Amount and Spatial Distribution of Public Open Space to Maximize the Net Benefits from Urban Recreation

By Kent Kovacs
Department of Agricultural and Resource Economics
University of California, Davis
kovacs@primal.ucdavis.edu

Selected Paper prepared for presentation at the AAEA
Annual Meeting, Providence, Rhode Island, July 24-27, 2005

I. Introduction

The rapid rise of land area in urban and metropolitan uses has meant a decline in natural lands within and around urban areas. The loss of open space has generated strong public support for growth management in cities. In 2003 and 2002, ballot measures generated $1.8 billion and $10 billion respectively for local and state land conservation bringing the full tally of funds since 2000 to $16.8 billion (Land Trust Alliance, 2004).

A substantial proportion of the benefits from these open spaces come from recreation. Eighty-three percent of the approximately 290 million people in the United States do walking for pleasure, 74% have family gatherings, 54% do picnicking, 52% do sightseeing, and 45% go wildlife viewing. Further, participation in these activities is growing. In the last nine years, 40 million more people began walking for pleasure, family gatherings grew by 36 million, picnicking grew by 20 million, and sightseeing grew by 24 million (National Survey on Recreation and the Environment, 2000).

The spatial distribution of the open spaces in metropolitan areas influences the aggregate net benefits and the equity of the net benefits from recreation received by residents. The focus of this paper is to find the amount of land in parks and the spatial distribution of the city parks to maximize the aggregate net benefits from recreation. The location of a city park influences the cost of a trip by residents to the park. The size of the park influence the net benefits from a park because, for example, a baseball field and a playground are facilities with large space requirements. Further, size influences the ability of parks to connect neighborhoods and house special events like concerts or street fairs. Finding, for a fixed amount of land in parks, the number of the parks to bring residents closer to parks while simultaneously not diminishing too much the size of the parks is a key tradeoff explored in this paper.

Most of the literature on the spatial distribution of public goods like parks is in economic geography and regional science. Gaussier (2001) examines the optimal location of landfills by weighing the cost of transporting the rubbish against the desire for consumers to stay far away from the landfill. A few papers in regional science look at the placement of parks in the standard monocentric city model, but there is no mention of the recreational benefits from these placements. Yang (1990) weighs the optimal size of a central park against the need for a central business district, and Lee (1997) weighs the configuration of a greenbelt against the increased cost of travel in a city.

Economic geography and regional science also have an empirical literature on the spatial distribution of public goods. Witten et. al. (2003) and Hewko et. al. (2002) improve upon measures of the accessibility of public goods, and Lindsey et. al. (2001) examine the accessibility to new kinds of public goods like greenways. A few papers go beyond descriptive statistics to examine the amount and distribution of public goods, but these papers usually look infrastructure rather than specific public goods. Weinhold and Reis (2001) consider the relationship between infrastructure growth and population growth in the Amazon.

Most of empirical literature in economics examines the determinants of the amount of infrastructure without any investigation of the spatial distribution of that infrastructure. Glaser et al. (1995) find that population growth is a stimulant to infrastructure growth, and Cutler and Glaeser (1995) find that ethnic segregation slows investment in public goods. Poterba (1998) finds that an elderly population lowers the investment in education, and Goldin and Katz (1998) find that cities with more educated people invest more in infrastructure. Since these studies do not examine specific public goods, the findings are difficult to use to make inferences about how those determinants would affect the amount and the spatial distribution of parks.

From the model developed to find the amount of land in parks and the spatial distribution of parks to maximize the net benefits from recreation, comparative statics for the influence of city characteristics on the optimal amount of land, number and sizes of parks are generated. Data on parks from metropolitan areas are collected, and the relationship between city characteristics and the amount of land, number and sizes of parks is examined empirically. By comparing the signs of the coefficient estimates to the signs of the comparative statics for the city characteristics, the presence of sub-optimal amounts of land or spatial distributions of parks, from the perspective of maximizing net benefits from recreation, in metropolitan areas is identified. The presence of sub-optimal amounts of land in parks or spatial distributions of parks should alert policy makers to consider changing the guidelines about the creation and placement of public open space.

II. The Model

City parks and recreation departments are responsible for deciding the amount of land for parks, the number of parks, and the size of parks within the area of a city. Along with the amount of land for parks, the division of that land into the different numbers and sizes of parks influences the net benefits people receive from recreation at the parks. More parks mean that travel costs to the parks fall. However, since some of the optimal amount of land in parks is taken away from other parks to create the new park, the size of the other parks fall, and the diminished size of the other parks reduce the net benefits of recreation from those parks to the public.

Suppose a city is a line with length L. Although cities are two-dimensional, a one-dimensional model does not take away from the main results unless the shape of parks is important. The shape of the parks, albeit potentially interesting, is much more mathematically cumbersome to represent. The population is homogenous and uniformly spread over a city. Not all individuals must have the same characteristics, but every neighborhood has the same mix of people that every other neighborhood does. In other words, on average, the population is homogeneous and uniformly spread over a city. Since many cities have neighborhoods significantly different from each other, the assumption is that the population of each neighborhood is homogeneous, and the planner, acknowledging the differences across neighborhoods, chooses the amount, number and size of parks for each neighborhood accordingly. The planner divides the land for parks into parks of equal size according to the rule, $A = ns$, where A is an endogenously determined amount of land for parks, n is the number of parks, and s is the size of each park.

The cost of a park is buying the land for the park and the cost of maintenance for the park. The amount of land for parks chosen by the planner influences the price of the land purchased for the parks. The more land created for parks the greater the purchase price for all land created for parks, i.e. $dp(A)/dA > 0$, since greater amounts of land in parks increase the scarcity of land for other developments, leading the price of land to rise.

Usually not all the land for parks in a city is chosen in a moment. City parks and recreation departments often have budget constraints allowing them only to buy some of the desired amount of land in a year. Further, since parks need to have trails made and buildings cleared, not all parks are instantly available once the land for them is purchased. Also, many cities are growing, and new parks are built in the suburbs many years after parks close to the city center were built. These are all short run constraints however since eventually the necessary funds for purchasing land for parks become available; the parks are eventually all built, and a city stops growing. If the price of land and the socioeconomic characteristics of a city do not change over time, then the short-run constraints may be ignored for modeling the long run optimal spatial distribution of parks.

Consider the placement of city parks in Diagram I. Each zone has length 2x* where x* is the maximum distance from a park that a family living in the city will visit the park. At any distance greater than x*, the travel cost from visiting the park exceeds the benefit. Consequently, families living outside the zones do not visit a park. A families' demand curve for trips to a park makes more explicit what the x* for a park is. In Diagram II, the demand for trips to a park has the choke price $a(s,\mathbf{z})$ where s is the size of the park and $\mathbf{z}$ is a vector of socioeconomic characteristics of the family.

While parks are often areas for gatherings by several families, this model looks only at the optimal placement of parks for recreation by single families. If gatherings by several families occur more often among families that are spatially close, then the benefits from the gatherings of several families would influence the optimal spatial distribution of the parks, but the model does not examine this facet.

The families' demand curve for trips to a park reflects all the trips the family ever takes to the closest park. Of course, not every trip to the park is the same since the time of the year and the activities performed make the trips a little different. The generic family demand for trips to a park is an agglomeration of these different types of trips to a park.

An assumption on the choke price function is that $\partial a(s,\mathbf{z})/\partial s > 0$. Park size induces parallel shifts up in the demand for trips to a park. The sign of $\partial a(s,\mathbf{z})/\partial z_i$ depends on the characteristic z_i. Specific socioeconomic characteristics mentioned later are income, education and population, and for these examples $\partial a(s,\mathbf{z})/\partial z_i > 0$. Education is believed to shift out demand for trips to a park since a more educated populous better understands the health benefits of recreation. Population is believed to shift out the demand for trips to a park since each family has more people than before.

Although all families within the zone of a park go to that park, there is assumed no congestion at the park reducing the benefit from a trip to the park. Since smaller parks have smaller zones around them, fewer families visit them making the relationship between congestion at a park and the size of the park uncertain. The cost per mile of travel to a park is assumed a constant. For neighborhood parks this assumption is the most reasonable. For larger parks that every household in the city visits, there is a greater likelihood that there is some delay on the highway or only an indirect route to the park raising the per mile travel cost for households living further away from the park.

If a person lives at a distance x^0 from the park, the cost of making a *round* trip visit to a park is $2kx^0$ where k is the constant cost per mile traveled. The triangle represented by the area above the cost per trip line but below the demand curve is the net benefit to a person living x^0 from a park. The net benefit is represented mathematically by $\frac{(a(s, z) - 2kx^0)^2}{2b}$, where b the slope of the demand curve. The distance x^*, the distance marking the boundary of the zone around the park, is defined by $a(s, z) = 2kx^*$ because for all distances $x > x^*$ the cost of a trip exceeds the benefit of a trip. Since the net benefits of a trip is negative for families living $x > x^*$ from the park, those families do not visit the park.

While the slope of the demand curve b represents the preference for trips to the park, the slope is also able to represent the population density of the city. The higher the population density of the city the more flat is the slope of the demand curve for trips to a park since from each location around the park more trips are taken to the park. The demand curves for trips to a park for every family at a location are horizontally summed resulting in a flatter representative demand curve for trips to a park at that location.

If there are individuals in the city with no net benefits from parks, then there should be no overlap in the park zones. Although some individuals between the parks gain because they are closer to a park when zones overlap, other people were already receiving those benefits before the parks were brought closer together. The result is no change in the overall net benefit to the public. However, the people not between the parks only lose when the parks are brought together. Therefore, the change in overall net benefits is negative if the park zones overlap.

If the parks are far enough apart that the zones are significantly separated, those families between the parks but now outside the zones lose benefits from being farther from the parks, but those not between the parks brought into the zones gain exactly the benefits lost by those formerly within the zones. The result of spacing the parks farther apart is no change in aggregate net benefits.

The aggregate net benefits from a *single* park is

$$\int_{0}^{a(s,z)/2k} \frac{2(a(s,z)-2kx)^2}{2b}\,dx = \frac{1}{b}\int_{0}^{a(s,z)/2k} (a(s,z)-2kx)^2 dx \qquad (1)$$

because the net benefits from everyone living the distance $x^* = (a(s,z)/2k)$ from either side of the park are summed.

Suppose the city planner is trying to find the amount of land, number and size of the parks in the city to maximize the overall net benefits to the public. The cost of a park includes the purchase of the land for the park and the cost of maintaining the park. Without any maintenance cost, the cost of parks is $p(A)A$, the cost of buying land for parks at the price

$p(A)$. The maintenance cost increases the cost of parks to $\gamma p(A)A$, with $\gamma > 1$. Maintenance cost is greater if the land for parks is costly to purchase because the city spends more on maintenance of the city's expensive assets.

The net benefits maximization problem for the planner is:

$$\max_{A,n} \quad \frac{n}{b} \int_{0}^{a(s,z)/2k} (a(s,z) - 2kx)^2 dx - \gamma p(A)A \quad \text{s.t.} \quad 2x^* n = \frac{a(s,z)n}{k} \leq L \qquad (2)$$

where the constraint is meant to ensure that the park zones do not overlap. The constraint says that sum of the lengths of all the zones around the parks is less than the length of the city.

However, for the rest of the paper, the constraint is assumed not to bind. If the constraint binds, the overall net benefits to the public are greater than if the constraint does not bind. The reason is that, since the zones overlap, some families willing to travel a greater distance to a park in fact only travel a shorter distance. Although overlap of the zones is not good if unnecessary, the necessary overlap of the zones means more benefits to the public than having no overlap of the zones for the same number of parks. If a necessary overlap of zones occurs from increasing the number of parks, there are even more net benefits resulting from the increase in the number of parks. By assuming the constraint does not bind, the comparative statics possibly suggest creating fewer parks than the true optimum. Nonetheless, for spatially large cities with expensive land, assuming the constraint does not bind is quite reasonable.

The net benefits maximization problem is simplified by integrating the quadratic net benefits term:

$$\max_{A,n} \quad \frac{10n\, a(s,z)^3}{24bk} - \gamma p(A)A \qquad (3)$$

Solving the maximization problem (2), the first order conditions are:

$$\frac{10a(s^*, z)^2}{8bk}\frac{\partial a(s^*, z)}{\partial s} - \gamma \frac{\partial p(A^*)}{\partial A} A^* - \gamma p(A^*) = 0 \qquad (4)$$

$$\frac{10a(s^*, z)^3}{24bk} - \frac{\partial a(s^*, z)}{\partial s}\frac{10a(s^*, z)^2 A^*}{8bkn^*} = \frac{a(s^*, z)}{3} - \frac{\partial a(s^*, z)}{\partial s}\frac{A^*}{n^*} = 0 \qquad (5)$$

Equation (3i) determines the optimal amount of land for parks, A^*. If k, b, or γ increase, less land is made into parks. If the cost of travel is too high, no individual receives a positive net benefit from going to the park, and there is no reason to make parks land. If the demand for trips to the park is very steep, the net benefits from parks fall off very quickly, and again there is no reason to make a lot of land for parks. If maintenance costs are high, less parks land is purchased because the cost of maintaining the land is too much.

The optimal amount of land in parks is at the intersection of the marginal benefit and marginal cost of land in parks. The marginal benefit of more land in parks, $\frac{10a(s, z)^2}{8bk}\frac{\partial a(s, z)}{\partial s}$, is from the increase in size of every park by the same amount, so that each trip to a park yields higher net benefits. The marginal cost of more land in parks, $\gamma \frac{\partial p(A^*)}{\partial A} A^* + \gamma p$, is the market price of land, the extra price paid for all the land in parks because more land is purchased, and the cost of the maintenance of the new parks land.

Diagram III and IV illustrate the marginal benefit and marginal cost curves for the optimal amount of land in parks. The marginal cost curves are the same in both diagrams. The marginal benefit curve in Diagram III assumes that $\frac{\partial a(s,z)}{\partial s}$ is declining, or that $\frac{\partial^2 a(s,z)}{\partial s^2} < 0$, and the marginal benefit curve in Diagram IV assumes that $\frac{\partial a(s,z)}{\partial s}$ is constant, or that

$\frac{\partial^2 a(s,z)}{\partial s^2} = 0$. Diagram III is more realistic since, after a certain size, a bigger park does not increase any longer the net benefits from a trip to the park. However, for the comparative statics done later, the assumption that $\frac{\partial a(s,z)}{\partial s}$ is constant is very handy, and Diagram IV shows the implication of this assumption for the optimal amount of land in parks, A** > A*.

Equation (3ii) determines the optimal number of parks, n^*. The parameters k, b, and γ drop from Equation (3ii), and consequently do not influence the optimal number of parks. The optimal number of parks is a tradeoff between the gain of the net benefits from making another zone and the loss of net benefits from making all the zones smaller. The land for parks is made into smaller parks until the marginal benefit the public receives from having a park "next door", $\frac{a(s,z)}{3}$, equals the marginal cost from making all the parks a little smaller, $\frac{\partial a(s,z)}{\partial s}\frac{A}{n}$.

Diagram V and VI illustrate the marginal benefit and marginal cost curves for the optimal number of parks. The marginal benefit curves are the same in both diagrams. The marginal cost curve in Diagram V assumes that $\frac{\partial a(s,z)}{\partial s}$ is declining, or that $\frac{\partial^2 a(s,z)}{\partial s^2} < 0$, and the marginal cost curve in Diagram VI assumes that $\frac{\partial a(s,z)}{\partial s}$ is constant, or that $\frac{\partial^2 a(s,z)}{\partial s^2} = 0$. Since the assumption that $\frac{\partial a(s,z)}{\partial s}$ is a constant has both $n^{**} > n^*$ and $A^{**} > A^*$, the optimal size of the parks may not be influenced by the assumption on $\frac{\partial^2 a(s,z)}{\partial s^2} = 0$, i.e.

$A^{**}/n^{**} = s^{**} \cong s^* = A^*/n^*$.

The signs of the comparative statics are unambiguous if several additional assumptions are made. The three assumptions are that $\frac{\partial^2 a(s,z)}{\partial z_i \partial s} = 0$, $\frac{\partial^2 p(A)}{\partial A^2} = 0$, and $\frac{\partial^2 a(s,z)}{\partial s^2} = 0$. Since the sign of the cross partials might be argued either way, assuming the cross partial is zero seems plausible. The second order sensitivity of the market price of land to the amount of parks land is about zero since the amount of parks land is often a small proportion of the total area of the city. The second order sensitivity of the choke price of demand for trips to the park size is probably negative since park size likely has a weaker influence on demand for trips if the park is already big. While the examination of the first order conditions investigated the importance of the assumption on $\frac{\partial^2 a(s,z)}{\partial s^2}$, the comparative statics assume that $\frac{\partial^2 a(s,z)}{\partial s^2} = 0$ to simplify exposition.

The comparative statics are summarized in Table I. The example socioeconomic characteristics, income, education and income generate parallel shifts outward in the demand for trips to a park, i.e. $\frac{\partial a(s,z)}{\partial z_i} > 0$.

The comparative statics suggest that the total amount of land in parks is positively influenced by the income, education and the population of the city. This finding makes sense since a higher demand for trips to parks should raise the amount of land for parks. Also, a quickly declining demand for trips, a high cost of travel, and high maintenance costs negatively influence the amount of land for parks. Again, this finding makes sense since low net benefits from trips to the park and the high costs of maintenance of the parks should dissuade planners from making more land for parks.

The number of parks is positively influenced by the income, education and the population of the city. Whenever a new park is made, the major increase in net benefits goes to the families lucky enough to have the new park placed directly next to them. If all families have more income, education and people, the increase in net benefits to those families directly next to the new park is even greater. However, the net benefits lost by other families because all other parks decrease in size are also greater if all families have more income, education and people. The comparative statics say that the larger gain to the families next to the new park exceeds the larger loss to the families next to the older parks.

The number of parks is negatively influenced by quickly declining demand, a high cost of travel, and high maintenance costs. A more quickly declining demand and a higher cost of travel result in lower net benefits from trips to the park for all families. In this case, the comparative statics say that the lower gain for the families next to a new park is less than the lower loss for the households next to the older parks. This comparative static result is consistent with the comparative static result of the influence of income, education and population on the number of parks.

The size of the parks is negatively influenced by the income, education and the population of the city. Since both the amount of land for parks and the number of parks rise with income, education and population, the comparative statics say that the optimal number of parks rise more than the optimal amount of land for parks. The more that demand for trips to a park shifts out the more that the land for parks should be spread into small bits throughout the city rather than placed in big clumps.

Perhaps, surprisingly, the size of the parks is not influenced by how quickly demand declines, the cost of travel, or the maintenance cost of parks. Both the amount of land for parks and the number of parks fall in a way that they exactly offset each other, and the size of the parks is unchanged. Since the quickness of the decline in demand, the cost of travel, and the maintenance cost of the parks are equally influential on a new park and the older parks, the influences cancel each other, and there is no effect on the optimal size of the parks. Only demand shifters influence the optimal size of parks.

Proposition: If $\frac{\partial a(\cdot)}{\partial z_i} > 0$, where z_i is a socioeconomic characteristic of a city like income or education, then an increase in z_i makes the optimal amount of land increase, the optimal number of parks increase, and the optimal size of the parks decrease. If an increase in z_i shifts out the demand for trips to a park, the net benefits from a new park increase, and the land for parks is optimally spread in small pieces throughout the city.

Although the assumption that families are homogeneous and uniformly located appears to be a serious shortcoming of the model, heterogeneity is in fact easily allowed for in the model by dividing the city into smaller geographic units, i.e. neighborhoods, with households within each piece homogeneous. The model predicts that the optimal amount of land, number of parks, and the size of parks vary across the neighborhoods of the city because of the differences in the socioeconomic characteristics and the density of the households.

The demand for parks has been simplified by not considering the influence of congestion at the parks or the presence of substitutes for parks. Congestion at the parks suggests that parks should be larger, but the loss of benefits from the congestion needs to measurable to know exactly how much larger to make each park. Substitutes for parks also have a potentially strong influence on the demand for trips to parks. Households substitute away from parks by buying houses with larger backyards or purchasing a membership at a country club. Substitutes for trips to the park shift the demand for trips to the park. Along with the socioeconomic characteristics of the households, substitutes for the parks are potentially important shifters of the demand for trips to parks.

The remaining sections investigate if the comparative static predictions for the optimal amount of land, number, and size of parks hold up empirically.

III. Data

The data on the parks, including the amount of land in parks, the median size of the parks, and the number of parks, are from the Microsoft software *MS Streets and Trips 2003*. The maps in *MS Streets and Trips 2003* come from Navigation Technologies. The accuracy of the maps is confirmed by having employees drive the roads everywhere in the US to update street names. The points of interest in the maps including parks, schools, cemeteries, and golf courses are purchased by Navigation Technologies from a database provider InfoUSA. The points of interest data are only purchased for major metropolitan areas and tourist destinations explaining why there are no parks shown in the maps of the software for many cities. The address information for the points of interest from InfoUSA is updated continuously throughout the year. The cities where the data on parks was collect are shown in the Appendix.

The area of each park in a city is measured, and the areas are summed to get the amount of land in parks. For rectangular parks, the length and width are measured to get the area of the park. For irregular shaped parks, the park is divided into rectangular bits and the area of the rectangular bits is summed. Other spatial information about the parks, like the distance to downtown and the extent that parks clump together, is also collected using the *MS Streets & Trips 2003* software. Distance to downtown is found by measuring a straight line from the edge of the park nearest the downtown to the downtown. The location of downtown is identified by entering the name of the city into *MS Streets & Trips 2003*; the location of the city name is deemed the downtown. The downtown found by this method is often near the city hall. The extent that parks clump together is found by measuring the distance between the nearest parks from closest edge to closest edge. The median of those distances is inverted to generate the variable of the extent that the parks clump together.

The data on the cities from 1970, 1980, 1990, and 2000 come from the Bureau of the Census. Most of the year 2000 city data is from the *County and City Data Book (CCD)*, a Census Bureau publication that provides data for a cross section of counties, metropolitan areas, and cities. "Cities" in the CCD are incorporated places that have a population of 25,000 or more. Additional city data for the year 2000 not available from the CCD like median income is found at an online source called Ersys.com. City data from 1970, 1980 and 1990 is also collected by Census, but the data are retrieved from CDs produced by a company called Geolytics. The CDs from Geolytics have data on a wide range of subjects at a more geographic specific level than city, and there is no need to supplement those data with other sources.

The data for the zip codes, only available for the year 2000, is taken from the American FactFinder found at the Bureau of the Census web site. The land areas of the zip codes, referred to as zip code tabulation areas (ZCTAs), are taken from the US Gazetteer File for 2000[1].

IV. Empirical Specification and Estimation

The comparative statics predictions of the relationship between the characteristics of a city and the amount and spatial distribution of parks are examined empirically. Unfortunately, the comparative statics predictions are from a model having assumptions not entirely consistent with the real world. In particular, the assumption that people have identical socioeconomic characteristics and are uniformly distributed in a city is not reasonable for all cities. Many cities have neighborhoods that are homogenous within but are significantly different from the other neighborhoods of the city. In addition to collecting socioeconomic data at the city level, socioeconomic data is also collected at the zip code level. The regions defined by zip codes contain more homogeneous groups of people than cities since the populations are usually smaller and the regions are more spatially compact if the zip code is densely populated.

Although the regions defined by zip codes may be more homogeneous, the government agency creating and spatially distributing parks may operate at a different spatial level, e.g. the city. However, the hypothesis is that the government agency recognizes the different regions of homogenous groups of people and spatially distributes parks accordingly. In other words, whether there is a government agency for each zip code of homogenous group of people or a single government agency for the entire city, the amount and spatial distribution of the parks in the city is the same.

The empirical model does not seek to identify all possible forces influencing the spatial distribution of parks. Planners creating and spatially distributing parks may react more strongly to city budget constraints, city growth projections, the irreversibility of park creation, or public interest groups when deciding the amount of land, number and size of parks in a city. The purpose of the empirical model is rather to learn if particular characteristics of a city are related to the amount and spatial distribution of parks in the way suggested by the model of net benefits maximization.

While the goal of the empirical model is fairly modest, the specification of the empirical model remains difficult since the coefficient estimates on the variables of interest are most likely distorted by omitted variable bias. For example, the relationship between the median income of a city in the year 2000 and the amount of land in parks is of interest. However, the median income of the city in the year 2000 is correlated with the median income of the city from 1990, and the median income of the city from 1990 certainly does influence the current amount of land in parks in the city. Fortunately, at the city level, for many of the variables of interest, historical values are known. The age and race of the population are strongly correlated with the variables of interest, but these city characteristics are omitted since these variables are highly collinear with median income and there is no compelling reason why these city characteristics should influence the amount and spatial distribution of parks.

The variables in the empirical models are:

lnpkarea, the natural log of the total area of parks in a city (sq. miles)
pknum, the number of parks in a city
lnmpksize, the natural log of the median park size in a city (sq. miles)
medinc, the first principal component of the median household income of a city for the years 1980, 1990, and 2000 (dollars)
College, the first principal component of the population holding a four year college degree in a city for the years 1970, 1980, and 1990
Pop, the first principal component of the population in a city for the years 1970, 1980, 1990 and 2000 (people)
trvltime, the commuting time to work deflated by an index of the distance to the CBD of the metropolitan area in the year 1990[2] (minutes)
Y2Kden, constructed variable that is the population of the city divided by the area of the city in the year 2000 (people/sq. mile)
landprice, the average monthly payment on a mortgaged house for a city in the year 1990 (dollars)
incsq, constructed variable that is the square of *medinc*
CtyAge, the age of a city since its incorporation (years)
ZipAge, a weighted (proportion of total homes) average of the age of the homes in the zip code (years)

The natural log is taken of the total area in parks and the median size of parks because the transformation increases the variation in the skewed component of those variables while not affecting the relevant ordering. The explanatory variables *trvltime* and *Y2Kden* are meant to represent the cost of travel in a city. The cost of travel along the highways and the main roads of a city are represented by *trvltime* since it is assumed that people take the highways and main roads on the way to work. The cost of travel along smaller streets is represented by the *Y2Kden* variable since a more densely populated city is likely to have more crowded streets and more intersections. Since population density also flattens the demand curve for trips to a park, the opposite effect of travel cost on the spatial distribution of parks is possible.

Although *landprice* is the monthly mortgage payments on a house, the housing cost is strongly correlated with the price of land. Indeed, if the land already has a residence on it, then monthly housing payments even better represents the cost to acquire the land for parks. However, to the extent that homes are larger and of better quality, the monthly housing payments reflect the value of the house rather than the value of the land that the house rests upon. Another reason for using the monthly mortgage payments on a house is that these payments are less linked to the amount of land in parks than the actual land price. Since mortgage payments reflect financing costs and housing quality in addition to the location value of a home, the mortgage payments make a good instrument for the price of land.

The empirical model has a familiar linear form. However, the *incsq* explanatory variable suggests nonlinearity exists in the relationship with income. The nonlinear relationship with income is probably because, at the low income range, people demand parks because they live in apartments and small houses. However, at the upper income range, people substitute towards large backyards and country clubs instead of parks. The substitution away from parks explains the nonlinearity in income that has a positive influence in the lower range but turns into a negative influence at the higher range.

$$\begin{aligned} lnpkarea_i &= \alpha_1 + \beta_{11} medinc_i + \beta_{21} College_i \\ &+ \beta_{31} Pop_i + \beta_{41} trvltime_i + \beta_{51} Y2Kden_i + \beta_{61} landprice_i \\ &+ \beta_{71} incsq_i + \beta_{81} CtyAge_i + \varepsilon_{1i} \end{aligned} \tag{6}$$

$$\begin{aligned} pknum_i &= \alpha_2 + \beta_{12} medinc_i + \beta_{22} College_i + \beta_{32} Pop_i \\ &+ \beta_{42} trvltime_i + \beta_{52} Y2Kden_i + \beta_{62} landprice_i + \beta_{72} incsq_i \\ &+ \beta_{82} CtyAge_i + \varepsilon_{2i} \end{aligned} \tag{7}$$

$$\begin{aligned} lnmpksize_i &= \alpha_3 + \beta_{13} medinc_i + \beta_{23} College_i \\ &+ \beta_{33} Pop_i + \beta_{43} trvltime_i + \beta_{53} Y2Kden_i + \beta_{63} incsq_i \\ &+ \beta_{73} CtyAge_i + \varepsilon_{3i} \end{aligned} \tag{8}$$

The main problem plaguing the empirical model estimation is collinearity among the explanatory variables. The collinearity lowers the *t*-statistics of the collinear variables making the separate influence of each collinear variable on the dependent variable difficult to identify. Since there is strong collinearity over time for the population, median income, and education variables, principal component analysis is applied to each time series of these variables to get a single composite index for each of the variables. The first principal component is the single composite index for each of these time series of variables. Each composite index represents the cumulative influence of the variables over time on the amount and spatial distribution of parks. However, Table II shows that collinearity lingers between the median income, education, and the price of land variables.

Further, collinearity limits more general specifications because other potentially influential variables like the age of the population or the percentage minority of the population are collinear with weighted median income and land prices. Accordingly, when the age of population is included in the empirical model, the variable only dilutes the influence that median income and land price have on the dependent variable making the results more ambiguous. Additionally, there is no compelling explanation of how these city characteristics should influence the amount and spatial distribution of parks.

Other problems estimating the empirical model are measurement error in the dependent variable and heteroskedasticity. Although the measurement error in the dependent variables is a nuisance, there is not much to do except to note that caution and consistency were applied in the dependent variable measurement.

The Breusch-Pagan test for heteroskedasticity finds heteroskedasticity in the park number and the median park size equations at the ten percent level. In the park number equation, the error variance is significantly influenced by the income, income squared and city age variables. In the median park size equation, the error variance is significantly influenced by population density. Since the nature of the pattern of the heteroskedasticity is not discernable, White's heteroskedasticity standard error correction is applied in the estimation of the equations. The results with White's correction are very similar to the results from regular OLS or more advanced procedures like robust regression where each observation is given different weights based off an iterative process.

V. Results

Additional variables were collected to examine how parks are spatially distributed within cities. Also, information on the number of other green spaces like schools, golf courses, and cemeteries was collected to learn about the substitutes for parks and other investments in public goods in the cities. The description of these variables are:

dwntwn, the median distance of the parks from downtown measured from the closest park edge (miles)
clump, constructed variable that is the inverse of the median distance to the next closest park measured from park edge to park edge (miles)
props, constructed variable that is the area in small parks divided by the total area in parks--a park is small if it is less than six and a half acres
propm, constructed variable that is the area in medium size parks divided by the total area in parks--a park is medium sized if it is between six and a half and sixty-five acres
propl, constructed variable that is the area in large parks divided by the total area in parks--a park is large if it is greater than sixty five acres
golf, number of golf courses in the city
schools, number of schools in the city
cemetery, number of cemeteries in the city

Summary statistics are in Table IV. The negative numbers for the average *lnpkarea* and *lnmpksize* are not alarming because most cities do not have more than one square mile of land in parks, and the median size of a park is certainly not more than one square mile. The reason there are less than five parks sometimes in very populous cities is that those cities usually have small areas with high land prices because they are located close to extremely populous cities like Chicago. Cities like those are certainly the exception in the sample. Although a high proportion of the total land in parks in a city is from large parks, the only reason is that large parks have a lot of land since most of the cities have almost entirely small or medium sized parks. Further, although the maximum values of median income, population, and city area in the year 2000 are quite large, those observations are exceptions with most values around the averages. The cities are fairly old since on average they have been incorporated for ninety-seven years.

The amount of land in parks and the median size of the parks are skewed before the natural log transformation spreads the variation in those variables. The remaining economic and demographic variables like median income, land price, and population have adequate variation for determining their influence on the amount of land, number and size of parks.

Recall that a key assumption in the theoretical model is that individuals are homogeneous and uniformly located within a city. In other words, all individuals have the same race, age, income, and costs of travel, etc. in the city. The median population is 60,270 and the median area is 22.5 square miles for the sample of cities in the year 2000. These cities are populous and spatially large enough that the assumption of homogeneity throughout the city is questionable. If the diverse array of individuals is evenly mixed, the city still fits the criterion of homogeneity. However, large cities with a diversity of people usually have ethnic neighborhoods where people of similar characteristics cluster together. Since the assumption of homogeneous individuals is dubious, the parks and socioeconomic data for the zip codes in the sample of cities were collected. The median population is 28,066 and the median area is 13 square miles for the zip codes in the sample of cities for the year 2000. Since the zip codes have on average a lower population and area than the cities, the assumption of homogeneous individuals in each zip code is more believable. However, whether a city planner contemplates the heterogeneity of a city before deciding on the amount and spatial distribution of parks is debatable.

Another key assumption is that the land for parks is split into parks according to the rule $A = ns$. The implication of the assumption is that all parks are the same size. The data indicate that parks are certainly not all the same size in cities although a few cities have only small or medium sized parks. The assumption that all parks are the same size makes sense in a city completely homogeneous because symmetry is a natural result if everyone is the same. While heterogeneity in the city is one possible explanation for the different sized parks, another good reason is that parks of different sizes offer different benefits, and the city planners make parks of different sizes to make the full range of benefits from parks available to the public. Although there is a range of park sizes, the empiric model examines if the median park size is influenced by the socioeconomic characteristics as predicted by comparative statics of the theory.

In Table V are the correlations among the variables representing the spatial distribution of parks and potential substitutes for parks like schools, cemeteries, and golf courses. These correlations are a snapshot of the main patterns of the spatial distribution of parks in the sample of cities.

There is a strong positive correlation between the number of parks and the total land in parks. This suggests that cities with a lot of land in parks usually have many medium and large sized parks. There is also a strong positive correlation between median park size and the total land for parks. This suggests that cities that have more land for parks tend to have larger parks. In sum, the pattern is that cities without much land for parks have a few fairly small parks, and cities with a lot of land for parks have many medium to large sized parks.

The correlation between the distance of the parks from downtown and total area in parks is strongly positive. Cities with a lot of land in parks have the parks further from downtown. The correlation between the distance of the parks from downtown and median park size is positive. Also, the correlation between the distance of the parks from downtown and proportion of parks land in large parks is positive. Both these correlations suggest that parks further from downtown are larger. In sum, cities without much land in parks have a couple small parks near downtown, and cities with a lot of land in parks have numerous medium to large parks located far from downtown.

The correlation between how much parks clump together and the number of parks is strongly positive. Also, cities with a lot of land in parks tend to have parks clumped. These correlation suggest that cities without much land in parks have a couple parks fairly spread out from each other near downtown while cities with a lot of land have numerous medium sized park located close to each other far from downtown. The correlations do not suggest any relationship between park size and the clumping of the parks. The zones of visitation to parks from the theory of this paper suggest that medium and large sized park should be far apart from each other, but the small parks may be either close or far from each other. The finding of zero correlation between park size and clumping of parks agrees with the theory in this paper of the zones of visitation to parks.

The number of schools is positively correlated with both the number of parks and the amount of land in parks. This suggests that cities that make significant investments in parks also invest heavily in schools. City planners appear to treat parks and schools like complements rather than substitutes. In general, cities prefer to have a portfolio of public goods to offer its residents rather than only one type. As for the number of golf courses and cemeteries, there is largely no relationship between these other greeneries and the number of parks. The finding of no relationship is not all that surprising since cemeteries and golf courses are usually privately run enterprises offering greenery only as a spillover from the main enterprise.

The correlations suggest that there are two main patterns to the layout of parks. Cities without much land in parks have a couple small parks separated from each other near downtown, and cities with a lot of land in parks have many medium sized parks close to each other but far from downtown. Cities without much land in parks usually have at least couple parks near the downtown, and cities with a lot of land in parks usually cluster them away from downtown in residential neighborhoods.

Tables VI and VII have the results from the estimation of the empirical model. The results in Table VI are at the spatial scale of city while the results in Table VII are at the spatial scale of zip code. The data at the smaller spatial scale of zip codes better matches the assumption in the theoretical model of homogeneity of socioeconomic characteristics throughout an area. The empirical model for the zip code data has a slightly different specification than the model for the city data. In particular, the variable *trvltime* is missing in the zip code model since zip codes do not have downtowns. Also, the variable *ZipAge* replaces the variable *CtyAge* since the age of the city containing the zip code is not necessarily the same as the age of the structures in the zip code. Estimation of the empirical model is done with the Intercooled STATA 8.0 software package[3].

Most of the variables are significant in the equation for the total amount of land in parks. The significant variables in Tables VI and VII for the total amount of land in parks are largely the same except that the variable for college education is found significant with the zip code data. Larger variation in the college education variable for the zip code data is the reason the variable for college education is significant in Table VII. Tables VI and VII both suggest that an inverse U-curve relationship between income and the amount of land for parks exists. At low levels of income, income has a positive influence on the amount of land in parks. While at high levels of income, income has a negative influence on the amount of land in parks. At high levels of income there is less demand for land in parks since city residents purchase houses with large backyards or memberships at a country club.

In Table VI, the significant variables at the 5% level of the estimated equation for the amount of land in parks are median income, population, population density, land price and median income squared. For the city data, the variation in the amount of land for parks is most strongly explained by population and population density since the coefficient estimates of these variables have the highest t-statistics. The strong explanatory power of population makes sense since many states have laws requiring a certain amount of land for parks be created for a certain amount of people. The strong explanatory power of population density suggests that planners consider the travel costs for recreation at the parks at the city level before deciding how much land to have for parks.

In Table VII, the significant variables at the 5% level of the estimated equation for the amount of land in parks are median income, percent college educated, population, population density, land price and median income squared. For the zip code data, the variation in the amount of land for parks is most strongly explained by the land price, population, and median income. Land price and median income probably have stronger explanatory power since the values better reflect the values of these variables throughout the zip code since zip codes are more homogeneous. For the city data, land price and median income are averages across a heterogeneous city diluting the estimated influence of these variables on the amount of land in parks. Note that population density has much weaker explanatory power than land price does for the zip code data. One explanation is that for the city data the variation in population density is actually important variation in the land price that the land price variable at the city level is not adequately representing.

For the equation of the total amount of land in parks, all the signs of the coefficient estimates shown in Tables VI and VII match the predicted signs of the comparative statics. Recall that the predicted signs of the comparative statics are dependent on three assumptions. In particular, the least credible assumption is that the second order sensitivity of the choke price for the demand for trips to the size of the park is zero. However, the predicted signs of the comparative statics for the amount of land in parks are not sensitive to this assumption. In other words, since the predicted signs of the comparative statics for the amount of land in parks are the most robust, the expectation is that the coefficient estimates for the amount of land in parks would most closely match the signs of the comparative statics for the amount of land in parks.

While most variables in the equation for the total amount of land in parks are significant, the only significant variable in Tables VI and VII for the number of parks equation is population. In Table VI there are no other variables significant at the 10% level for the number of parks equation, and in Table VII there are no other variables having *t*-statistics for their coefficient estimates greater than one for the number of parks equation. Clearly, no inverse U-curve relationship between income and the number of parks emerges from the results in Tables VI and

VII. The strong explanatory power of population is not that surprising since planners are often concerned about equitably providing land for parks in a city. Since population density is held constant in the equation for the number of parks, an increase in population implies more people in a city with a larger area. In order to have the land for parks equitably provided in a city with a greater population across a larger area, the number of parks has to increase.

The sign of the coefficient estimate for population in the number of parks equation matches the predicted sign from the comparative statics. However, many of the variables thought to influence the number of parks from the theory are shown to have no influence on the number of the parks in the empirics. The signs from the comparative statics for the number of parks are sensitive to the least credible assumption that the second order sensitivity of the choke price for the demand for trips to the size of the park is zero. If the assumption is relaxed to allow the second order sensitivity of the choke price for the demand for trips to the size of the park to be negative, the predicted signs of the comparative statics for income and education become ambiguous. Ambiguous predicted signs from the comparative statics fit with the poor explanatory power that income and education have on the number of parks observed in the empirics.

Other potential explanations for the poor explanatory power of most variables in the number of parks equation are factors not incorporated into the theory of this paper. One factor is that the presence of parks in a city slows traffic in the city. If traffic is correlated with other socioeconomic variables of the city, city planners' concern about traffic congestion may make the influence of the socioeconomic characteristics on the number of parks differ from the comparative statics predictions. Another factor is the large transition cost of spatially rearranging land in parks. The number of parks observed in the city or zip code data may have been optimal at the time the parks were built, but the demographics of the cities change over time until that number of parks is no longer optimal. However, the large transition cost of spatially rearranging the land in parks prevents the city planner from rearranging the land in parks towards the new optimal number of parks.

Tables VI and VII for the park size equation have the same significant variables except that Table VI has the significant variable *trvltime* not present in the empirical model for the zip code data. The significant variables of the park size equation that Tables VI and VII share are income and the land price. Although there is some indication that an inverse U-curve relationship between income and park size exists, the relationship is weaker than the inverse U-curve relationship found between income and the total area of parks land. Since there is a positive correlation between park size and the total area of parks along with no correlation between park size and the number of parks, the weak inverse U-curve relationship observed between park size and income may simply be a diluted form of the inverse U-curve relationship between the total area of parks land and income.

For the park size equation in Tables VI and VII, the variables income, land price and *trvltime* are significant at the 5% level. The explanatory power of income on park size is in fact stronger in Table VI than in Table VII. Since large parks are likely to be used by residents across a city, perhaps planners consider the city level income more than the local neighborhood income before choosing the park sizes of an area. Nonetheless, the strong explanatory of income in Table VII suggests that planners are paying attention to the more localized income levels too. The explanatory power of land price on park size is equally strong in Tables VI and VII. Unexpectedly, the land price, or opportunity cost of land, influences the median park size. One explanation is that many parks are created from land that cannot function in more profitable uses such as the conversion of many abandoned train tracks into parks. While the planner creates parks from these random bits of land, the planner is unwilling to purchase land at a high price adjacent to these random bits to make the parks bigger. In this fashion, the opportunity cost of land influences the park size. Also unexpectedly, the variable *trvltime*, the commuting time to the CBD, influences the median park size. This finding may reflect the decision by city planners to make few large parks when commuting time is high since there is the concern that large parks slow traffic.

None of the signs of the coefficients for the significant variables in Tables VI and VII for the park size equation match the predicted signs from the comparative statics. Although the comparative statics sign for the influence of income on park size is negative, a definitively positive sign for the coefficient on income is found. The sensitivity of the comparative statics sign to the assumption that the second order sensitivity of the choke price for the demand for trips to the size of the park is zero may explain partly why the opposite sign for the coefficient is found. However, the violation of that assumption alone should only make the income variable insignificant rather than significantly positive. Another potentially violated assumption used to unambiguously sign the comparative statics is that the cross partial of the sensitivity of the choke price for the demand for trips to income and park size is zero. If income does in fact increase the desirability of making parks larger, the sign of the comparative statics for the influence of income on park size potentially turns positive. If all the assumptions to unambiguously sign the comparative statics are in fact correct, the mismatch of signs for the coefficient estimates and the comparative statics may be because of factors like traffic congestion or the dynamics of park creation not incorporated into the theoretical model.

The results from the amount of parks land equation are promising. Most variables in the equation are significant at the 5% level, and the R^2 of 0.465 suggests a reasonably strong fit of the equation. Also, the signs of the coefficient estimates match the predicted signs of the comparative statics suggesting that the theoretical model is capturing most of relevant factors determining the amount of land in parks. Further, the truth of the assumptions to sign the comparative statics for the amount of land in parks is strengthened.

However, for the equations of the number of parks and the park size, most variables are insignificant and the fit is not good. Although the basic socioeconomic variables of a city do well at explaining the amount of land in parks, there are other important influences on the number of parks and the size of parks not embodied in those basic socioeconomic variables.

Since the signs of the coefficient estimates and comparative statics rarely match for the park number and park size equations, one explanation is that either the signs of the coefficient estimates, the signs of the comparative statics, or both are wrong. Bias in the signs of the coefficient estimates may exist because other relevant variables such as for the travel costs in the city or the fixed costs of spatially rearranging parks are omitted. The signs of the comparative statics may be wrong either because the assumptions of the theoretical model to sign the comparative statics are incorrect or because other factors, such as the concern of planners about traffic congestion, not included in the theoretical model, but if included would change the signs of the comparative statics. Alternatively, all the signs of the coefficient estimates and comparative statics may be correct, but planners are not creating the numbers and sizes of parks to maximize the social benefits of the city residents. Possible explanations for why planners would create park numbers and sizes different from the socially optimal are political susceptibility to interest groups or incorrect information about the city residents.

VI. Conclusion

This paper develops a theory to explain how city characteristics influence the spatial distribution of parks for maximizing the net benefits to the public from recreation at the parks. Both the net benefits from the recreation and the cost of creating and maintaining the parks are explicitly incorporated into the theory. A few simple assumptions on the preferences for trips to a park and the market for land allow the comparative statics for a change in city characteristics on the amount of land in parks, the number of parks, and the size of parks to be unambiguously signed.

Empirics to examine if the city characteristics influence the spatial distribution of parks in the way suggested by the comparative statics have also been done. The empirical examination is done at two spatial levels, city and zip code, in an attempt to match the data to the assumptions of the theory in the best way possible. The results of the empirics at the two spatial levels are similar to each other.

The theory suggests that the amount of land, the number, and the size of the parks are all sensitive to shifters of demand for trips to a park. Outward shifts in demand for trips result in more land for parks, a higher number of parks, and smaller parks. Changes in city characteristics, such as the travel costs to the parks and the maintenance cost of parks, unrelated to the size of the parks influence the optimal amount of land and number of parks but have no influence on the size of the parks.

The results of the empirical analysis follow closely to the results from the theoretical model for the influence of the city characteristics on the amount of land in parks. In particular, the city characteristics of population, income, education, and land price are shown to have significant explanatory power for the amount of land in parks. These city characteristics influence the amount of land in parks in the way suggested by the theory. Most city characteristics have poor explanatory power of the number of parks except for population. Likewise, most city characteristics have poor explanatory power of the size of parks except for median income and land price, and these significant variables have different effects on the size of parks than predicted from the theoretical model.

The less encouraging results of the empirical analysis for the influence of city characteristics on the number and size of parks are probably because the assumptions on preferences for the demand for trips in the theoretical model are not completely correct. Unfortunately, there is little research examining the influence of spatial characteristics of public open space on the demand for recreation to better guide these assumptions. There is also the concern that the city characteristics change across time, and the current spatial distribution of parks is meant for a city with characteristics from many years ago because the redistribution of parks land is too costly. Nonetheless, the mismatch of the results of the empirics and theory suggests that in many cities a redistribution of land in parks is possible to increase the net benefits to the residents from recreation at the parks.

References

Forest Service and NOAA. “2005 National Survey on Recreation and the Environment: Planning for the Eighth National Recreation Survey” http://www.srs.fs.usda.gov/trends/Nsre/NSRE200562303.pdf (2003).

Gaussier, N. (2001). "The Spatial Foundations of Obnoxious Goods Location: The Garbage Dumps Case." Regional Studies **35**(7): 625-36.

Glaeser, E., S. J., et al. (1995). "Economic Growth in a Cross-Section of Cities." Journal of Monetary Economics **34**: 117-143.

Goldin, C. and K. L. (1998). Human Capital and Social Capital: The Rise of Secondary Schooling in America, 1910-1940, NBER Working Paper No. 6439.

Halstead, J. M., J. Whitcomb, et al. (1999). "Economic Insights into the Siting Problem: An Application of the Expected Utility Model." Agricultural and Resource Economics Review **28**(1): 65-75.

Haughwout, A. (2002). "Public Infrastructure Investments, Productivity and Welfare in Fixed Geographic Areas." Journal of Public Economics **83**(3): 405-28.

Hewko, J., K. E. Smoyer-Tomic, et al. (2002). "Measuring neighborhood spatial accessibility to urban amenities: does aggregation error matter?" Environment and Planning A **34**(7): 1185-1206.

Kemmerling, A. and A. Stephan (2002). "The Contribution of Local Public Infrastructure to Private Productivity and Its Political Economy: Evidence from a Panel of Large German Cities." Public Choice **113**(3-4): 403-24.

Lindsey, G., M. Maraj, et al. (2001). "Access, equity, and urban greenways: An exploratory investigation." Professional Geographer **53**(3): 332-346.

Land Trust Alliance. “LandVote 2003: American voters care about saving the special places in their communities.” http://www.lta.org/publicpolicy/statelocalnews.htm (May 24, 2004).

Lee, C. M. and M. Fujita (1997). "Efficient Configuration of a greenbelt: theoretical modelling of greenbelt amenity." Environment and Planning A **29**: 1999-2017.

Martin, R. (1999). "The new 'geographical turn' in economics: some critical reflections." Cambridge Journal of Economics **23**(1): 65-91.

Poterba, J. (1994). "Demographic Structure and the Political Economy of Public Education." NBER Working Paper No. 5677, July 1998.

Weinhold, D. and E. J. Reis (2001). "Model Evaluation and Causality Testing in Short Panels: The Case of Infrastructure Provision and Population Growth in the Brazilian Amazon." Journal of Regional Science **41**(4): 639-57.

Witten, K., D. Exeter, et al. (2003). "The quality of urban environments: Mapping variation in access to community resources." Urban Studies **40**(1): 161-177.

Yang, C. H. and M. Fujita (1983). "Urban spatial structure with open space." Environment and Planning A **15**: 67-84.

Diagram I

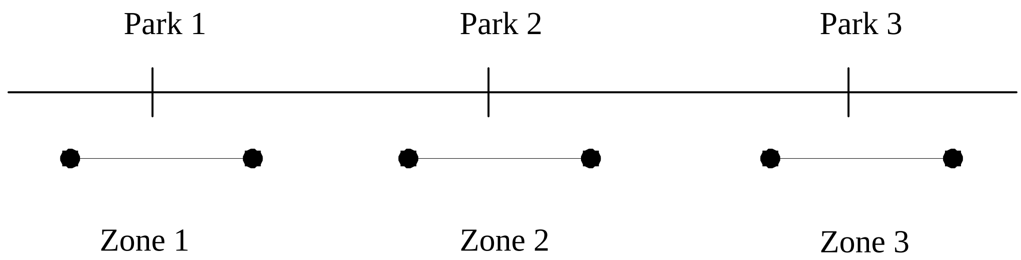


Diagram II

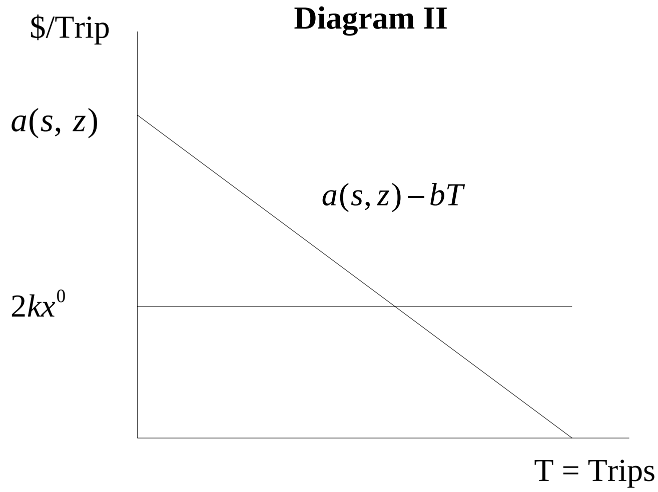


Diagram III

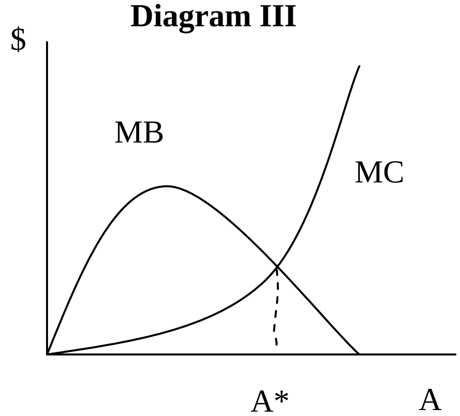
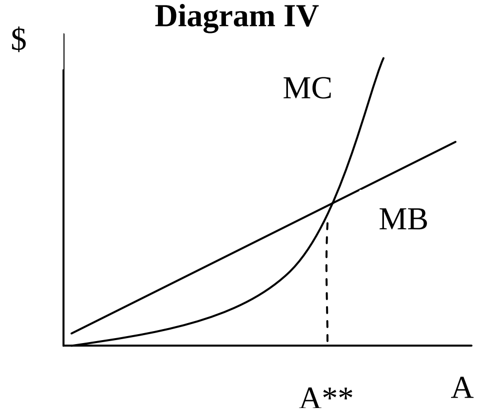
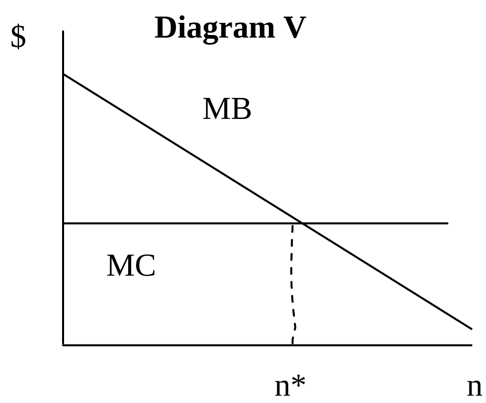


Diagram IV





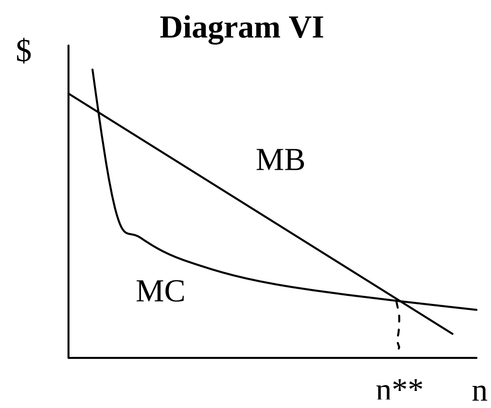


Table I: Predictions of the comparative statics

		Exogenous					
		Income	Education	Population	*b*	*k*	γ
Endogenous	Total land in parks: *A*	(+)	(+)	(+)	(-)	(-)	(-)
	Number of parks: *n*	(+)	(+)	(+)	(-)	(-)	(-)
	Park size: *s*	(-)	(-)	(-)	No Effect	No Effect	No Effect

Table II: Collinearity among explanatory variables

	medinc	*College*	*landprice*	*incsqr*
medinc	1			
College	0.5525	1		
landprice	0.6482	0.7109	1	
incsqr	0.9898	0.5766	0.6691	1

Table III: Summary statistics

Variable	Mean	Std. Dev.	Min	Max
lnpkarea	-0.25	1.20	-3.86	2.57
pknum	16.39	9.86	2	43
lnmpksize	-3.66	0.74	-5.52	-1.87
distdwntn	1.93	0.58	1.02	4.47
clump	2.19	0.85	0.5	4.25
props	0.059	0.13	0	1
propm	0.43	0.30	0	1
propl	0.50	0.33	0	0.98
Medinc2K	35756	8883	19127	60545
College2K	11.19	4.63	3.9	22.76
Pop2K	65773	25364	35420	147595
trvltime	18.51	2.43	13.4	24.32
Y2Kden	3040	1915	540	13850
Y2Karea	30.34	27.30	3.9	177
landprice	795	200	477	1497
CtyAge	97	42	34	197
schools	18.48	11.38	3	68
cemetery	1.83	1.85	0	9
golf	2.07	1.67	0	9
Number of Observations: 66				

Table IV: Spatial variable and park substitute correlations

	pknum	*lnpkarea*	*lnmpksize*	*dwntwntn*	*clump*	*props*	*propm*	*propl*	*schools*	*cemetery*	*golf*
pknum	1										
lnpkarea	0.542	1									
lnmpksize	-0.084	0.372	1								
dwntwntn	0.2	0.319	0.274	1							
clump	0.526	0.259	-0.01	-0.197	1						
props	-0.114	-0.543	-0.54	-0.213	0.089	1					
propm	-0.17	-0.604	-0.163	-0.212	-0.1	0.015	1				
propl	0.204	0.767	0.366	0.27	0.05	-0.41	-0.918	1			
schools	0.435	0.348	-0.054	0.266	0.048	-0.14	-0.236	0.27	1		
cemetery	0.184	0.192	-0.066	0.248	-0.095	-0.142	-0.198	0.237	0.610	1	
golf	0.179	0.081	-0.108	-0.013	-0.034	0.009	-0.116	0.101	0.313	0.128	1

Table V: Testing the theoretical model -- Cities

OLS with White's robust standard errors

Explanatory Variables	*lnpkarea*	*pknum*	*lnmpksize*
constant	1.66 (1.36)	30.78 (2.18)	0.583 (0.76)
medinc	0.428 (3.36)	1.79 (1.67)	0.333 (4.84)
College	0.145 (1.05)	0.719 (0.61)	-0.013 (0.17)
Pop	0.211 (4.00)	2.69 (4.59)	0.029 (0.85)
trvltime	-0.0281 (0.5)	-0.281 (-0.5)	-0.115 (-2.74)
Y2Kden	-0.00026 (-3.97)	-0.00084 (-1.63)	0.000034 (0.7)
landprice	-0.0026 (-2.02)	-0.00728 (-0.82)	-0.00245 (-3.5)
wincsq	-0.087 (-2.8)	-0.01 (-0.03)	-0.0376 (-1.69)
CtyAge	0.0063 (1.87)	-0.011 (-0.4)	-0.0018 (-0.97)
F-statistic	13.85	5.35	8.68
R-squared	0.465	0.277	0.353
t-statistics in parentheses Observations: 66			

Table VI: Testing the theoretical model -- Zip Codes

OLS with White's robust standard errors

Explanatory Variables	*lnpkarea*	*pknum*	*lnmpksize*
constant	-3.96 (-3.51)	0.598 (0.18)	-3.99 (-4.57)
medinc	0.00011 (3.49)	0.00054 (0.55)	0.000063 (2.62)
College	0.0349 (2.36)	0.035 (0.66)	-0.0021 (-0.18)
Pop	0.000025 (3.51)	0.00018 (5.33)	-0.000008 (-1.52)
Y2Kden	-0.0001 (-1.81)	-0.00004 (-0.22)	0.000031 (-0.7)
landprice	-0.0021 (-3.99)	-0.0016 (-0.71)	-0.0012 (-3.05)
wincsq	-6.42e-10 (-2.54)	-5.7e-11 (-0.007)	-3.25e-10 (-1.66)
ZipAge	0.006 (0.56)	-0.004 (-0.11)	-0.0053 (-0.61)
F-statistic	7.47	7.01	3.67
R-squared	0.245	0.234	0.137

t-statistics in parentheses
Observations: 169

Appendix

Alameda, CA	Apple Valley, MN
Arlington Heights, IL	Auburn, WA
Baytown, TX	Bedford, TX
Belleville, IL	Berwyn, IL
Blaine, MN	Bloomington, MN
Blue Springs, MO	Boca Raton, FL
Bolingbrook, IL	Boulder, CO
Boynton Beach, FL	Bradenton, FL
Bremerton, WA	Brooklyn Park, MN
Buffalo Grove, IL	Burnesville, MN
Cleveland Heights, OH	Concord, NC
Coral Gables, FL	Daytona Beach, FL
Dearborn Heights, MI	Dearborn, MI
East Lansing, MI	Edmond, OK
Edmonds, WA	Euless, TX
Florissant, MO	Fort Collins, CO
Galveston, TX	Gastonia, NC
Grapevine, TX	Greeley, CO
Gresham, OR	Hammond, IN
High Point, NC	Hillsboro, OR
Independence, MO	Kansas City, KS
Kenner, LA	Kent, WA
Kentwood, MI	Kokomo, IN
Leavenworth, KS	Lees Summit, MO
Longmount, CO	Loveland, CO
Midwest City, OK	Norman, OK
Novi, MI	Olathe, KS
Olympia, WA	Racine, WI
Rockhill, SC	Salem, OR
Santa Rosa, CA	Sante Fe,NM
St. Charles, MO	St. Peters, MO
Tigard, OR	Waukesha, WI
Wauwatosa, WI	West Allis, WI

Footnotes

[1] http://www.census.gov/tiger/tms/gazetteer/

[2] The index is formed as follows -- (miles from CBD, value of index): {[0-5],1}; {(5-10],1.15}; {(10-20],1.2}; {(>20),1.25}.

[3] STATA has a wide array of estimation routines often used in econometric analysis.