local public goods.

The *generalized* Henry George Theorem allows for multiple sources of spatially localized increasing returns to scale and of spatially localized decreasing returns to scale. Sources of spatially localized increasing returns to scale include, in addition to pure local public goods, increasing returns to scale in production (which can be internal to the firm, external to the firm but internal to the industry (so-called localization economies), or external to the firm but internal to the city (so-called urbanization economies)) and localized congestible facilities with decreasing long-run average costs. Sources of spatially localized decreasing returns to scale, in addition to land scarcity, include localized disamenities such as air pollution and localized congestible facilities with increasing long-run average costs. The generalized Henry George Theorem also allows for distortions. The generalized Henry George Theorem states that *in any constrained Pareto optimal* (which allows for unalterable distortions) *and nontrivial* (neither indeterminate, completely agglomerated, nor completely dispersed) *allocation of population in a spatial economy, the*

aggregate shadow losses from the increasing returns to scale activities (losses evaluated at social opportunity costs or shadow prices) *just equal the aggregate shadow profits from the decreasing returns to scale activities.*

This paper has two aims. The first is to provide an intuitive and nontechnical presentation of the Theorem and to demonstrate how very general the Theorem is. The second is to address the question that forms the title of the paper: Does the Henry George Theorem provide a practical guide to optimal city size? This question entails two subquestions: Can the Theorem in principle be applied to infer in what ways the distribution of population over a system of cities is distorted? And if the answer to this question is affirmative, are the data available that would be needed to calculate whether an actual city is over- or underpopulated, or could these data be collected? In addressing these questions, the paper will draw heavily on Kanemoto, Ohkawara, and Suzuki (1996a), which uses the Henry George Theorem as the conceptual basis in estimating whether Tokyo is too large.

The rest of the paper is organized as follows. Section II will present perhaps the simplest model illustrating the basic Henry George Theorem. Section III will explore the Theorem's generality. Section IV will briefly summarize the literature on overpopulation and then critically review the procedure employed by Kanemoto, Ohkawara, and Suzuki (1996a) and compare it to alternatives. Section V will conclude.

II

The Economics of the Basic Henry George Theorem

THE BASIC HENRY GEORGE THEOREM was first presented in Flatters, Henderson, and Mieszkowski (1974) in the context of a regional economic model. This section presents the Theorem in the context of an urban economic model, which is a simplified version of that presented in Arnott and Stiglitz (1979).

We start off with the simplest model that illustrates the Theorem—a circular, monocentric city with a point central business district (CBD), identical individuals, and fixed lot size.

The geography of the economy is a featureless plain extending indefinitely far in every direction. There is a mean density of population over the plain. The technology is everywhere the same. A generic good is produced under constant returns to scale, to the sole factor, labor, of which each individual inelastically supplies one unit. One unit of the generic good can be transformed into one unit of consumption good, one unit of transport service, or one unit of a pure local public good, which provides benefits to all the city's residents without congestion but to no one living outside the city. The technology of production requires that all residents in a city work at that city's point CBD. Individuals have identical tastes and derive utility from consumption, lot size, and the local public good. Tastes are such that it is efficient to provide each individual with a lot of unit size.

The economy is centrally planned by a benevolent despot who decides how to allocate the economy's resources so as to maximize the common utility level of its residents. In particular, she must decide on the population of each city, as well as the allocation of output produced by the city's residents between consumption, transport service, and the local public good. In deciding on each city's population, she faces a tradeoff; in a more populous city, the cost of the local public good is divided among a larger population but at the same time average commuting distance is increased.

The following notation is employed:

N	population of a city
Y	per capita output
C	per capita consumption
P	units of the pure local public good
t	units of transport service required to move an individual a unit distance
x	distance from the CBD
b	distance of the urban boundary from the CBD
U	utility function
s(x)	shadow land rent at x
r(x)	market land rent at x

The planner chooses the geometry of cities so as to minimize aggregate commuting costs at every level of population. It is assumed that the optimal allocation entails no aggregate land scarcity, so that cities are circular. Furthermore, land in nonurban use is unutilized and therefore has no scarcity rent. For a city of population N, the resource constraint is

$$NY = NC + \int_0^b (tx)2\pi x dx + P. \tag{1}$$

This indicates that aggregate output of the generic good goes toward aggregate consumption, aggregate transport services, and units of the pure local public good. N residents require N units of land, which entails an urban radius of $b(N) = \left(\frac{N}{\pi}\right)^{1/2}$. Evaluating Equation (1) yields

$$NY = NC + \frac{2}{3} t\pi^{-1/2} N^{3/2} + P. \tag{2}$$

To simplify the algebra, let $m = \frac{2}{3} t\pi^{-1/2}$. The planner chooses N, C, and P to maximize utility $U = U(C, P)$, subject to the resource constraint (written for convenience in per capita terms). The corresponding Lagrangean is

$$L = U(C, P) + \lambda\left(Y - C - mN^{1/2} - \frac{P}{N}\right), \tag{3}$$

for which the first-order conditions are

$$N: \lambda\left(-\frac{m}{2} N^{-1/2} + \frac{P}{N^2}\right) = 0 \tag{4}$$

$$C: U_c - \lambda = 0 \tag{5}$$

$$P: U_p - \frac{\lambda}{N} = 0. \tag{6}$$

From Equation (4),

$$P = \frac{m}{2} N^{3/2} = \frac{ATC}{2}, \tag{7}$$

where *ATC* is aggregate transport costs in the city. Thus, whatever the level of public good, optimal population is that for which expenditure on the public good equals one-half aggregate commuting costs.

The shadow rent on land at a distance x from the CBD is the resource saving from having an extra unit of land there. Moving an individual from the boundary of the city to the extra unit of land at x would result in a resource saving $t(b - x)$. The shadow rent on land at the boundary of the city equals the shadow rent on land in nonurban use, which equals zero. Hence,

$$s(x) = t(b - x). \tag{8}$$

Integrating $s(x)$ over the area of the city gives aggregate shadow land rents (ASLR):

$$\begin{aligned} \text{ASLR} &= \int_0^b s(x)2\pi x dx \\ &= \frac{m}{2} N^{3/2}. \end{aligned} \tag{9}$$

Comparing Equations (9) and (7) yields

$$P = \text{ASLR}, \tag{10}$$

which is the basic Henry George Theorem (HGT hereafter) with a zero opportunity rent on land in nonurban use. Note that the first-order conditions for C and P were not used in deriving this result. Thus, the Theorem holds for any level of the pure public good, not just the optimal level that satisfies the Samuelson condition.

The above result can be illustrated diagrammatically. Hold P fixed, and plot as a function of N the aggregate amount of the consumption good after allocation of units of the generic good used up in the production of the public good and of transport services. From Equation (2):

$$NC = NY - P - mN^{3/2}.$$

Since lot size and the level of the public good are fixed, utility is maximized by maximizing per capita consumption. Now, this per capita variable is an average, and at an interior optimum of an average, the average equals the marginal. The average equals

$$\frac{NC}{N} = Y - \frac{P}{N} - mN^{1/2} \tag{11}$$

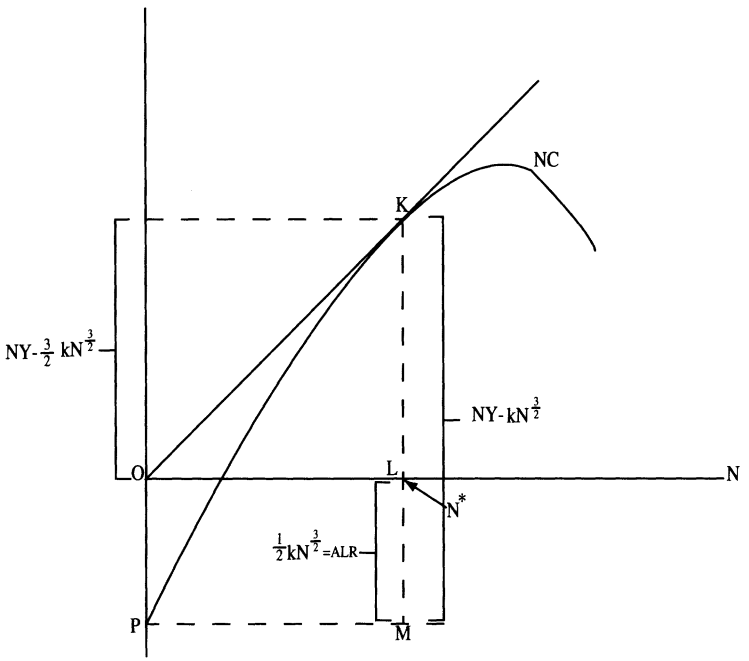
while the marginal equals

$$\frac{dNC(N)}{dN} = Y - \frac{3}{2}mN^{1/2}. \tag{12}$$

Equating the two yields the basic HGT. Diagrammatically, per capita consumption at a particular N is given by the slope of the ray from the origin to the point on the NC function corresponding to that N, while the marginal equals the slope of the total curve at that N, and the two are equal where the ray from the origin is tangent to the NC function. In terms of Figure 1, per capita consumption is maximized

Figure 1

Diagrammatic depiction of the Henry George Theorem.



at N^*. The distance KL equals NC at N^* and also the marginal of NC at N^* times N^*. From this, the HGT follows immediately.

The basic HGT can also be derived as the solution to a dual problem. Let U^* be the utility obtained at the optimum. The optimal allocation can be solved as the minimum per capita resource cost of providing this level of utility. Now, at a local interior minimum of average cost, there are locally constant returns to scale. The Product Exhaustion Theorem[1] states that with constant returns to scale and marginal-cost/shadow pricing, the value of output equals the value of inputs or, alternatively, that shadow profits are zero. Applying the Product Exhaustion Theorem to the above economy implies that, in a city of optimal population size, the shadow profit from production equals zero. There are three goods in the economy: the consumption good, the pure local public good, and lots. The consumption good is produced under constant returns to scale, the pure local public good under increasing returns to scale, and lots under decreasing returns to scale (since the lots are produced with land and transport services, and a doubling of population requires a doubling of land and a more than doubling of transport services). Since the consumption good is produced under constant returns to scale, the shadow profits from its production are zero. Since the pure local public good is by definition uncongestible, its marginal cost price is zero, so that under marginal-cost pricing the production of the public good generates a shadow loss equal to the resource cost of its production, which in the model is simply the level of the public good. Since the marginal lot is at the boundary, the marginal cost of a lot equals the travel services used up in commuting to the boundary, which is increasing in population. Thus, the shadow profit made in the production of lots equals (the cost of commuting to the boundary lot times population) less aggregate transport costs. Since a lot's shadow rent equals the inframarginal surplus associated with its production (in other words, the difference between the marginal-cost price and the cost of producing it), aggregate shadow land rents equal the shadow profit made in the production of lots. The aggregate shadow profit from production therefore equals the shadow profit from the production of lots minus the shadow loss in the production of the public good. Thus, in a city of optimal population size for

which the shadow profits from production are zero, the shadow loss from the production of the public good equals the shadow profit from the production of lots, which in the context of the model implies Equation (10).

The above argument can be generalized to economies that are like the one above but contain multiple groups of individuals in specific proportions.[2] A Pareto optimal allocation is one that minimizes the average resource cost of providing individuals in group *j* with exogenous utility level $\overline{U}_j$, for j = 1, —, J, in those proportions. Since the allocation is average-cost minimizing, the same general line of argument applies. Aggregate shadow profits are again zero, with shadow profits from the production of lots equaling the shadow losses from production of the pure local public good. There are, however, important qualitative differences between the optimal allocation with identical individuals and a Pareto optimal allocation with multiple groups of individuals. First, while the solution to the above model with identical individuals entails a single type of city, the solution to the extended model with multiple groups of individuals in specific proportions entails a different city for each group. Instead of an optimal city, there is an optimal system of cities corresponding to each Pareto optimal allocation. We shall refer to a spatial unit that is replicated at an optimum as a *spatial unit of replication*. Second, although with homogeneous individuals there is a unique Pareto optimal allocation so that there is no ambiguity in referring to *the* optimal allocation, with multiple groups all the allocations corresponding to points on the utility-possibility frontier are Pareto optimal so that an optimal spatial unit of replication must be defined with respect to a vector of utility levels, $\overline{U}_j$. In the next section, we shall show that the above line of argument extends in many other ways as well.

Note that Figure 1 and the argument corresponding to it would be unchanged if OP were a fixed cost associated with the production of the generic good (with constant marginal costs, as in the basic model) rather than the level of the public good. This suggests that the HGT generalizes to situations in which the source of increasing returns to scale is increasing returns to scale in production rather than a pure local public good; that this is indeed the case will be argued in the next section.

Thus far we have dealt with planning allocations. The optimal allocation with identical individuals can be decentralized as a quasi-competitive equilibrium. Let the generic good be the numeraire. Then, for each city, if the planner chooses a common level of the pure local public good, announces a wage equal to the (value of the) marginal product of labor, sets market rents equal to the shadow rents computed as above, uses the land rents to finance the public good (or alternatively assigns ownership shares in urban land to urban residents and then applies a confiscatory tax to urban rental income), and then assigns individuals to cities of optimal population size (conditional on the level of the public good), the government's budget will balance and the optimal allocation (conditional on the level of the public good) will be supported as a quasi-competitive equilibrium. All markets will clear, and no one will have an incentive to move from one city to another. Furthermore, if the government additionally chooses the level of the pure local public good so as to satisfy the Samuelson condition, the resulting optimal allocation will be supported as a quasi-competitive equilibrium. A quasi-competitive equilibrium as implicitly defined above differs from a conventional Arrow-Debreu competitive equilibrium in two respects: first, the government provides a pure local public good and finances it from land rents; and second, the government initially assigns individuals to cities but allows them to move subsequently (though they will choose not to do so). If the government did not initially assign individuals to cities, migration-related market failures of the type discussed in Kanemoto (1980) might result.

The optimal allocation with identical individuals may also be attained as a city developer equilibrium. The argument was first put forward by Henderson (1985). One variant runs as follows. Introduce a class of economic agents called city developers. Each city developer sets up a city by renting the land required (in the above model, the rent on nonurban land was set equal to zero to simplify the algebra), offering individuals a specific level of utility in return for their working in his city, and organizing his city so as to maximize net city surplus (the value of output produced by the city residents minus the cost of providing them with the contracted level of utility),

which possibly entails restricting population size. The city developer clearly has an incentive to operate efficiently in order to provide the contracted level of utility at minimum resource cost. Furthermore, competition between developers will drive up the utility level to the point at which a city developer who operates efficiently makes zero profit—zero net city surplus. In the corresponding allocation, households receive the optimal utility level that is provided at minimum cost per capita, and cities are of optimal size. Another variant of the argument entails developers employing the market mechanism in their cities, decentralizing production of the generic good, instituting a market for land, and purchasing units of the public good. Competition between city developers results in the optimal allocation, in which the aggregate land rents each receives just cover the cost of the pure, local public good he provides.

The model employed in this section to illustrate the basic HGT makes a number of simplifying assumptions in addition to identical individuals—fixed lot size; constant transport services used up per unit distance travelled; linear production possibilities between the pure local public good, the consumption good, and transport services; a zero opportunity rent on land in nonurban use; a single consumption good; a single pure local public good; and globally constant returns to scale in production at a point CBD using only labor as input. It can be shown (see, for example, Papageorgiou and Pines 1998) that the basic HGT continues to apply with variable lot size, nonlinear transport costs and multiple transport modes, a positive opportunity rent on nonurban land, multiple consumption goods, multiple pure local public goods, multiple factors of production, including possibly land, and alternative specifications of the spatial characteristics of production internal to the city, consistent with polycentricity. Two qualifications are needed, however. First, the combination of multiple goods and multiple factors may generate local nonconvexities in production sets (see Wilson 1987), which may result in multiple city types at the optimum, even when individuals are identical, resource endowments uniform, and the production technology the same everywhere. The basic HGT continues to hold but for each spatial unit of replication rather than for each city. Second, with a

positive opportunity rent on nonurban land, the aggregate land rents in the Theorem are replaced by differential land rents—aggregate land rents less the aggregate opportunity cost of urban land.

This concludes our review of the basic HGT. In the next section we consider generalizations of the basic HGT, in particular the generalizations to economies with alternative sources of increasing and decreasing returns to scale, overlapping jurisdictions, spatial inhomogeneity, durable housing and infrastructure, and distortions.

But before proceeding, now is an opportune point for a brief digression on Henry George and the Henry George Theorem. Henry George (1880) argued that the benefits of technical progress, which he regarded as largely atmospheric and nonappropriable, are capitalized into the value of land. His argument for a single tax ran along the following lines. The taxation of land is efficient, neither generating static deadweight loss nor discouraging technical progress. That landowners as a group reap all the benefits of technical progress as windfall gains is unfair. Equity can be efficiently improved by imposing a tax on land, with the revenues being spent on public services that benefit everyone. George furthermore argued that since a tax on land would generate sufficient revenue to finance public services, it is *the single tax* needed. To my knowledge, there is no suggestion in George's writings that a confiscatory tax on land raises just the right amount of revenue to finance pure public goods in a city of optimal population and is therefore the single tax needed to finance them. Apparently, therefore, George's writings did *not* anticipate the Henry George Theorem. Rather, the Henry George Theorem provides a different argument for the single tax.

III

The Generalized Henry George Theorem

RECALL THAT THE GENERALIZED HGT STATES that for any Pareto optimal allocation in a spatially homogeneous economy with a "nontrivial" pattern of agglomeration (which excludes locational indeterminacy due to the absence of transport costs, complete dispersion, and agglomeration of all activity at a point in space), aggregate shadow profits are zero for each spatial unit of replication. The argument runs

as follows. Associated with any Pareto optimal allocation is a vector of utilities, one level of utility for each type of individual in the economy, and a particular distribution of individuals by type. Any such allocation minimizes the per capita resource costs of achieving this vector of utilities; if it did not, the allocation would not be Pareto optimal. Thus, any nontrivial Pareto optimal allocation corresponds to an average cost minimum and hence locally constant returns to scale. The Product Exhaustion Theorem then implies that at any Pareto optimal allocation shadow profits are zero, with the shadow profits from the decreasing returns to scale activities being exactly offset by the shadow losses from the increasing returns to scale activities. Furthermore, since the economy is spatially homogeneous, it is divided into identical spatial units of replication, in each of which therefore shadow profits are zero.

The previous section focused on a special case of the generalized HGT in which individuals are identical and the source of increasing returns to scale is pure local public goods. With identical individuals, the Theorem also holds when the source of increasing returns to scale is instead increasing returns to scale in production, whether these are internal to the firm, external to the firm but internal to the industry, or external to the firm but internal to the city or even group of cities. Indeed, this is the variant of the generalized HGT that was discovered independently by Serck-Hanssen (1969), Starrett (1974), and Vickrey (1977).[3] In this situation, the Theorem is that, in a city of optimal population size, aggregate land rents equal shadow losses in production. Since shadow losses in production equal the (local) degree of returns to scale times the shadow value of output, this variant of the Theorem typically states that, in a spatially homogeneous economy with identical individuals and in which the single source of increasing returns to scale is increasing returns in production, in a city of optimal population size aggregate land rents equal the value of output times the degree of returns to scale in production. Arnott (1979a) demonstrated that the Theorem holds as well when the source of increasing returns to scale is decreasing long-run average costs for a congestible facility. Since the Theorem concerns the relationship between different aggregate economic values, evaluated at the shadow prices/social marginal costs corresponding to an

optimal allocation, its application to congestible facilities entails their use being priced at marginal social cost and their capacity size being optimal. Accordingly, this variant of the Theorem states that, in a city of optimal population size, aggregate land rents equal the loss incurred by the congestible facility when its capacity is optimal and its use priced at marginal social cost. Since this loss equals the cost of constructing optimal capacity minus the revenue raised from marginal-cost pricing its use, this variant of the Theorem may alternatively be stated as follows: under the specified conditions, aggregate land rents plus the revenue raised from marginal-cost pricing the congestible facility just cover the cost of constructing optimal capacity. Arnott (1979a) also showed that, when there are multiple sources of increasing returns to scale, the relationship between aggregates is additive; thus, for example, if there are pure local public goods *and* increasing returns to scale in production *and* congestible facilities, the Theorem states that, in a city of optimal population size, aggregate land rents plus the revenue raised from marginal-cost-pricing use of the congestible facilities cover the shadow losses from production plus the costs of constructing congestible facilities of optimal capacity plus expenditure on local public goods.

In the above paragraph, it was implicitly assumed that each city has only one congestible facility of each type. Realistically, however, not only are there typically multiple types of congestible facilities in a city (schools, swimming pools, parks, etc.) but also it may be efficient to have multiple facilities of a particular type, either because the cost-minimizing scale of the facility is smaller than the city's population or because residents' costs of travel to use the facility are reduced by having multiple facilities, each perhaps having its own jurisdiction, such as a school district. These complications were considered by Hochman, Pines, and Thisse (1995), who showed that the generalized HGT continues to apply but that decentralization of a Pareto optimal allocation requires a metropolitan government.

With multiple goods and increasing returns to scale in production, Pareto optimality typically entails cities differing in industry structure. If the only source of increasing returns to scale is increasing returns to scale internal to individual firms or individual industries, cities are completely specialized in production. More realistically, there are

complementarities in production (backwards, forwards, and sideways linkages) between industries, with the form of the complementarities determining the pattern of co-location of industries across cities. In our large economy, there would be multiple New Yorks specializing in the FIRE industries and fashion, multiple Los Angeleses specializing in entertainment-related industries, and so on.

We have noted that when there are multiple types of individuals the HGT also applies to all Pareto optimal allocations. Thus, in principle, population heterogeneity causes no problems. It does, however, raise two issues concerning the Theorem's practical relevance. First, decentralization of a Pareto optimal allocation may entail redistribution between groups, which may lead one to reasonably question the political feasibility of such lump-sum redistribution.

Second, as heterogeneity of consumption goods, pure local public goods, congestible facilities, factors, and individuals increases, intuitively the size of a spatial unit of replication should increase; rather than a single city type, there will be systems of cities with different cities of different sizes, different population compositions, different industrial mixes, different mixes of public goods, and so on. Even with relatively little heterogeneity, the population size of a spatial unit of replication may be large relative to the population of a small country, so that the country can accommodate only say 1.4 average-cost-minimizing spatial units of replication. This gives rise to what is referred to in the literature as *the integer problem*, that the population of the country is not close to an integer multiple of the average-cost-minimizing population of a spatial unit of replication. What, then, will be the *constrained* optimum allocation of population in the country, taking into account the overall population constraint? It may entail only one spatial unit, two identical spatial units, each accommodating half the population, or two quite different spatial units. The Henry George Theorem is derived on the assumption that the economy is indefinitely large. Does the Theorem still hold when this is not the case? Imagine plotting the average cost of efficiently providing a vector of utilities, each element corresponding to a group of individuals, holding the proportions in each group constant, as a function of the overall population of the country. This function will have global minima at 1.0, 2.0, . . . times the population size of an

average-cost-minimizing spatial unit of replication, but it may well have other local minima at noninteger values. Thus, even when the country's population is less than the population size of a spatial unit of replication, one cannot say a priori whether a constrained Pareto efficient allocation entails operation at a point of increasing or decreasing average cost, nor therefore whether aggregate shadow profits for the country as a whole will be positive or negative. With greater heterogeneity, the integer problem may be significant even for a large country. The degree to which in practical situations the integer problem causes the HGT to be violated has not been investigated empirically.

The HGT is derived on the assumption that land is homogeneous, but in reality locations differ in terms of fertility, natural amenities such as visual beauty and climate, and natural accessibility such as access to the sea or a navigable river. How do these Ricardian differences in land affect the Theorem qualitatively, and how important are they quantitatively? To my knowledge, this question has not been investigated in the literature.

Thus far, we have considered only static economies. Does the Theorem extend to intertemporal economies? In the absence of any state dependence, the answer is obviously affirmative, since the economy can be optimized anew every period. The most obviously important source of state dependence is durable capital, and in the urban context the most important forms of durable capital are buildings and public infrastructure such as roads. Arnott and Kraus (1998) have shown that the Mohring-Harwicz-Strotz self-financing results for congestible facilities, which were originally established for static economies, generalize to dynamic economies, but with the results in discounted terms. In light of the close similarity between the generalized Henry George Theorem and the Mohring-Harwicz-Strotz results, which is explored in Berglas and Pines (1981), it is natural to conjecture that the generalized Henry George Theorem extends to intertemporal settings. Consider, for example, extending the basic static model of the previous section to allow for durable housing and durable pure local public goods (such as a lighthouse). For the static model, the result was that in a city of optimal population size, aggregate land rents equal expenditure on the pure, local public good. The

analog in a dynamic model with durable housing would be that, in cities whose population size follows the optimal trajectory, aggregate land *values* equal the discounted present value of expenditure on the pure, local public good.[4]

The final generalization to be considered in this section is distortions. As indicated earlier, the generalized Henry George Theorem applies for any Pareto optimal allocation in a large spatial economy with a nontrivial pattern of agglomeration. Furthermore, when a Pareto optimal allocation in such an economy is decentralized as a quasi-competitive equilibrium, market prices equal the corresponding shadow prices, so that the Theorem holds when aggregates are evaluated at market prices. The presence of unalterable distortions, however, generally precludes decentralization of Pareto optimal allocations and causes shadow prices to deviate from market prices. It might appear therefore that the generalized Henry George Theorem does not hold in distorted economies; if it does not, the practical relevance of the Theorem is dubious since uninternalized production externalities, interaction externalities, and congestion externalities are very important features of all real-world cities. Fortunately, the Henry George Theorem can be adapted to cover distorted economies. The intuition is that a *constrained* Pareto optimal allocation will still be an average cost minimum when evaluated at shadow prices. The presence of distortions introduces additional constraints, so that the planning problem becomes one of minimizing resource costs per capita of providing the exogenous vector of utilities to the various groups, holding fixed the proportion of individuals in each group, as before, but now subject to the additional constraints imposed by the distortions. This corresponds to minimizing average costs measured in terms of shadow prices. Thus, *in distorted urbanized economies the generalized Henry George Theorem continues to hold when the aggregate magnitudes are valued at shadow prices.*

Shadow prices can be calculated in computer simulation models of the urban economy, but such models have not yet been developed to the point where one can be confident that the calculated numbers correspond well to actual shadow prices. In applying the generalized HGT to estimate whether a particular city has a

larger-than-optimal or smaller-than-optimal population, or whether the size distribution of cities is distorted in a particular way, it may therefore be necessary to value the economic aggregates in the Theorem at market prices. The question then arises: *In distorted urbanized economies, does the Henry George Theorem continue to hold when aggregate magnitudes are valued at market prices, and if not, is there a systematic bias in the use of market rather than shadow prices?* It is easy to construct numerical examples indicating that the answer to the first question is in general negative, though there are some special cases in which, despite distortions, market and shadow prices coincide. The answer to the second question depends on the nature of the distortions, as well as other aspects of the urban economy.

There are two major distortions associated with urban economies. The first is agglomeration externalities in production, the second unpriced transport congestion. Consider first agglomeration externalities in production. Insufficient empirical research has been done in measuring these externalities to identify the channels through which they operate. In some theoretical specifications, the externalities operate via labor; in this case, the market undervalues labor relative to output and other inputs. In some other theoretical specifications, the externalities operate via output, in which case the market undervalues all inputs relative to output, so that the ratio of the value of inputs to the value of output is higher when evaluated using shadow prices than at market prices. Consider next unpriced transport congestion. Unpriced transport congestion results in the market undervaluing the cost of transport services. This distorts land use in a complicated way (see Kanemoto 1980; Arnott 1979b), in general causing the market rent on land to exceed the shadow rent on land at some locations and to fall short of it at others; whether aggregate shadow land rents are greater or less than aggregate market rents depends in a complicated way on road capacity at various locations, the technology of congestion, the shape of the city, and the the properties of the demand for land.

There are other distortions. Many, such as those deriving from asymmetric information, are aspatial, and it is hard to see what

systematic bias such distortions could introduce. Some other distortions, such as externalities associated with incompatible land uses, would appear to be too localized to introduce any systematic bias. There are, however, two other classes of distortion that are likely to substantially affect the relationship between shadow and market land rents. The first is land use controls. By introducing a constraint on the profit-maximizing use of the land, a binding land use control applied to a single plot of land causes the rent on that land to fall. This, however, is a partial equilibrium result. Furthermore, it should be recognized that many land use controls are imposed in response to externalities and may therefore reduce the difference between market and shadow rents. At a general equilibrium level, how land use controls affect the relationship between aggregate market land rents and aggregate shadow land rents depends in a complicated way on the properties of the urban economy and the form of controls. The second is land-related taxes, of which the property tax is particularly important in North American cities. Because land is inelastically supplied, taxes on land rent are fully backward-shifted. If, therefore, there were no distortions, application of a land rent tax would result in shadow rents equalling the gross-of-tax market rents and therefore exceeding the net-of-tax market rents. The property tax is not, however, simply a land rent tax; it taxes not only "pure" land rent, but also structure rent and rent attributable to improvements to the land. In addition, it is a tax on land *value* rather than on land rent, and a tax on land value has the effect of increasing the discount rate in land use decisions.

Distortions will drive a wedge not only between market and shadow land rents but also between market and shadow prices for other goods, which will cause the shadow valuation of other aggregates to differ from the corresponding market valuation.

In conclusion, in a constrained Pareto optimal economy, in which individuals are efficiently allocated over cities subject to the constraints introduced by distortions, little can be said in general concerning how close the generalized Henry George Theorem, with the corresponding aggregates valued at market rather than shadow prices, comes to holding.

IV

Applying the Generalized Henry George Theorem to Estimate Whether Cities Are Over- or Underpopulated

THE FIRST SUBSECTION WILL PROVIDE a summary review of the literature in urban and regional economics related to over- and underpopulation. The second subsection will summarize the approach taken by Kanemoto, Ohkawara, and Suzuki (1996a) in attempting to determine whether Tokyo is too large. And the third will discuss alternative approaches to estimating whether cities are too large or too small, including that employed by Kanemoto, Ohkawara, and Suzuki (1996a).

A. Over- and Underpopulation of Cities

The standard approach to estimating whether cities are over- or underpopulated entails looking at the market failures associated with individual migration decisions. The line of argument was originally developed in the context of rural to urban migration. The original, crude form of the argument was that, since a peasant does not face the full social costs of moving to a city, there will be excessive migration and cities will be too big. This argument is flawed in three respects: it fails to look at the benefit side; it fails to consider whether the peasant faces the full social costs of living in the countryside; and it concerns the distribution of population between the countryside and the city, implying nothing about whether individual cities are too large or too small or whether the size distribution of cities is distorted in a particular way. Furthermore, many policy applications of the argument have failed to distinguish between marginal and average costs.

The argument can readily be modified to eliminate these flaws, as is done in Papageorgiou and Pines (2000). It is socially optimal for a peasant to move from the countryside to the city if

$$MSB_u - MSC_u > MSB_r - MSC_r, \tag{20}$$

that is, if the marginal social surplus (defined as marginal social benefit minus marginal social cost) generated by the peasant is greater

in the city (indexed by u) than in the countryside (indexed by r). The peasant, however, will base his decision on private rather than social values; in particular, he will choose to migrate from the countryside to the city if

$$\mathrm{MPB}_u - \mathrm{MPC}_u > \mathrm{MPB}_r - \mathrm{MPB}_r. \tag{21}$$

Rural-urban migration will occur to the point where the marginal migrant is indifferent between migrating and not migrating. Migration will be excessive if it is optimal for this marginal migrant to stay in the countryside. Thus, rural-urban migration is excessive if

$$(\mathrm{MSB}_u - \mathrm{MPB}_u) - (\mathrm{MSC}_u - \mathrm{MPC}_u) > (\mathrm{MSB}_r - \mathrm{MPB}_r) - (\mathrm{MSC}_r - \mathrm{MPC}_r). \tag{22}$$

To apply this rule requires estimating the divergence between the social and private values of the marginal benefit and cost of urban residence and of rural residence. The same line of argument can be applied to cities at different levels in a hierarchy of cities.

There is another, quite different argument that atomistic migration will lead to cities that are too big, which is reviewed in Kanemoto, Ohkawara, and Suzuki (1996a). Consider an economy with identical individuals. As the urbanized population grows, more and more cities will be developed. Let $U(N)$ represent utility as a function of a city's population size, which is assumed to have an inverted U shape, and N^* denote optimal city size at which utility is maximized. With atomistic migration, the first city will continue to grow beyond optimal population size, until the population N' is reached, such that $U(0) = U(N')$, at which time an atomistic migrant has an incentive to leave the city, providing the seed for a new city to form. As population continues to grow, the population of the newly created city will grow and that of the first city shrink, until both cities are of size N^*. As population grows further, both cities will continue to grow until each city has a population N', at which time a third city will form, and so on, ad infinitum. According to this line of reasoning, with more than one city, each city's population is nearly always larger than optimal.[5]

No one, to my knowledge, has formally extended the above argument to an economy with a hierarchy of cities differing in their industrial mix, with the largest city or cities containing the industries with

the strongest scale economies. However, Kanemoto, Ohkawara, and Suzuki (1996a, p. 20) argue informally that there is a presumption at least that cities at the top of the hierarchy tend to be more overpopulated than cities further down the hierarchy: "With a hierarchical structure, a new city in a certain [level of the] hierarchy usually comes from a city one level lower. This means that it is relatively easy to increase the number of cities at a lower level of the hierarchy. In contrast, it is extremely difficult to add a new city at the highest level of the hierarchy. . . . [According to this argument], divergence from optimal city size [tends to be] larger for larger cities."

B. The Argument in Kanemoto, Ohkawara, and Suzuki (1996a)

Kanemoto, Ohkawara, and Suzuki (1996a) attempt to infer whether Tokyo is too large. The conventional way to approach the problem would be to examine the divergence between the private and social marginal benefits and costs at different levels of the urban hierarchy, as described in the first paragraph of the previous subsection. Kanemoto, Ohkawara, and Suzuki instead approach the problem from the perspective of the Henry George Theorem. In the paper's central model, only one private good is produced, land is homogenous, all individuals are identical, all the profits from decreasing returns to scale activities are implicitly assumed to be manifest as aggregate land rents, and the only source of increasing returns to scale is Marshallian economies of scale in production with respect to urban population. The last assumption implies that a representative firm has a production function of the form $f(k,n,N) = g(N)h(k,n)$, where k is the capital employed by the firm, n the number of workers employed by the firm, and N the city population, and where $h(\cdot)$ exhibits constant returns to scale and $g(\cdot)$ is increasing in city population. In terms of cost functions, this means that the individual firm faces a horizontal marginal cost curve, with the level of marginal cost falling as city population increases. Under competition, each firm pays its workers the private marginal product of labor, gh_n, which falls short of the social marginal product, $g'h + gh_n$. By employing an extra worker, a firm not only increases its own output but also increases the productivity of all other firms; since the firm ignores this positive externality, the

social marginal product of labor exceeds the private marginal product. Since, however, the ratio of the market wage to the private marginal product of labor equals the ratio of shadow wage to the social marginal product of labor, which equals the ratio of the marginal product of capital to the rental rate, production efficiency obtains. Thus, the Marshallian externality operates on the margin of the city's population. The Marshallian externality is assumed to be the only distortion in the urban economy. Since all individuals are identical and only one good is produced, optimal city size is well defined.

It is shown in the Appendix that the HGT applies in this economy, with aggregates being valued using *market* prices. With the assumed form of increasing returns to scale, the result is that, in a city of optimal size, aggregate market land rents equal the value of output times the local degree of increasing returns to scale (or shadow losses from production). At first glance, this is surprising since there is the uninternalized Marshallian externality, and in general with distortions the HGT holds when aggregates are valued at shadow prices but not when they are evaluated at shadow prices. The reason the Theorem holds with aggregates valued at market prices is that, since it does not upset production efficiency, the Marshallian externality operates only on the margin of population, and the HGT applies with the optimal population, not with the competitively determined level of population.

Kanemoto, Ohkawara, and Suzuki consider as well a somewhat more sophisticated model in which there is in addition social overhead capital. The paper allows social overhead capital to vary in its degree of publicness, ranging from being a pure local public good to being a publicly provided private good. If social overhead capital is priced at marginal social cost, the HGT continues to apply, albeit as a different variant since there is an additional source of economies of scale: aggregate market land rents equal shadow production losses plus the cost of the social overhead capital net of the revenue raised from marginal-cost pricing its use (which equals the cost of the social overhead times its degree of privateness).

Kanemoto, Ohkawara, and Suzuki then *assume* that when the model is generalized in such a way that the optimum is characterized by a system of cities, the optimum is characterized by the HGT

holding *for each individual city* and hence *for each level of the urban hierarchy.* This is not strictly correct. When the optimum is characterized by a system of cities, the HGT applies to each spatial unit of replication but not generally to each city.

Kanemoto, Ohkawara, and Suzuki ignore the complications introduced by heterogeneity in individual tastes and skills, as well as the integer problem. If actual heterogeneity were considered, the population of a spatial unit of replication would likely considerably exceed the population of Japan. Thus, the paper's implicit assumption that a Pareto optimal allocation across cities of Japan's population would result in the HGT holding for the country as a whole is open to question, and the paper's assumption that it would hold for each individual city even stronger.

Japan is one of the very few jurisdictions that systematically collects comprehensive data on land values. Since data on land values are available but not data on land rents, Kanemoto, Ohkawara, and Suzuki develop a formulation of the HGT in terms of values rather than rents. In terms of the paper's model, this formulation would (according to an argument advanced earlier in the paper) state that *aggregate land values equal the discounted present value of shadow production losses.* Since the future time path of the value of output by city is not known, Kanemoto, Ohkawara, and Suzuki make the implicit assumption that the ratio of discounted present value to the value at a particular point in time of shadow production losses is constant across cities. This would hold if the possibly time-varying rate of growth in shadow production losses were constant over cities. This implicit assumption is supported by Eaton and Eckstein (1997), who document that on average the growth rate of population in Japanese cities over the 20th century was essentially independent of city size. Applying this assumption gives a weakened form of the HGT: that the *ratio* of aggregate land values in a city to the city's shadow production losses is constant for different levels of the urban hierarchy. This is the form of the Theorem that Kanemoto, Ohkawara, and Suzuki test.

Kanemoto, Ohkawara, and Suzuki actually put forward other reasons for the form of the Theorem they test. It has already been mentioned that they argue that the processes of migration and city

formation under competition are likely to result in cities on average being overpopulated at all levels of the urban hierarchy, which according to their argument would result in aggregate land rents exceeding shadow production losses at all levels of the urban hierarchy. They then *assume* that, taking the distortions that cause the overpopulation as given, a constrained Pareto optimal allocation of population across cities would be characterized by balanced overpopulation, in particular a constant ratio of aggregate land rents to shadow production losses at all levels of the urban hierarchy.

According to elementary valuation theory, the ratio of an asset's rent to its value equals the interest rate minus the growth rate in rents. But the data that Kanemoto, Ohkawara, and Suzuki employ to test for Tokyo's overpopulation come from the mid-1980s, a period during which land values in Japan rose rapidly, reaching 5.35 times the value of GNP in 1990. Mera (in Mera and Renaud 2000) and others have partly explained this rapid run-up in land values as a result of tax policy changes that resulted in land being used as a tax shelter against succession duties and partly as a speculative boom. Both explanations provide good reasons to believe that aggregate land values were well above the levels consistent with elementary valuation theory and, if the HGT holds, well above the discounted present value of shadow losses in production. Neither explanation for the overvaluation of land suggests that the degree of overvaluation should be related to the level of the urban hierarchy. In applying the HGT to investigate whether Tokyo is too large, this overvaluation of land can be taken into account by examining whether the *ratio* of aggregate land values to the shadow losses from production is larger for Tokyo than for cities at other levels of the urban hierarchy.

Thus, in applying the HGT to determine whether Tokyo is too large, Kanemoto, Ohkawara, and Suzuki make several simplifying assumptions: they ignore heterogeneity in land, population, and goods, thus implicitly assuming that it does not affect the form of the HGT tested; the only distortion they model explicitly is the Marshallian externality in production; they assume that the Theorem applies to each level of the urban hierarchy rather than, as theory indicates, to each spatial unit of replication; they ignore the integer problem; they assume that inefficiencies in migration and city formation would result in a

constrained efficient allocation characterized by balanced overpopulation across levels of the urban hierarchy, implying constancy of the *ratio* of aggregate land rents to shadow losses across different levels of the urban hierarchy; they assume that the ratio of aggregate land values to aggregate land rents is constant across levels of the urban hierarchy; and they assume that the Theorem holds when aggregate values are measured using market rather than shadow prices, which is true in undistorted economies and for the particular distorted economy they consider, but not generally for distorted economies. The string of assumptions underlying the form of the test they employ to determine whether Tokyo is too large is indeed a long one.

Commenting on the data used in the paper and on the econometric specification employed is beyond the scope of this current paper. In all the specifications employed that exclude social overhead capital, the authors find that the ratio of aggregate market land values to shadow losses in production (measured as the current market value of output times the estimated local degree of increasing returns to scale) is not systematically larger for Tokyo than for other cities. The results obtained when social overhead capital is treated are broadly similar. They interpret these results to imply that, if Tokyo is overpopulated, it is no more overpopulated than the average Japanese city; in this sense, Tokyo is not too large.

C. Alternative Approaches to Estimating Whether Cities Are Overpopulated

Public policy in many countries discourages rural-urban migration or growth of the country's largest cities in the belief that the unregulated market leads to overurbanization or excessive concentration of the urban population in the largest cities. These policies are based on perception rather than sound empirical work. Planners tend to subscribe to the overurbanization and overconcentration hypotheses, while economists tend to be agnostic but at the same time to argue that planners undervalue the wisdom of the market and overstate its failures. Empirical work holds the promise of resolving these differences and leading to more enlightened policy.

There was an empirical literature on optimal city population size in the early 1970s. This literature defined optimal population to be

that for which the per capita cost of public services is minimized, which was typically found to be around 250,000. We now have a considerably more sophisticated conception of optimal city size (or more generally the optimal size of a spatial unit of replication) that incorporates not only the per capita cost of public services but also traffic congestion, taxation, and economies of scale in production, and the earlier empirical literature provides little insight into optimal city size according to this conception.

As was discussed in Section IVA, there are two alternative but not inconsistent approaches to estimating deviations between the optimal pattern of settlement and the market-determined pattern. The first looks at deviations between the marginal social and marginal private benefits and costs of the marginal migrant; models in this vein may exclude space, treat it implicitly, or treat it explicitly. The second draws on the Henry George Theorem, with Kanemoto, Ohkawara, and Suzuki being the first and to my knowledge only example.

It would be easy to dismiss the conclusion reached by Kanemoto, Ohkawara, and Suzuki on the basis of the long chain of questionable assumptions made in deriving the overpopulation criterion tested. These assumptions include both essential simplifying assumptions of the underlying model, such as identical individuals, and assumptions made in moving from the model to the overpopulation criterion applied, including the use of market rather than shadow prices. But quantification is an essential element of enlightened policy making, and if the procedure employed by Kanemoto, Ohkawara, and Suzuki is the soundest available, its conclusion should be respected until a sounder procedure is developed. Is there a sounder procedure? The alternative broad approach of examining the divergence between marginal social and marginal private costs and benefits is very general. One important difference between the two approaches is that the concepts of marginal social benefit and marginal social cost incorporate some distributional assumption (in standard cost-benefit practice, a dollar is valued the same to whomever it is given or from whomever it is taken), whereas the HGT applies for any Pareto optimal allocation. Put alternatively, the HGT applies for *any efficient* allocation of population over cities, whereas the notion of optimal city size in the social versus private costs-and-benefits approach is welfarist. Another

important difference concerns the amount of information required to implement the two approaches. The basic approach of Kanemoto, Ohkawara, and Suzuki requires remarkably little information—all that is required are data for each city on aggregate land values and the aggregate value of output, and estimates of the local degree of returns to scale in aggregate urban production, which can be simply obtained using cross-city data on wages and capital-labor ratios.[6] The social versus private costs-and-benefits approach is informationally far more demanding, particularly when the distributional assumption that "a dollar is a dollar" is relaxed. The informational economy of the Kanemoto, Ohkawara, and Suzuki approach is however somewhat illusory, since it derives from the *assumption* that the properties of urban distortions are such that the HGT applies when evaluated using market rather than shadow prices. Once this assumption is relaxed, shadow prices must be estimated, and doing so is just as informationally demanding as estimating social costs and benefits under the dollar-is-a-dollar assumption. The social versus private costs-and-benefits approach has the advantage that is does not rely on the assumption that the integer problem can be ignored; the quantitative importance of the integer problem is a completely open question. The social versus private costs-and-benefits approach has the added advantage that it does not require information on land rents or land values, which are typically unavailable and difficult to estimate.[7] Thus, both approaches have advantages and disadvantages. Since the two general approaches are consistent conceptually, both being derivable from the maximization of social welfare subject to the same complete list of constraints, it would be fruitful to pursue both in empirical work and then to work toward reconciling the empirical results deriving from specific applications of each approach. It might be found, for example, that the ratio of aggregate shadow land rents to aggregate market land rents varies so much from city to city that Kanemoto, Ohkawara, and Suzuki's procedure of approximating aggregate shadow land rents by aggregate market land rents is invalid; or it might be found that the ratio of aggregate shadow land rents to aggregate market land rents is similar across cities, in which case aggregate shadow land rents can be estimated by inflating or deflating aggregate market land rents. It might also be found that the integer

problem can safely be ignored, which would bolster the case for the HGT approach. Or it might be found that the integer problem is so important as to nullify the usefulness of the HGT approach.

V

Conclusion

THE *BASIC* HENRY GEORGE THEOREM states that, in an arbitrarily large, spatially homogeneous economy composed of identical individuals, in which the single source of increasing returns to scale is a pure local public good, the single source of decreasing returns to scale is the production of lots via commuting costs, labor is the only factor of production, and the distribution of economic activity over space is nontrivial, optimal city size is well defined and is characterized by aggregate land rents equalling expenditure on the pure local public good.

While the basic HGT is a striking result and elegant in its simplicity, it only hints at the generic nature of the result. The *generalized* HGT gives the general result that in any large, spatially homogeneous economy with a nontrivial distribution of economic activity over space, for any Pareto optimal allocation aggregate profits are zero for each spatial unit of replication. This holds with heterogeneous individuals, multiple consumption goods, multiple factors of production, multiple local public goods, and multiple sources of increasing and decreasing returns to scale. Furthermore, where there are unalterable distortions in the economy such as unpriced transport congestion, the Theorem continues to hold, subject to two qualifications; first, it holds for aggregate *shadow* profits but not for aggregate market profits, and second, it holds for *constrained* Pareto optimal allocations, where the constraints are the unalterability of the distortions.

The basic intuition for all HGT results is that, for large economies, in either a Pareto optimal or a constrained Pareto optimal allocation, the per capita shadow costs of providing each of the economy's groups of individuals a given level of utility is minimized. Since there are locally constant returns to scale at an interior average (here with respect to the population of a spatial unit of replication) cost minimum, and since shadow profits are zero at an average cost

minimum, shadow profits are zero at any interior Pareto optimal or constrained Pareto optimal allocation.

This paper first derived the basic and generalized Theorems in the context of especially simple models and then explained why each variant of the Theorem generalizes in the way it does. The paper then asked whether the Theorem can be employed to identify how the equilibrium distribution of population over space deviates from a Pareto optimal or constrained Pareto optimal allocation. These questions were addressed through examination of a paper by Kanemoto, Ohkawara, and Suzuki that applies the Theorem empirically with the aim of determining whether Tokyo is too large. This paper also reviewed alternative conceptual approaches employed in the literature to infer how the equilibrium distribution of population differs from an efficient or optimal distribution.

I identified the long string of questionable assumptions that Kanemoto, Ohkawara, and Suzuki (1996a) make in moving from the generalized HGT to the test they employ to identify whether Tokyo is too large. I then argued that, despite these weak links, the result obtained by Kanemoto, Ohkawara, and Suzuki that Tokyo is not too large should be accorded considerable weight in policy circles until other papers are written that provide a superior procedure. An alternative conceptual approach to identifying over- or underpopulation, which could be applied empirically, is to compare marginal social versus private costs and benefits. I identified advantages and disadvantages of the two empirical procedures. Kanemoto, Ohkawara, and Suzuki's procedure is based on stronger assumptions but is also less informationally demanding. Also, tests based on the HGT are tests for an *efficient* allocation of population, whereas tests based on the other procedure incorporate equity considerations since their derivation entails distributional assumptions. Finally, I argued that since the two procedures are derived from the same conceptual foundations, they are essentially complementary. While neither has yet been developed to the point where empirical conclusions based on them are compelling, having them compete in the marketplace of ideas will lead to both being improved.

Does the Henry George Theorem provide a practical guide to

optimal city size? The jury is not yet in, but the approach is sufficiently promising to merit further exploration.

Notes

1. And corollaries of the Theorem state that, with increasing (decreasing) returns to scale and marginal-cost pricing, shadow profits are negative (positive).

2. To my knowledge, Schweizer (1986) was the first work to extend the HGT to heterogeneous individuals.

3. Who first discovered the Theorem is unimportant; it was "in the air." Serck-Hanssen was the first into print, but unpublished notes in the Vickrey archives suggest that Vickrey had derived the Theorem at least before Serck-Hanssen's article was published.

4. At settled locations there may be no vacant land on the basis of which land value can be measured. At such locations, differences in accessibility are capitalized into property values. The relevant variant of the Henry George Theorem would then relate the discounted present value of expenditure on the pure, local public good to the sum of land values at locations where land value can be inferred plus the sum of property value accessibility premia at those locations where land value cannot be inferred.

5. The strength of the argument is weakened slightly if one admits city developers who form new cities of size N‴ less than N^*. In this case, the first city will continue to grow beyond optimal population size until the population reaches N'' at which U(N″) = U(N‴), etc., but the same general conclusion applies.

6. The more sophisticated approach of Kanemoto, Ohkawara, and Suzuki entails measuring the value of social overhead capital, as well as its degree of privateness. It seems reasonable to assume that the degree of privateness of overhead capital is more or less constant across cities, in which case this parameter could be straightforwardly estimated.

7. *Olcott's Land Values* provides especially detailed estimates of Chicago land values dating back over 100 years.

References

Arnott, R. J. (1979a). "Optimal City Size in a Spatial Economy." *Journal of Urban Economics* 6: 65–89.

——. (1979b). "Unpriced Transport Congestion." *Journal of Economic Theory* 21: 294–316.

Arnott, R. J., and M. Kraus. (1998). "Self-Financing of Congestible Facilities

in a Growing Economy." In *Topics in Public Economics.* Eds. D. Pines, E. Sadka, and I. Zilcha. New York: Cambridge University Press.

Arnott, R. J., and J. E. Stiglitz. (1979). "Aggregate Land Rents, Expenditure on Public Goods, and Optimal City Size." *Quarterly Journal of Economics* 93: 471–500.

Berglas, E., and D. Pines. (1981). "Clubs, Local Public Goods and Transportation Models: A Synthesis." *Journal of Public Economics* 15: 141–162.

Eaton, J., and Z. Eckstein. (1997). "Cities and Growth: Theory and Evidence from France and Japan." *Regional Science and Urban Economics* 27: 443–474.

Flatters, F., V. Henderson, and P. Mieszkowski. (1974). "Public Goods, Efficiency, and Regional Fiscal Equalization." *Journal of Public Economics* 3: 99–112.

George, H. ([1880] 1956). *Progress and Poverty.* New York: Robert Schalkenbach Foundation.

Henderson, J. V. (1985). "The Tiebout Model: Bring Back the Entrepreneurs." *Journal of Political Economics* 93: 248–264.

Hochman, O., D. Pines, and J. Thisse. (1995). "On the Optimal Structure of Local Governments." *American Economic Review* 85: 1224–1240.

Kanemoto, Y. (1980). *Theories of Urban Externalities.* Amsterdam: North-Holland Publishing.

Kanemoto, Y., T. Ohkawara, and T. Suzuki. (1996a). "Agglomeration Economies and a Test for Optimal City Sizes in Japan." Unpublished ms.

——. (1996b). "Agglomeration Economies and a Test for Optimal City Sizes in Japan." *Journal of the Japanese and International Economies* 10: 379–398.

Mera, K., and B. Renaud. (2000). *Asia's Financial Crisis and the Role of Real Estate.* Armonk, NY: M. E. Sharpe.

Papageorgiou, Y. Y., and D. Pines. (1998). *An Essay on Urban Economic Theory.* Boston: Kluwer Academic Publishers.

——. (2000). "Externalities, Indivisibility, Nonreplicability, and Agglomeration." *Journal of Urban Economics* 48: 509–535.

Schweizer, U. (1986). "General Equilibrium in Space and Agglomeration." *Fundamentals of Pure and Applied Economics* 5: 151–185.

Serck-Hanssen, J. (1969). "The Optimal Number of Factories in a Spatial Market." In *Toward Balanced International Growth.* Ed. H. Bos. Amsterdam: North-Holland Publishing.

Starrett, D. (1974). "Principles of Optimal Location in a Large Homogeneous Area." *Journal of Economic Theory* 9: 418–448.

Vickrey, W. (1977). "The City as a Firm." In *The Economics of Public Services.* Eds. M. Feldstein and R. Inman. London: Macmillan.

Wilson, J. D. (1987). "Trade in a Tiebout Economy." *American Economic Review* 77: 431–441.

APPENDIX

The Henry George Theorem Holds in the Model of Kanemoto, Ohkawara, and Suzuki (1996a) with Aggregates Valued at Market Prices

To simplify, we particularize the model of Kanemoto, Ohkawara, and Suzuki (1996a) assuming unit lot size. Per capita consumption is

$$\frac{1}{n}(g(N)h(k, n) - rk) - mN^{1/2}, \tag{A.1}$$

per capita production minus per capita rental cost, where r is the exogenous rental price of capital, minus per capita transport costs. The optimal population is that population at which per capita consumption is maximized, and is therefore characterized by the first-order condition with respect to N:

$$\frac{1}{n}g'h - \frac{m}{2}N^{-1/2} = 0. \tag{A.2}$$

Multiplying Equation (A.2) by N^2 yields

$$\frac{N}{n}\left(\frac{g'N}{g}\right)gh - \frac{m}{2}N^{3/2} = 0. \tag{A.3}$$

Now, since $g'N/g \equiv \alpha$, the local degree of increasing returns to scale at the optimum, and since, with the generic good as the numeraire, $\frac{m}{2}N^{3/2}$ is aggregate shadow land rents and $\frac{N}{n}gh$ the aggregate shadow value of output, Equation (A.3) can be rewritten as

$$\alpha(\text{ASVO}) = \text{ASLR}, \tag{A.3′}$$

which is the familiar version of the Henry George Theorem with increasing returns to scale in production, with aggregates valued at shadow prices.

We need to show that the same relationship holds with aggregates valued at market prices. With Marshallian economies of scale, a quasi-competitive equilibrium exists with firms making zero profits. With the generic good as numeraire, the left-hand side of Equation (A.3′) equals the market value of output times the local degree of increasing returns to scale. The market rent on land at the boundary of the city is zero. Furthermore, individuals are willing to pay a premium t to live a unit distance closer to the city center, since this is the saving in commuting costs from doing so. As a result, market land rents coincide with shadow land rents so that aggregate land rents are the same whether evaluated at market or shadow prices. Thus, Equation (A.3′) continues to hold when the aggregates are valued at market prices.