"The spatial organization of cities:

Deliberate outcome or unforeseen consequence? "

Revised May 2004

By Alain Bertaud
duatreb@msn.com
http://alain-bertaud.com

Introduction

Urban spatial structures are shaped by market forces interacting with regulations, primary infrastructure investments and taxes.

Urban spatial structures are usually the unintended result of unforeseen consequences of policies and regulations that were designed without any particular spatial concerns. However, different urban spatial organizations perform differently. For instance, some urban shapes are unfavorable to the development of public transport; others tend to increase the efficiency of public transport while reducing residential floor consumption. Urban spatial structures are very resilient and they evolve only very slowly. For this reason, a city's spatial structure significantly reduces the range of available development options..

It is not possible to define an optimum city shape because city development objectives change with time. However, it is possible to identify the type of city shape that would be consistent with a specific objective. Typically, mayors are obliged to pursue several objectives at the same time. The choice of the appropriate trade-offs between several often conflicting objectives is a political decisions, not a technical one. This choice is best left to elected officials. Urban planners, however, should constantly monitor the impact that specific policies may have on city shape. They should be aware of the effect of the most common planning tools – land use regulations, infrastructure investments and taxation – on the spatial organization of a city. They should make sure that the urban shape resulting from their actions will be consistent with the objectives set by elected officials.

Urban shapes are path dependent. The spatial structure of a large cities evolves very slowly and can evolve only in a few directions. On a large scale, it is never possible to bring back to nature the land that has been already developed. Planners should therefore have a good understanding of the potentials and liabilities inherent to the current spatial organization of the city in which they work. This paper is demonstrating a number of tools and spatial indicators to apprehend a city's spatial structure and to help formulate its potential for different development objectives.

Summary

The raison d'être of large cities is the increasing return to scale inherent to large labor markets.

Cities' economic efficiency requires, therefore, avoiding any spatial fragmentation of labor markets. In simpler terms, it means that all the locations where jobs are offered should – at least potentially – be physically accessible from the place of residence of all households within about an hour travel time. This requirement should be borne in mind when evaluating alternative urban shapes. Any type of spatial organization implying that residence and jobs should be matched individually – i.e. that workers need to have a good access only to their current job location – contradicts our premises that large competitive labor markets are efficient and that this efficiency alone justifies the complexity and high operating costs of large cities.

Spatial indicators allow to compare cities' structures and to monitor the evolution in time of individual cities spatial organization.

Urban spatial structures can be defined and compared by using a number of indicators related to average land consumption, to the spatial distribution of population and to the pattern of daily trips. Comparing the value of these indicators among cities shows amazingly resilient common features – such as the negatively sloped density and land price gradient – but also variations of several orders of magnitude – such as the land consumption per person between Asian and North American cities.

Some spatial structures are more compatible than others with environmental and social objectives

It is possible to establish linkages between spatial structure and city performance in various sectors? In this paper we look more particularly at the link between city shape and (i) transit use and motorization, (ii) air pollution due to transport and (iii) poverty. We found that dense contiguously urbanized and dominantly monocentric cities are favorable to transit and may significantly reduce trip length and as a consequence the total amount of pollutant emitted by transport. However, in the absence of adequate traffic management in the central parts of cities, the concentration of pollution might be higher in dense dominantly monocentric cities.

Dense monocentric cities have typically higher land prices and therefore tend to reduce the housing floor space and land consumption of the poor while they tend to provide better and cheaper access to most of the jobs.

Can urban planners influence a city spatial structure? Should they?

Should urban planners attempt to change a city's spatial structure in order to improve a city's performance in particular sector such as transport, environment or access to jobs by the poor? The chances to do so are rather limited and they are long range, but they nevertheless exist. A planner disposes of three tools to influence city shape: land use regulations, infrastructure investments and taxation. However, to be able to use these tools coherently, clearly established objectives must have been formulated by elected officials. Because there is no optimum city shape per se, a city shape can be "improved" only in relations to priority objectives. Priorities, however, may change with time, while cities shapes are very resilient.

Inadvertent changes in city shape caused by poorly conceived regulations or infrastructure investments are much more common than voluntary shape changes. Planners should conduct an audit of existing regulations to find if their combined effects on city shape are consistent with the municipal priorities.

Do cities' shape tend to converge toward a standard spatial organization?

Is there a global trend in the evolution of urban spatial structures? From the available empirical evidence it seems that large cities tend to become less monocentric and that as a consequence the share of transit is eroding in most cities of the world, in spite of heavy investments and subsidies. On the other hand, in cities of Europe and Asia, which have a deliberate policy to provide adequate services in high density core and to invest in urban amenities – urban design, new theaters, museums, pedestrian streets, etc. – land prices in the city center tend to increase. This would indicate that the monocentric model is not dead or even dying and that the center of large cities can provide attractions which cannot be matched in the suburbs. However, city centers of large cities, however prestigious or attractive, contain only a fraction of the total number of jobs.

Telecommuting, which theoretically will do away with the need of face to face contact for a large variety of urban activities like jobs, shopping, education, and entertainment, has not yet had a marked effect on the structure of any city. It is all the more important to monitor the evolution of city shapes and the spatial aspects of the land market to detect any changes in locational demand due the information revolution.

The growth of large cities is a self generated phenomenon, it was never deliberately planned and was often actively discouraged

The growth of the large metropolitan areas and of megacities, occurring in 2002 on every continent, had never been stimulated by deliberate policies or even accepted as the unavoidable consequence of economic development.

In the 60s and 70s there was a consensus among municipal officials, urban planners, and municipal engineers that it was desirable to stabilize the size of large cities. It was thought that large cities of several million people would be unmanageable and unlivable. At the same time, there was a concern for "geographical equity", an abstract concept implying that all geographical areas of a country should grow at about the same rate. The growth and dominance of large metropolitan areas were considered abnormal inequitable and malignant in nature.

National urbanization policy tended to promote the growth of small towns and discourage the growth of large urban areas. This consensus on the negative social and economic effect of large cities cut across ideologies and the same negative bias against large cites could be found in the Soviet Union, in Communist China, in Cuba as well as in most market economy countries of Western Europe, Asia and America. The United States seems to have been the only country which escaped the trend, but probably only because the Federal government did not have the constitutional right to impose a national policy on city development. On the other hand, states perceived demographic growth as a competitive advantage. It would not have occurred to the state of California to consider its growth inequitable as compared to West Virginia's.

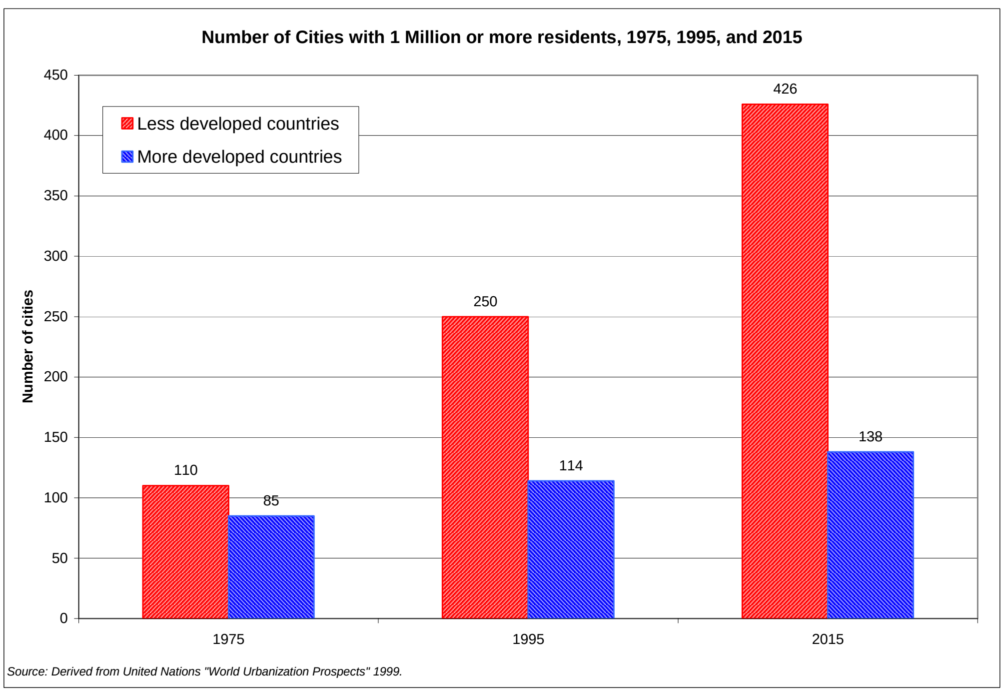


Figure 1: Growth of cities larger than a million people between 1975 and 1995 Source: Derived from United Nations, World Urbanization Prospects, 1999 Revision (2000).

In spite of this universal bias against them, during the last 30 years large cities did grow at a rapid pace as shown on *Figure 1* and are projected to grow further. This is also true for Megacities. According to United Nations reports on urbanization, in 1975, less than 2 percent of the global population resided in cities of 10 million or more residents. The proportion now exceeds 4 percent, and is projected to top 5 percent by 2015, when almost 400 million people will live in megacities.

Large labor markets are the only raison d'être of large cities

The fact that large cities have grown and keep growing, in spite of national policies which were biased against them, suggests that some potent economic reasons might be behind this growth. Large cities become more productive than small cities when they can provide larger effective labor markets. A large literature looks at cities as mainly labor markets like Ihlandfeldt, (1997) and the classic Goldner (1955), arguing that labor markets have increasing return to scale, which would explain the existence of megacities in spite of the difficulties in managing them. A large unified labor market is the raison d'être of large cities. Prud'homme (1996) provides a convincing explanation for the growth of megacities in the last part of the twentieth century: Megacities' capacity to maintain a unified labor market is the true long run limit to their size. Market fragmentation due to management or infrastructure failure should therefore result initially in economic decay and eventually in a loss of population[1]. In this paper, I am considering the spatial structure of a city as the possible cause of labor markets consolidation or fragmentation. It is obvious that the fragmentation of labor markets might have many different other causes, for instance, rigidity of labor laws or racial or sex discrimination.

This paper is focusing on the evaluation of the performance of various type of spatial organization. In evaluating spatial structures we will have to bear in mind that any shape whose effect is to fragment labor market will not be economically viable in the long run. This is an important reminder, as many planners pretend to solve the logistic problem posed by cities by proposing a spatial organization based on clusters of self sufficient "urban villages".

A viable type of urban structure should therefore allow complete labor mobility within a metropolitan area. Households, whatever their location within the metropolitan area, should be able to reach within a reasonable time (say less than 1 hour) all the locations where jobs are offered.

---

[1] I am certainly not implying here that the quality of infrastructure creates urban growth or that an infrastructure break down is the only reason why a city would shrink in size. Exogenous economic factors are of course determinant. But if infrastructure is not a sufficient reason to explain growth, the lack of it may explain stagnation in spite of favorable exogenous economic conditions.

Defining urban spatial structures

In order to evaluate the performance of various urban spatial structures it is necessary to establish some indicators which could be used to measure some of the most important spatial characteristics. Because we aim at an empirical analysis we will limit ourselves to the indicators which can be easily obtained in most cities by using census data, land use plans and satellite imagery.

To simplify the analysis we will consider only three aspects of urban spatial structures:

- The pattern of daily trips
- The average built-up density
- density profile and density gradient

Pattern of daily trips

Traditionally, the monocentric city has been the model most widely used to analyze the spatial organization of cities. The works of Alonso (1964), Muth (1969), and Mills (1972) on density gradients in metropolitan areas are based on the hypothesis of a monocentric city. It has become obvious over the years that the structure of many cities departed from the mono-centric model and that many trip-generating activities were spread in clusters over a wide area outside the traditional CBD.

As they grow in size, the original monocentric structure of large metropolises tends with time to dissolve progressively into a polycentric structure. The CBD loose its primacy, and clusters of activities generating trips are spreading within the built-up area. Large cities are not born polycentric; they may evolve in that direction. Monocentric and polycentric cities are animals from the same specie observed at a different time during their evolutionary process. No city is ever 100% monocentric, and it is seldom 100% polycentric (i.e. with no discernable "downtown"). Some cities are dominantly monocentric, others are dominantly polycentric and many are in between. Some circumstances tend to accelerate the mutation toward poly-centricity – historical business center with low level of amenities, high private car ownership, cheap land, flat topography, grid street design –; others tend to retard it – historical center with high level of amenities, rail based public transport, radial primary road network, and difficult topography preventing communication between suburbs.

A monocentric city can maintain a unified labor market by providing the possibility of moving easily along radial roads or rails from the periphery to the center (see Figure 2 (a)). The shorter the trip to the CBD, the higher is the value of land. Densities, when market driven, tend to follow the price of land, hence the negative slope of the density gradient from the center to the periphery.

The growth of polycentric cities is also conditional on providing a unified labor market. Some urban planners often idealize polycentric cities by thinking that a self-sufficient community is likely to grow around each cluster of employment. According to them, a number of self-sufficient "urban villages" would then aggregate to form a large polycentric metropolis (Figure 2, (b)). In such a large city, trips would be very short;

ideally, everybody could even walk or bicycle to work[2]. To my knowledge, nobody has ever observed this phenomenon in any large city. A metropolis constituted by self sufficient "urban villages" would contradict the only valid explanation for the existence and continuous growth of large metropolitan areas: the increasing returns obtained by larger integrated labor markets[3]. The urban village concept is the ultimate labor market fragmentation. Although there are many polycentric cities in the world, there is no known example of an aggregation of small self-sufficient communities. In spite of not being encountered in the real world, the utopian concept of a polycentric city as a cluster of urban villages persists in the mind of many planners. For instance, in some suburbs of Stockholm urban regulations allow developers to build new dwelling units only to the extent than they can prove that there is a corresponding number of jobs in the neighborhood. The satellite towns built around Seoul and Shanghai are another example of the urban village conceit: surveys are showing that most people living in the new satellite towns commute to work to the main city, while most jobs in the satellite towns are taken by people living in the main city.

In reality, a polycentric city functions very much in the same way as a monocentric city: jobs, wherever they are, attract people from all over the city. The pattern of trips is different, however. In a polycentric city each sub-center generates trips from all over the built-up area of the city (see Figure 2 (c)) Trips tend to show a wide dispersion of origin and destination, appearing almost random. Trips in a polycentric city will tend to be longer than in a monocentric city, ceteris paribus. For a given point in the city, the shorter the sum of the trips to all potential destinations, the higher should be the value of land. A geometrically central location will provide trips of a shorter length to all other location in the city. Therefore, we should expect polycentric cities to also have a negatively sloped density gradient, not necessarily centered on the CBD but on the geometric center of gravity of the urbanized area. The slope of the gradient should be flatter, as the proximity to the center of gravity confers an accessibility advantage that is not as large as in a monocentric city. The existence of a flatter but negatively sloped density gradient in polycentric cities can be observed in cities that are obviously polycentric, like Los Angeles or Atlanta.

---

[2] This is an extreme version of views expressed, for example, by Cervero (1989)
[3] Many papers such as Carlino (1979) and Sveikauskas (1975) document these increasing return to size.

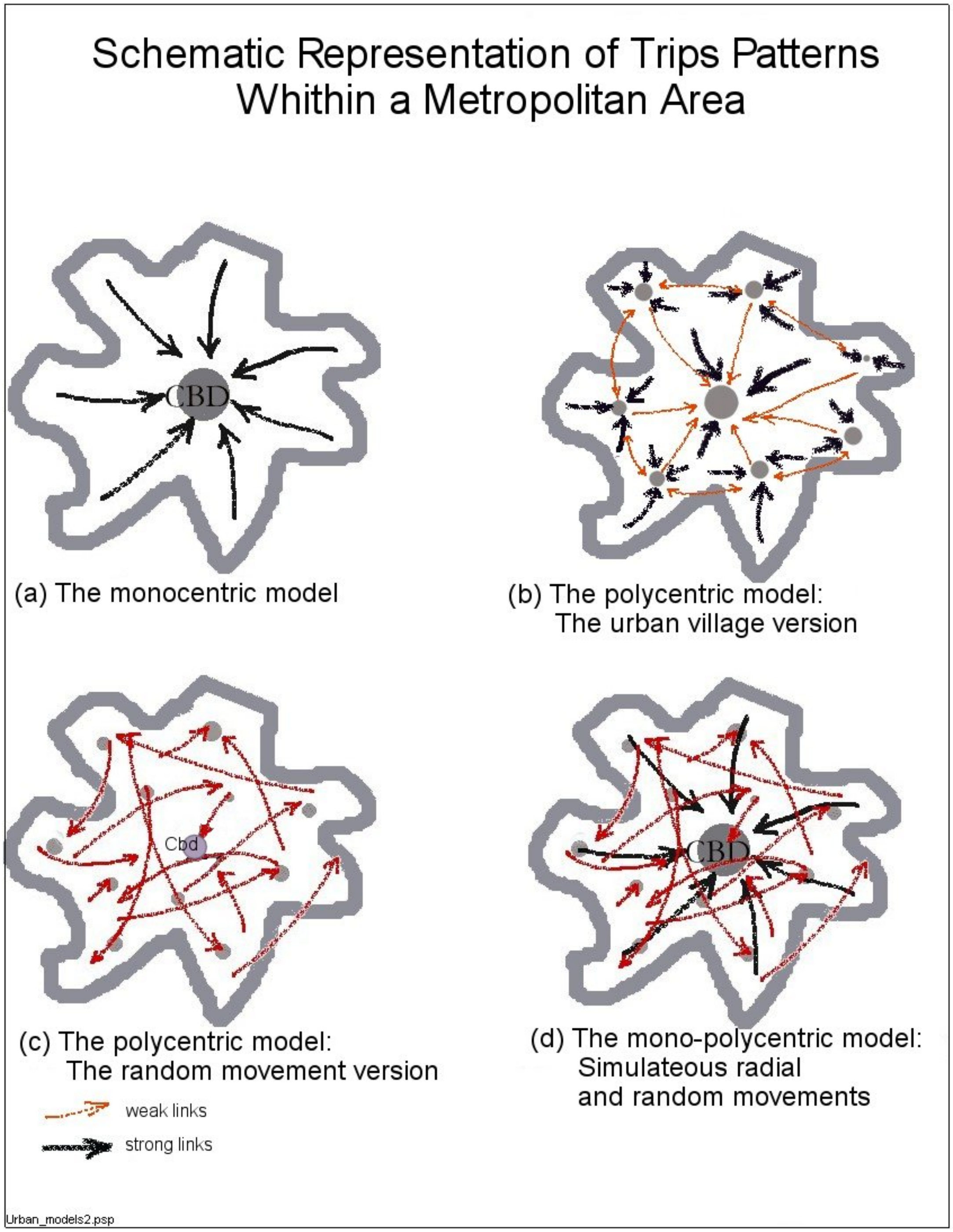


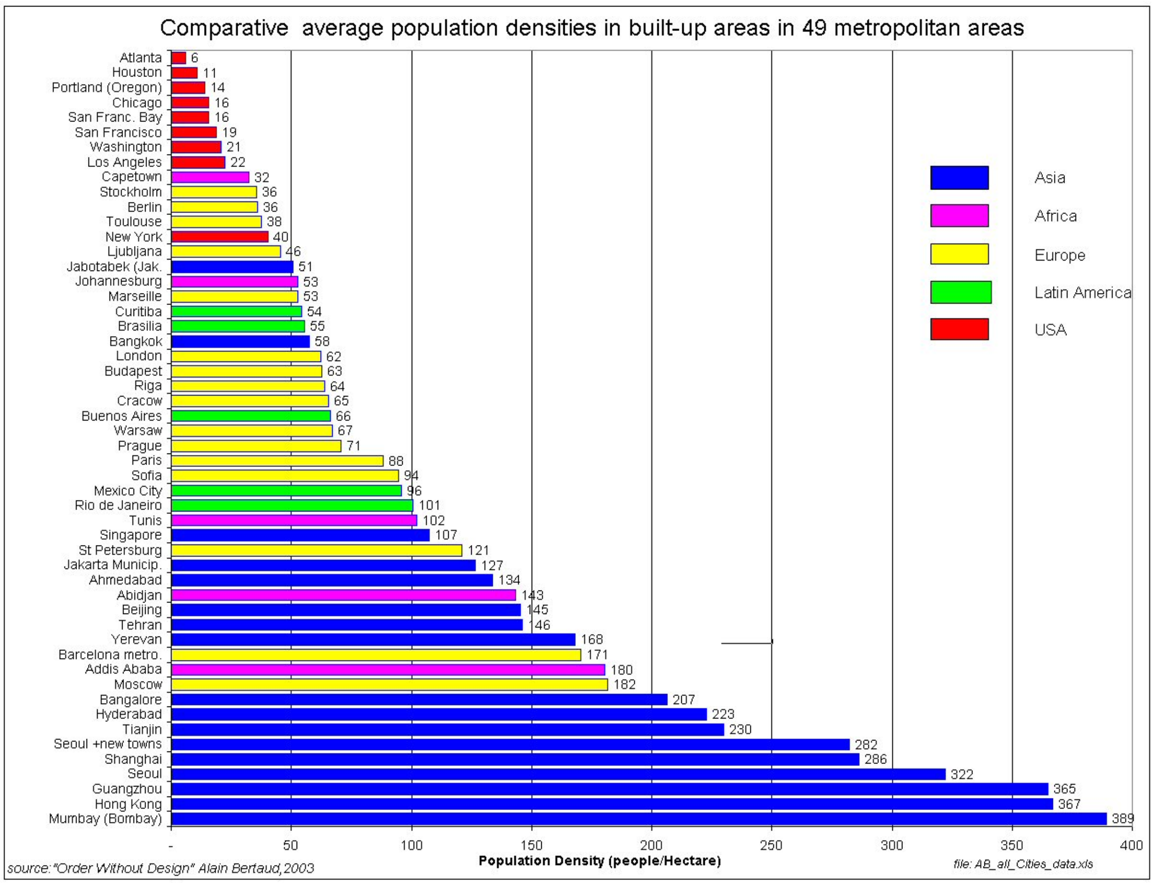
Figure 2: Pattern of daily trips

Land consumption: Comparative Average built-up densities

The amount of land consumed is an important parameter in defining an urban structure. The current concern for "sprawl" is in fact a concern for an over consumption of land by large cities. An accurate standardized measurement of urban land consumption is indispensable to address the issue of sprawl.

Land consumption (area of land per person) is usually measured by its inverse, population density (number of person per unit of land). Density is often measured as population divided by an administrative boundary, say, municipal limits. This measure of density is not very useful as municipal limits may include a large amount of vacant land, or even bodies of water. The only way to obtain a meaningful measure of density is to divide population by the built-up area which is consumed by urban activities. The densities mentioned in this paper have all been measured by dividing population by built-up area. Built-up area is defined as including all uses with the exception of contiguous open space larger than 4 hectares, agricultural land, forests, bodies of water and any unused land. In addition, land used by airports and by roads and highways not adjacent to urban used land is not included in the area defined as built-up area.

Figure 3: Average population densities in built-up areas in 46 Metropolitan areas



A comparison between the built-up densities of 49 cities around the world shows differences of several orders of magnitude (Figure 3). One should note that there is no clear correlation between density and income or between density and population size. However, one could make a case for a correlation between the density of a city and its location on a continent. US cities have the lowest densities; African, European, and Latin American cities have medium range densities, while Asian cities have high densities. This may suggest that densities may be strongly influenced by cultural factors. This is not surprising, as urban densities are largely influenced by the real estate markets, and therefore by consumers' trade-offs between commuting distance and land area consumed. The way households make these trade-offs are obviously influence by culture.

The cities whose densities are shown on Figure 3 are all reasonably successful cities, some might be better managed than others, but the great majority of them constitute the prime economic engine of the country to which they belong. This would suggest that – given the wide range of densities encountered – there is no "right, "correct", "manageable" or "acceptable" range of density per se. None of the cities in the sample shown on Figure 3, representing together about 250 million people[4], can be said of having a too low or too high density hindering its development or manageability.

Density profiles

The density profile within a city's built-up area is a convenient and simplified way to shows how the population is distributed within a metropolitan area.

Density profiles are based on density maps, themselves based on census data. Census data provides information about the spatial distribution of people when they are at home, say between midnight and 6 in the morning; census do not provide any information on where people are during the day. Where people are between midnight and 6 in the morning is important because it is the starting point of the daily trips discussed in the section above. What we call the spatial distribution of population is therefore an image of the location of the majority of a population of a city between about midnight and 6 in the morning. It is important to note that when urban planners show density maps they are of course showing the densities around midnight, not the density during the day.

The profile of density provides an image of the distribution of densities by distance from a central point which is usually the center of the central business district (CBD). In the large majority of cities the profile of density follows approximately a negatively sloped exponential curve as predicted by the model developed by Alonso (1964); Mills (1967); and Muth (1969). We can see that this is verified by a sample of 9 cities selected among US, European and Asian cities (Figure 4).

[4] Or about 10% of the world total urban population in 1990

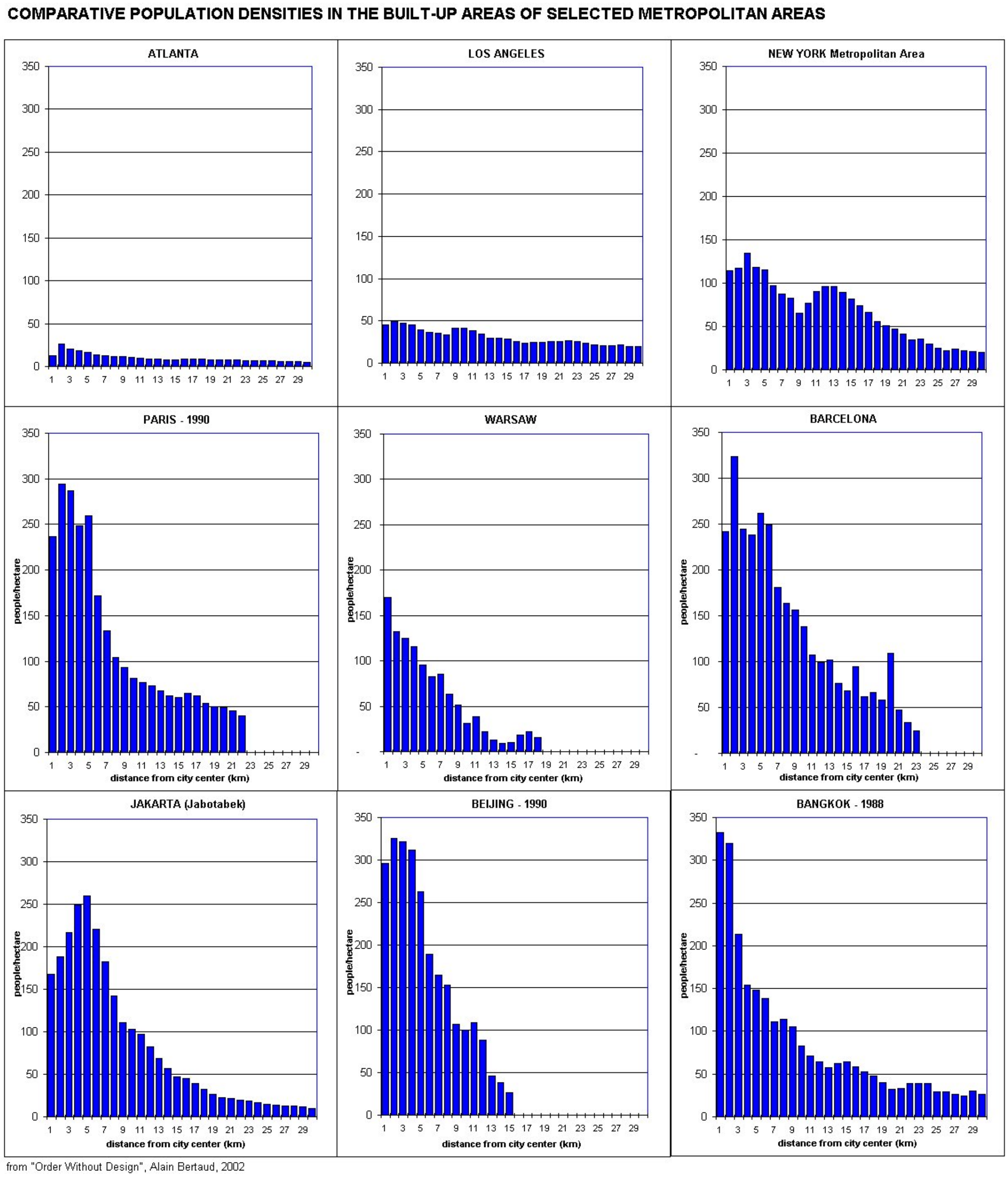


Figure 4: Profile of density of 9 Cities

The very large difference in absolute densities around the CBD between US and Asian and European cities can be related to the pattern of daily trips. Dominantly monocentric cities tend to have much higher densities close to the CBD than cities that are dominantly polycentric – such as US cities. The 6 non US cities shown on Figure 4 have densities within 4 km of the CBD ranging from 170 to 320 people per hectare (p/ha) compared to a range between 20 p/ha (Atlanta) to 120 p/ha (New York). We will see below that the spatial structures of monocentric high density cities are more compatible with the development of an effective public transport system than those of low density polycentric cities.

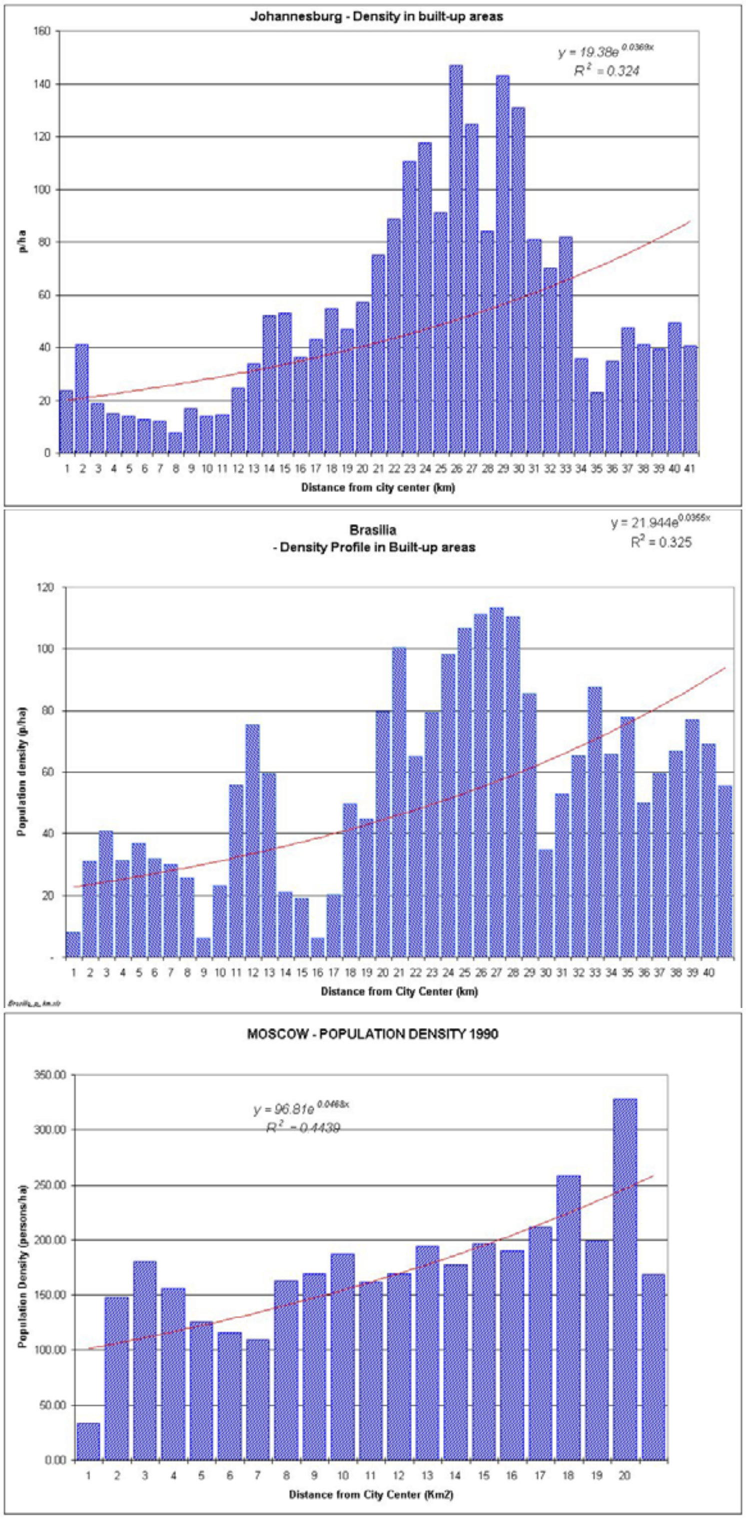


Figure 5: Brasilia, Johannesburg and Moscow density profile

The negatively sloped profile of density – as seen on Figure 4 – is generated by market forces as demonstrated by Alonso, Mills and Muth. This profile is so resilient that even cities with an historical interruption of land markets often shows a negative sloped gradient as it is the case for Warsaw and Beijing on Figure 4. However, a few cities where markets were interrupted or absent for long period show a positively sloped gradient (see Figure 5)

While a high or a low density does not have necessarily negative effect per se, a positively sloped density gradient constitutes always a liability for a city. For a given average density, in a city with a positive gradient, the median distance per person to the CBD will always be longer than in an equivalent city with a negative gradient. It is reasonable to infer that in a city with a positively sloped gradient all trips would be longer.

Moscow, Brasilia and Johannesburg are cities that seem to have very little in common except a history of disturbed land market. Whether the interruption was caused by Marxist ideology in Moscow, by a morbid cult of design in Brasilia or by Apartheid in Johannesburg, is irrelevant, the spatial outcome is similar. The positively sloped density profile reveals this common part of their history.

Linkages between spatial structure and transport efficiency

The type of urban structure often defined the most efficient mode of transport. The type of spatial structure, i.e. the degree of monocentricity and the density have a direct impact on trip length, on the feasibility of transit or private cars being the dominant mode of transport , and finally on pollution.

Densities, monocentrism and trips length

For a given population, the higher the density, the smaller is the built-up area. Providing the build-up area is roughly contiguous – i.e. not formed of large isolated areas like satellite towns – trips will be shorter in length in cities with high densities than in cities with low density. The comparison of the built-up area of 2 cities like Atlanta and Barcelona with similar population (about 2,5 million in 1990) but very different average density illustrates this point (Figure 6). In Atlanta the longest possible distance between 2 points within the built-up area is 137 km, in Barcelona it is only 37 km. The short trip distance due to high density in Barcelona makes it possible for a significant number of trips to be done by foot or bicycle, within Barcelona municipality, 20% of trips are made by walking. In Atlanta, the number of walking trips is so insignificant that it is not even recorded!

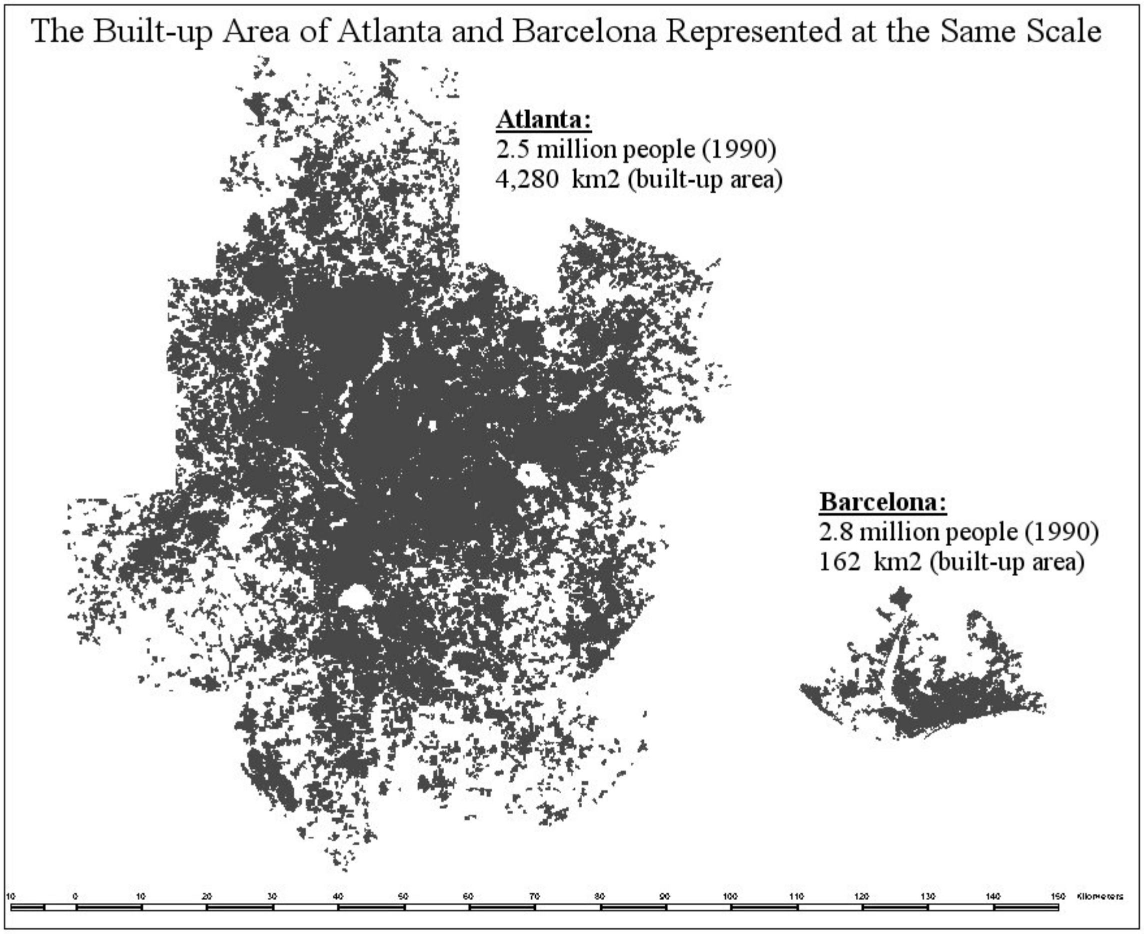


Figure 6: The Built-up Area and Barcelona represented at the same scale

But density is not the only factor which influences trip length. In a dominantly monocentric city, trips are usually shorter as the majority of trips are from the periphery to the CBD. In most dominantly monocentric city, the center of gravity of the population coincide with the CBD, this is the case in New York, London, Paris, Moscow, Shanghai etc. In this case the larger the proportion of trips to the CBD the shorter the trips will be as by definition the center of gravity is the point from which the sum of distance weighted by population is the shortest.

The effect of the spatial distribution of density on trip length is often underestimated. The theoretical graphs of Figure 7 show the large variations in trip length produced by different spatial arrangement for a imaginary city whose population and built-up area is kept constant, These variations in trip length occur whether trips are radial or from random origin and random destination.

Let us assume an imaginary city of 1 million people with an average density of 100 people per hectare, i.e. a built-up area of 100 square kilometers. To limit the number of possible shapes the variations will be limited to those who are inscribed within a square of 12 by 12 kilometers. Let us then test the variation of distance per person to the CBD and the average distance per person between random points for 20 variations of typical spatial structures, keeping the average density, population and built-up area constant. The variables are the density of sub-areas, the location of sub-areas with different densities and the shape of the built-up area within the limit of a square of 12 kilometers side. The results are shown on Figure 7.

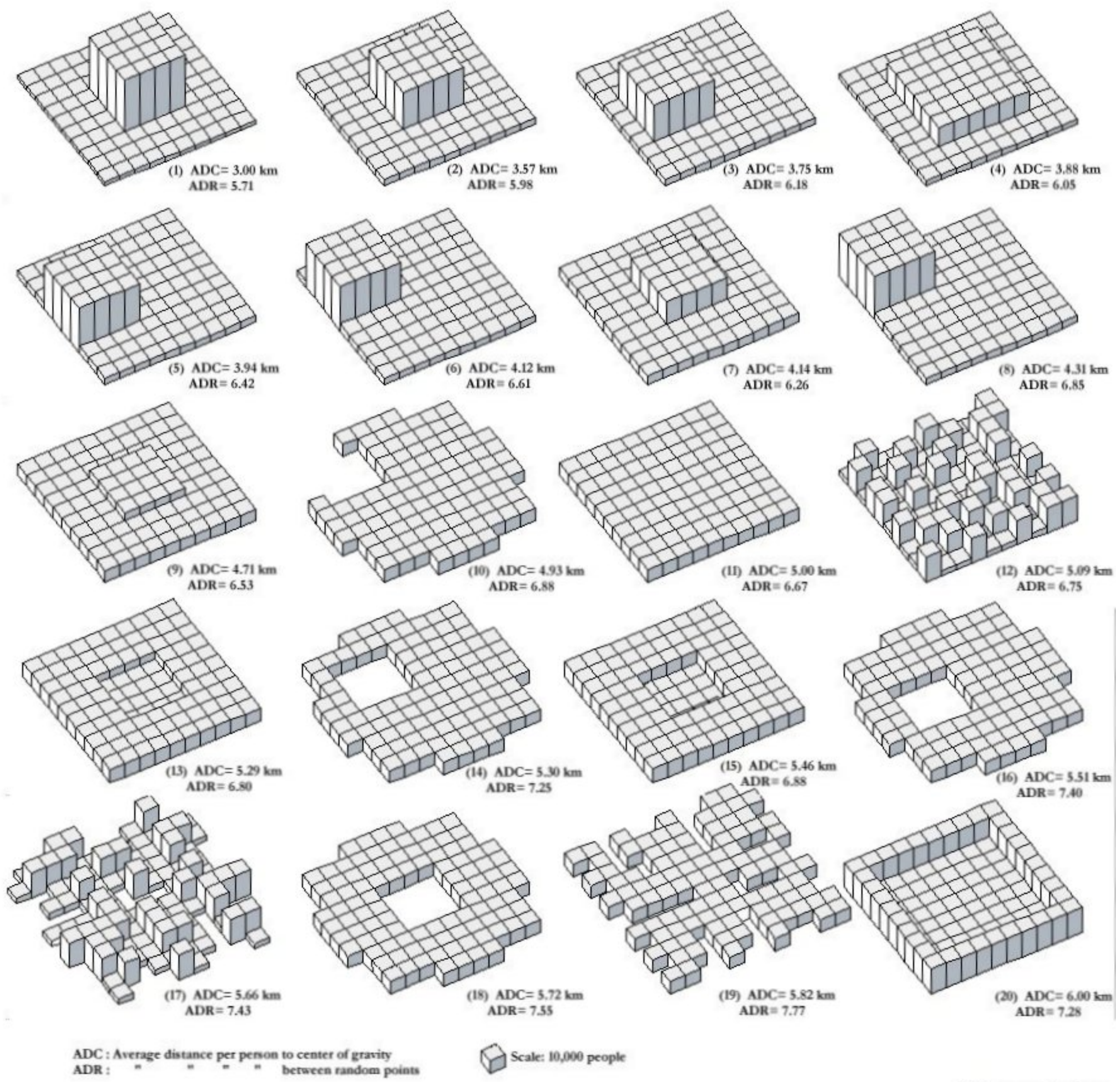


Figure 7: Schematic representation of different distribution of density in a city with constant average density and built-up area

The spatial organization types shown on Figure 7 are presented by order of decreasing performance for average distance to the CBD. The results allow us to draw 3 observations:

a. The variation of performance between types is large. The distance to the CBD double between layout 1 to layout 20 (from 3 to 6 kilometer, although the shape itself stay inscribed within a square of 12 by 12 km).

Between cities of identical average density, the distribution of local densities is therefore a very important factor in determining the length and the therefore the costs of trips and transport networks.

b. The variation in the distance to the CBD is much larger between different spatial arrangement than between the distance to the CBD and the distance between random points for a given shape. Shape itself is more important in city performance than whether a city is monocentric or polycentric.

c. While a poor performer for the distance to the CBD will generally be a poor performer for average distance between random locations, the correspondence is not linear. Some types of spatial arrangements, which are favorable to monocentric movements, are not favorable to random movements. For instance, layout ranked 13 for distance to CBD performs better for random movement than the layout ranked 8.

Compatibility of private cars with high densities

In this paper, densities are expressed in peoples per hectare of land (p/ha). But they can also be expressed in square meter of land per person. For instance the average density of Atlanta, 6 p/ha corresponds to a consumption of land of 1,666 m2 per person, and Barcelona's density of 171 p/ha corresponds to a consumption of 58 m2 per person. In the CBD of the European and Asian cities whose density profile is shown on Figure 4, the density is around 250 p/ha corresponding to a land consumption of 40 m2 per person.

A private car, which move around and park in a city needs at least about 40 m2 of land space. We can see that in Atlanta a car will occupy only a small fraction of the land available per person, only about 0.4%. By contrast, in the center of an Asian or European city, a private car would require about the same space as a person. The more cars are introduced in the CBD of dense cities the more they compete for space with people, not only with pedestrians but also with commerce, open space and all sorts of amenities. If private cars moving around and parking in a CBD were charged a market rent for the space they occupy, the problem of the allocation of land between cars and other activities would be solved. Unfortunately, this is not the case. Many cities have free or subsidized parking along the curbs of many city streets, and sometime even subsidized municipal parking off streets in downtown areas! Because of the political difficulty and at time the high transaction costs of charging market rents for the land occupied by cars in downtown areas, it is therefore often necessary to allocate land in an administrative way between cars and other urban activity. Hence the necessity of creating pedestrian only streets, and restricting free access to cars within historical areas as it has been done successfully in the historic centers of Cracow and Riga, for instance. The latest experiment in road pricing in the center of the city of London is another example of the difficulty of allocating space efficiently between car and people.

There is definitely a threshold of density beyond which private cars access should be severely restricted or even banned. In lower density area, the low land rent price does not justify the transaction cost of having paid parking; in dense downtown areas with high land rent, having cars pay a market rent for the land they occupy is the only way to obtain an efficient land allocation. The failure to charge for parking and street space may in the long run destroy the amenities of downtown areas, precisely because of the misallocation of land between those who pay market rents (shops, office housing) and those who do not (cars) . Singapore is the only city which has attempted to reflect the true costs of introducing cars in downtown area. The technology used allows to fine tune the pricing of downtown access while at the same time keeping the transaction cost low. Ironically, Singapore, by Asian standards is not particularly dense (see Figure 3).

Transit compatibility with various density levels and trip patterns

While we have seen that high densities are incompatible with the use of private car, the reverse is true for transit. Transit is incompatible with low densities and with spatial structures that are dominantly polycentric.

Table 1: Recommended Densities for Transit Operation

Recomended built-up and residential densities for various level of transit services

	Built-up Density p/ha	Residential Density p/ha
Boris Pushkarev and Jeffrey Zupan (1982)		
1 Bus: intermed serv, 1/2 mi between routes, 40 buses/day 7 du/res ac	**29**	**42**
2 Bus: freq serv, 1/2 mi between routes, 120 buses/day 15 du/res ac	**62**	**89**
3 Light rail: 5 min peak headways, 9 du/res ac, 25 - 100 sq mi corridor	**37**	**53**
4 Rapid tr: 5 min peak headways, 12 du/res ac, 100 - 150 sq mi corridor	**50**	**71**
The Institute of Transportation Engineers (1989) recommends the following minimums:		
1 1 bus/hour, 4 to 6 du/res. ac, 5 to 8 msf of commercial/office	**21**	**30**
2 1 bus/30 min, 7 to 8 du/res ac, 8 to 20 msf of commercial/office	**31**	**44**
3 Lt rail and feeder buses, 9 du/res ac, 35 to 50 msf of commercial/office	**37**	**53**
Peter Newman and Jeffrey Kenworthy (1989)		
"public transit oriented urban lifestyle"	**35**	**50**

Extracted from **John Holtzclaw, "Using Residential Patterns and Transit To Decrease Auto Dependence and Costs"; Natural Resources Defense Council, June 1994**

1 acre= 0.405 ha
Persons/dw= 2.4
% residential 70%

There is a purely geometric explanation to why low densities are incompatible with transit. Transit stations or bus stops have to be accessible by people walking from or to their residence or from or to their job or whatever other activity requires the trip. The walking speed of human beings in a city is limited by physiology to about 4.5 km an hour. Acceptable walking time to a transit station[5] varies with culture and income but surveys are showing that most people would not walk more than 10 minutes to a transit station or to a bus stop (although the acceptable walking time is usually higher for a metro station than for a bus stop). A 10 minute walking distance at 4.5 km/hour speed corresponds to a rounded maximum accessibility radius of about 800 meters to a bus stop or a metro station. A radius of 800 m in a street grid pattern will correspond to a catchment area varying between about 110 and 128 hectares depending on the arrangement between transit stop interval and transit line distances. As a rule of thumb I

[5] In the following section, I will use the words "transit station" to designate a metro station, a light rail station as well as a bus stop.

will use maximum catchment area of 120 ha per transit stop. When the number of people living, or working or shopping within this 120 ha fall below a certain threshold, transit becomes unpractical to the user and financially unfeasible to the operator. There seems to be a consensus among various researchers and operators that the density threshold for transit is around 30 p/ha (see table 1)

The comparison between Atlanta and Barcelona will illustrate in a more concrete manner the problems raised by low densities for transit operation. Atlanta and Barcelona had nearly the same population in 1990: Atlanta had 2.5 million people while Barcelona had 2.8 million.

Barcelona’s metro network is 99 kilometers long and 60% of the population lives at less than 600 meters from a metro station. Atlanta’s metro network is 74 km long – not so different from Barcelona – but only 4% of the population live within 800 meters from a metro station! Predictably, in Atlanta only 4.5% of trips are made by transit vs. 30% in metropolitan Barcelona.

Suppose that the city of Atlanta would want to provide its population with the same metro accessibility as Barcelona does (60% of the population within 600 meter from a metro station), it would then have to build an additional 3,400 kilometers of metro tracks and about 2,800 new metro stations. This enormous new investment will allow Atlanta metro to potentially transport the same number of people that Barcelona does with only 99 kilometers of tracks and 136 stations!

The example above illustrates the constraint that low density imposes on the operation of transit. I have been comparing metro track length and stations but a comparison between bus lines length and number of bus stops in Barcelona and Atlanta would have given the same results. With its low density of 6 people per hectare – compared to Barcelona 171 p/ha – Atlanta would have difficulties developing a viable form of transit, i.e. a transit system that is convenient for the consumer and financially viable for the operator.

In the case of Atlanta, the very low density precludes developing transit as an alternative transport to the automobile. “Encouraging” higher density, as many reports are fond of recommending, is not feasible either. To reach the 30 p/ha threshold over a period of 20 years, assuming that the historical population growth rate of 2.7% per year continues uninterrupted, the current built-up area would have to shrink by 67 %. In other words, about 67% of the existing real estate stock would has to be destroyed, the land over which it lays has to revert to nature and its population and jobs have to be moved into the 33% of the city which would remain.

The example of Atlanta shows how an existing spatial structure constrains the number of alternative strategies available to guide a city development. The lack of spatial analysis often leads to recommending unfeasible strategies, i.e. strategies which are incompatible with current urban structures.

Density is not the only spatial factor which constrains the development of transit; a dominantly polycentric structure is also a hindrance to transit operation. In monocentric cities most trips have multiple origins (the suburbs) but have one group of “clustered” destination (the city center); In polycentric cities most trips have multiple origins and multiple destinations. As a consequence, in a dominantly polycentric there is a multiplicity of routes with few riders. As a consequence, transit systems can operate efficiently in monocentric cities but are difficult to operate in polycentric cities.

Pollution and Spatial structure

The amount of air pollution generated by urban transport depends on the length, speed and number of motorized trips and the type of vehicles. For a given urban population, the length and number of motorized daily trips are closely correlated with the average population density in built-up areas, and the spatial distribution of trip destinations and origins.

Therefore, low density, polycentric types of urbanization have a double effect on pollution generated by transport: first, it increases trip length compared to denser more monocentric structures, and second, it increases the number of motorized trips as the proportion of transit trips and walking trips decrease with density.

However, engine technology and fuel types play also an important role in the amount of vehicular pollution and can counteract or attenuate the effect of unfavorable spatial structure. The comparison between Atlanta and Barcelona shows an interesting example of the impact of technology on urban air pollution. In 1999 the average yearly level of Nitrogen oxides was 47 $\mu g/m^3$ in Atlanta compared with 55 $\mu g/m^3$ In Barcelona. Air pollution due to traffic in Barcelona is higher than in Atlanta, is spite of the fact that Barcelona has a density 28 times higher than Atlanta and that 30% of trips are using transit and 10% are walking trips. This is apparently due to laxer emission standards for vehicles, in particular to the use of diesel fuel for cars (about 55% of private cars use diesel in Barcelona). In addition, vehicles tend to be older in Spain than they are in the US. Depending on the age of vehicle the amount of pollutant emitted can vary from 1 to 10. So, in some cases, air pollution might be more sensitive to the age and quality of vehicles than to the spatial structure.

High density monocentric structures certainly tend to decrease the total amount of pollutant emitted by transport compared to low density polycentric structures. However, the level of pollution exposure in dense monocentric center cities areas might be higher because of the more intense and slower traffic. Strict ban of on street parking to increase the speed of traffic flows and general traffic management measures are necessary to decrease high pollution exposure in central city areas.

Spatial structure and poverty

The type of urban spatial structure affects the welfare of the poor in a number of ways. In countries where the poor cannot afford individual means of transportation or where the large size of the city precludes walking as a mean of getting to jobs, dense monocentric cities are more favorable to the poor because they reduce distance and because they allow an efficient network of public transport.

The poor have a better chance to have a good access to most of the jobs in dense, monocentric cities. However, land is usually much more expensive in dense monocentric cities and as a consequence the poor would be able to afford less land and floor space for housing than in more spread polycentric cities.

High density housing requires a much higher quality of infrastructure and of urban services than low density housing. For instance, a leaky sewer in a low density settlement (say 50 p/ha) do not cause as much a health hazard than the same leaky sewer in a dense settlement (say, 800 p/ha , a common density in residential areas of many Asian cities). In the same way, a deficient solid waste collection system is less damaging to the environment in low density settlements than in high density ones.

To summarize: the poor are better off in a dense monocentric city when it comes to job access; however, in this type of city the poor are more likely to consume less land and floor space than in low density polycentric cities and the quality of their environment might be worse.

.

Urban laws regulating densities reduce the locational choices of the poor.

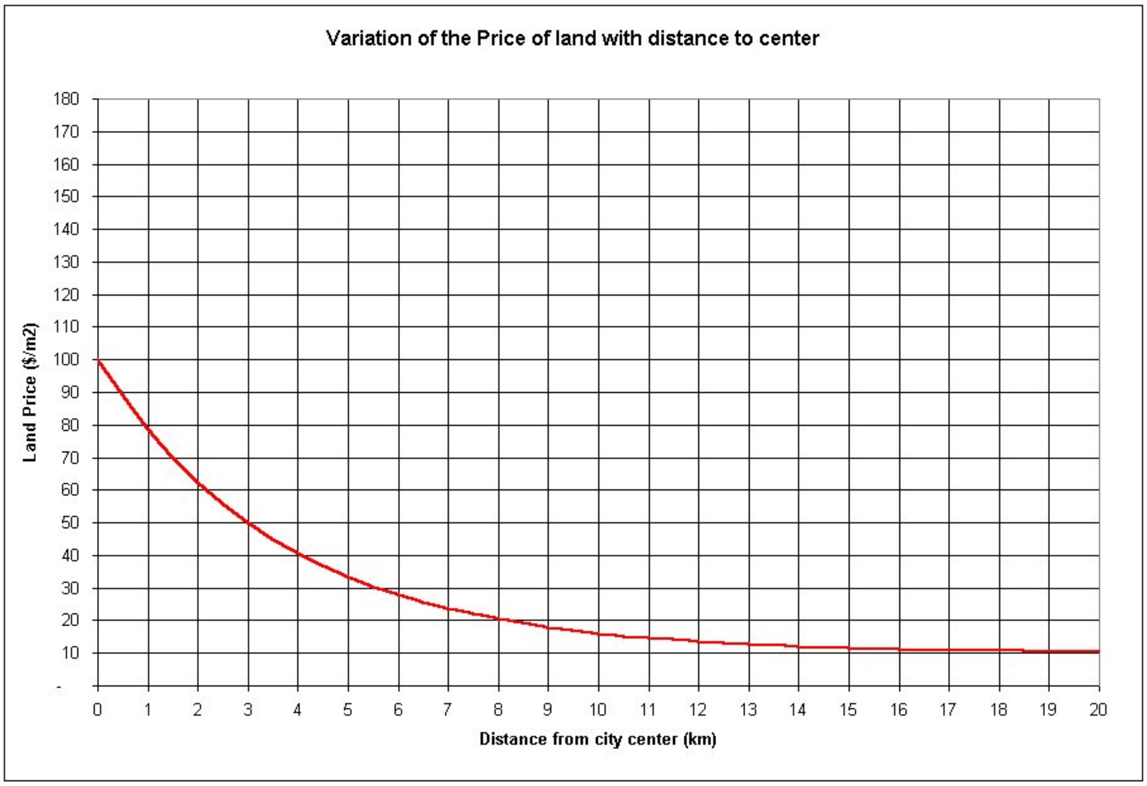


Figure 8: Land prices by distance to the city center in a monocentric city

It is important for the poor to make trade-off as do higher income groups when selecting a residential location. Residential mobility – defined as being able to change residential location to maximize its own welfare – is even more important for the poor than for the non poor. However, many well intentioned land use laws and subsidized housing program tend to severely limit the residential mobility of the poor.

Providing it can be retailed in non lumpy quantities, urban land is always affordable to most income groups. For a given price of land, each income group will adjust its consumption of land (and therefore density) and makes its own trade-off between distance to work and land and floor space consumption. However, land use regulations by establishing minimum plot sizes and floor area ratio tend to make land consumption lumpy, thus reducing the location choice of households who can only afford quantity of land below the arbitrary legal minimum. Many land use regulations have the effect of segregating the poor in areas which might not the best for their welfare. The theoretical example below will illustrate this point.

In a city where the land price profile follows the classical model established by Alonso, Muth and Mills as shown on Figure 8 where distance from the center in kilometers is represented horizontally and land price in US$ on the vertical axis.

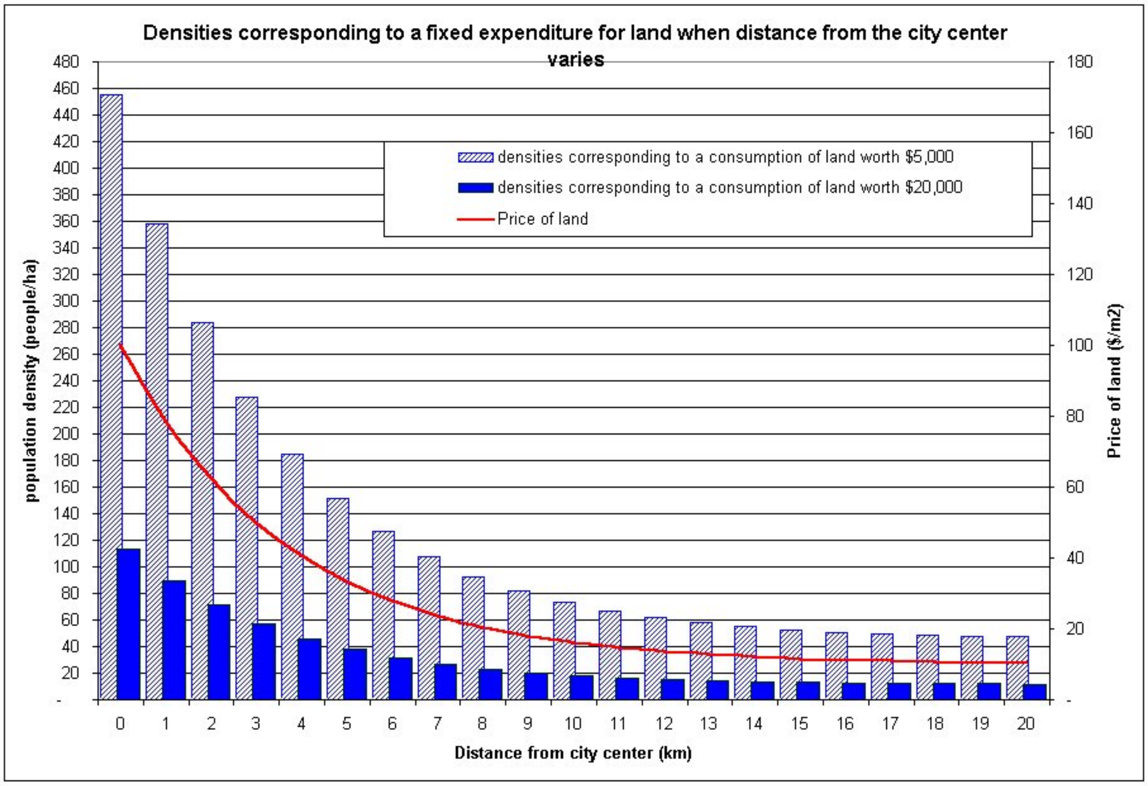


Figure 9: Affordable density for 2 income groups by distance from the city center

Let us assume that two income groups, A and B, are able to pay respectively $5000 and $ 20,000 for land. The affordable density for each group will vary by distance to the center as represented on Figure 9 where the left axis represent densities and the right axis land prices. It should be noted that there is a minimum theoretical density affordable for each group at any distance from the center. This does not mean that each group bid price will necessary follows the density curve for each group. The 2 density curves for each group represent only the density – and therefore the area of land – which can be purchased for the amount of money each group is willing to pay for land. For instance, the group A can afford to live at 3 km from the center at a minimum density of about 220 p/ha (45 m2/person) , while at the same distance group B can afford a minimum density of 60 p/ha (166 m2/person).

Let us now assume that a well intentioned urban planner draws a zoning plan covering the entire city containing – as zoning plans always do – restriction on minimum plot size, floor area ratio, set backs etc. The cumulative effect of all these restrictions will result in a de facto upper limit on densities set within the boundaries of each zone. The upper limit on density by distance to the center resulting from the zoning plan is represented by the red doted line on Figure 10. It can be seen that the result of the zoning plan is to exclude group A from most areas of the city except between a distance of 3 and 4 km and beyond 14 km from the center. In this particular case, group A would be practically relegated to the periphery of the city. An alternative for group A would be to bypass the effect of regulations by switching to the informal sector, but in doing so group A will lose a part of its property rights.

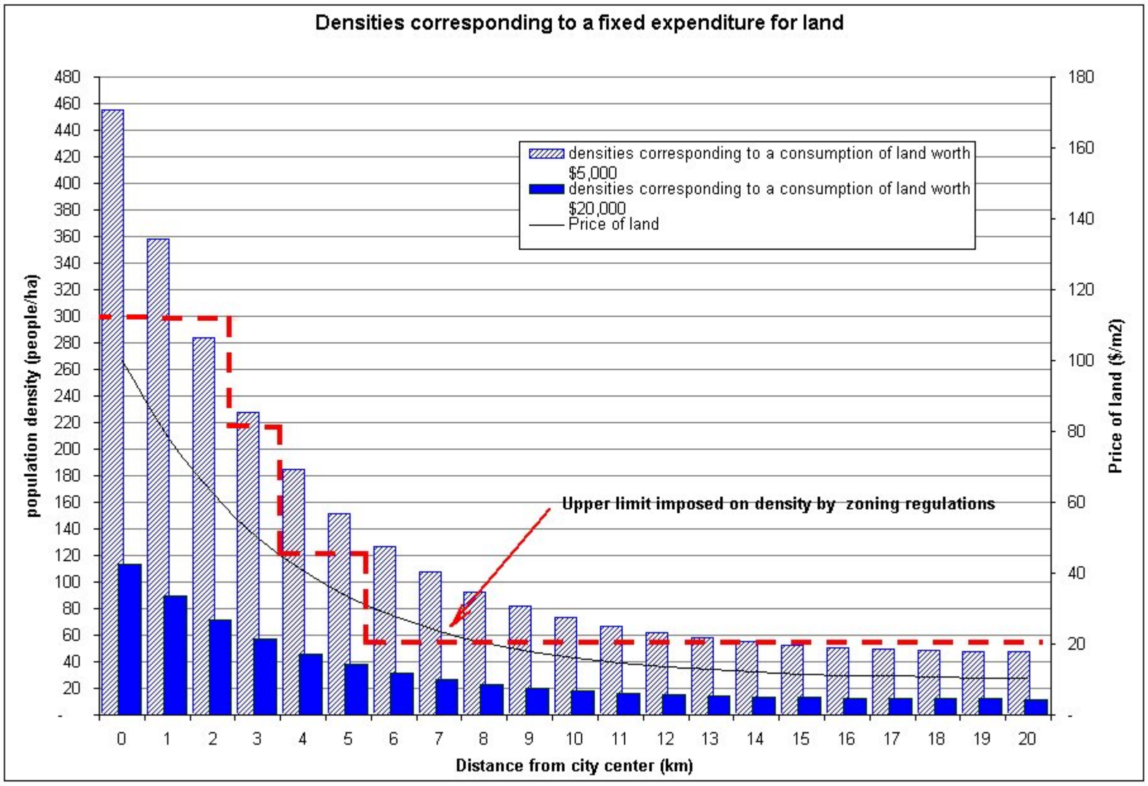


Figure 10: Zoning restriction and affordable densities

It must be noted that in this example, group B will be likely to be a strong supporter of the new zoning as it does not affect its residential mobility and its prevent group A from overcrowding the schools and infrastructure of the most desirable part of the city.

The example above is theoretical but it reflects the realities in many cities of the world. In Brasilia, for instance, one of the most carefully planned cities in the world, the majority of the poor are located in the far periphery while higher income groups are clustered around the center (Figure 11).

Subsidized low income housing projects most of the time reduce the residential mobility of the poor. The designers of the low income housing projects have to make the trade off between distance and plot area in advance of project construction. Low income households might have made a different trade off. Low income households are therefore tied to a location by the subsidy which goes with the project. Providing housing vouchers to poor households is the only way to direct housing subsidies without reducing the residential mobility of subsidy beneficiaries.

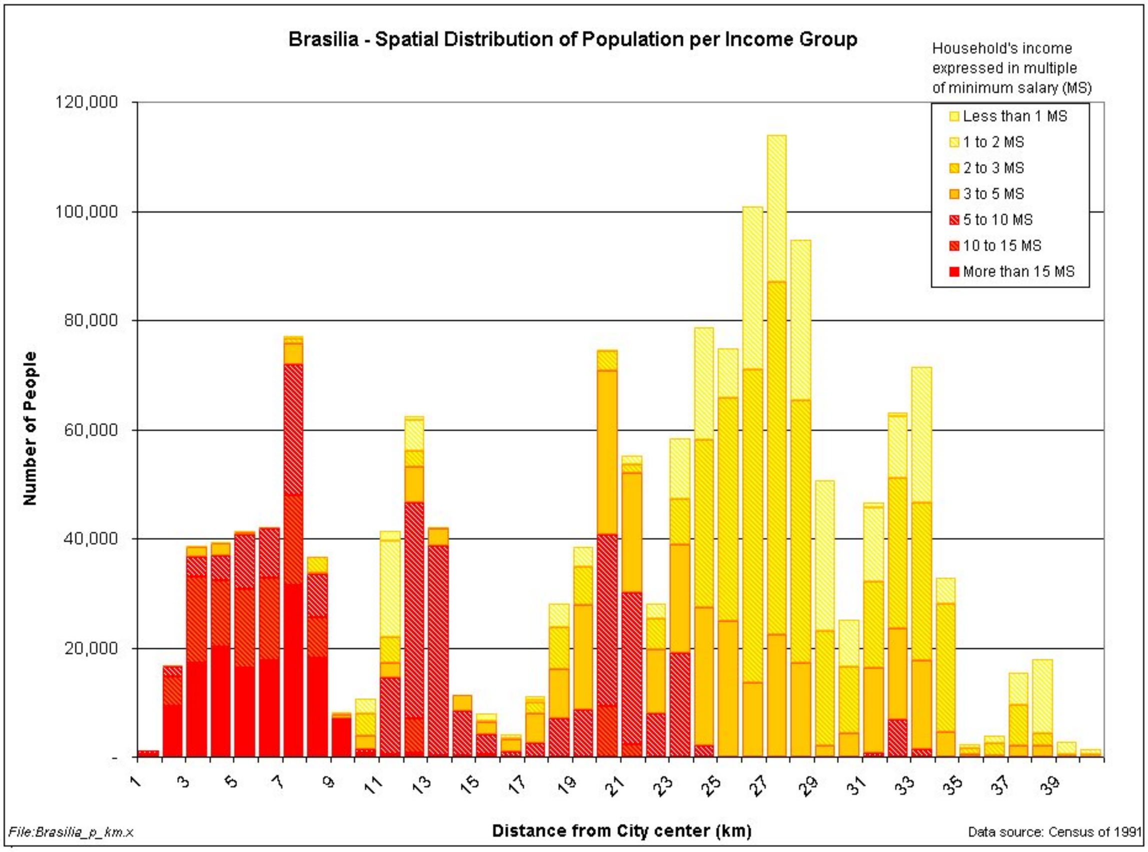


Figure 11: Brasilia, spatial distribution of population per income group

Should Municipalities attempt to change their cities' spatial structure?

Because of the different advantages inherent to some spatial structures, as discussed above, it might seem logical to attempt to change an existing urban shape in order to meet some specific objectives, as for instance a reduction of motorized trips. In doing so the following two points need to be taken into account:

First, there is no optimal spatial structure among many types of spatial organization. However, a positive density gradient and a dispersed non contiguous urbanization are clearly more costly to operate and have many negative environmental side effects and should be avoided..

Second, urban spatial structures are very resilient and are path dependent, i.e. it is easier to decrease density than to increase it, and it is easier for a monocentric city to become polycentric than the opposite.

Urban planners can influence city shapes only indirectly. Market forces in the long run are building cities (with few unfortunate exceptions like Brasilia). But market forces respond to constraints constituted by regulations and taxation and to opportunities provided by the network of primary infrastructure built by the state. Planners have therefore only 3 tools at their disposal to influence urban spatial structures: land use regulations, infrastructure investments and taxation. Figure 12. shows a schematic view of the interaction of markets and government action in shaping urban structures.

In order to influence the evolution of a city spatial structure these 3 tools should be carefully coordinated and be internally consistent to meet a common spatial objective. This consistency is very rare as regulations, infrastructure investments and taxes are often designed at different level of government and for very different purpose which have nothing to do with a city spatial structure.

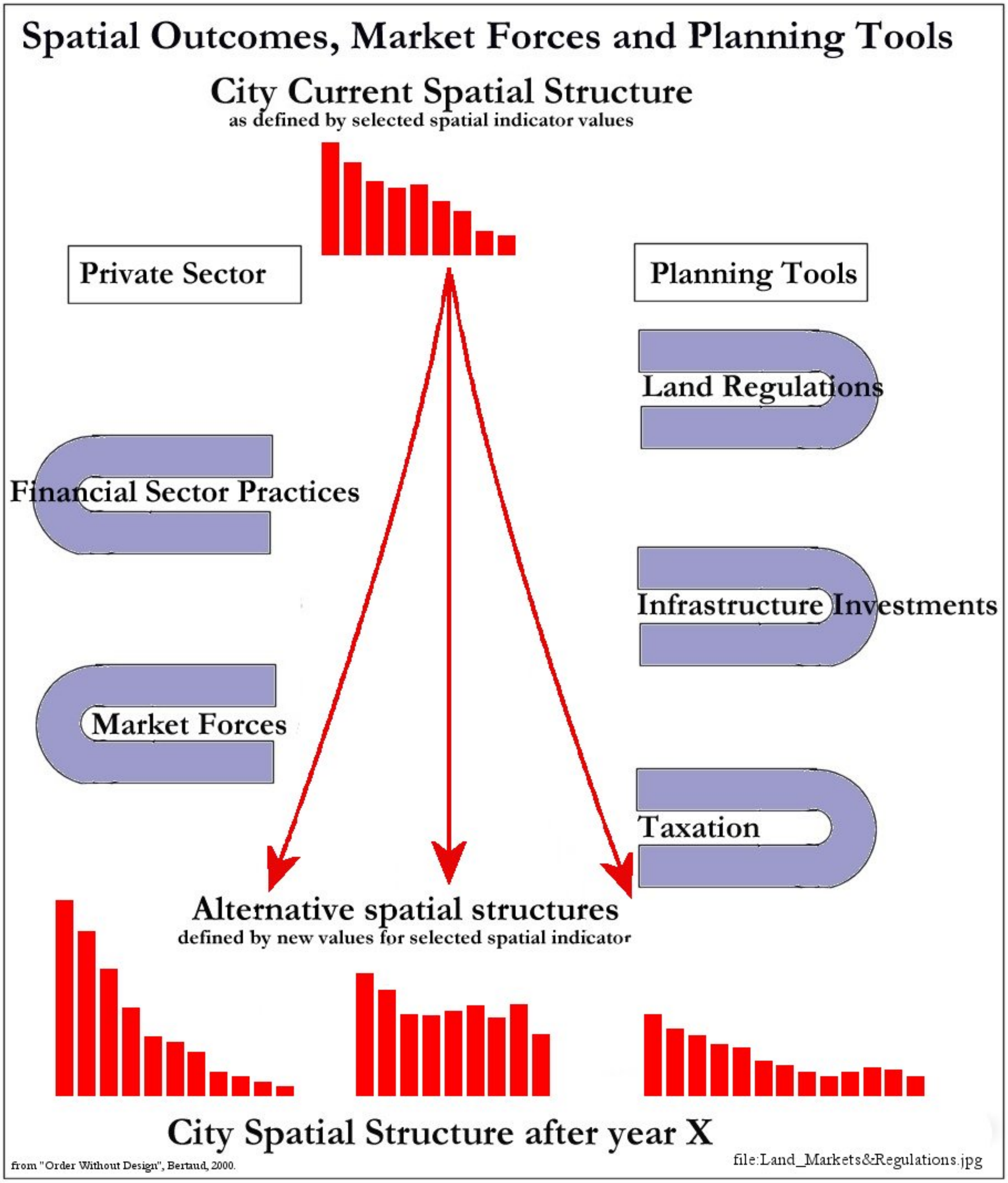


Figure 12: Interaction between market forces and government action

To summarize, urban planners should monitor urban structures to be aware of spatial trends and to know the limitations imposed by the current structure on policy options. In few cases, it will be possible to influence spatial trends in a limited way. A more ambitious outcome would require a strong, long term and continuous political determination combined with a dedicated and well integrated team of urban planners, municipal and transport engineers and financial planners. These conditions are seldom met during a period long enough to influence city shape. Curitiba is an example of a city where these conditions were met and where without doubt the current city structure is the result of a concerted long range effort. It is however not clear whether the resulting spatial structure has resulted in a welfare increase for the majority of the inhabitant of Curitiba, compared to what it would have been if the structure had been more market driven.

Is there a global trend in the evolution of urban spatial structures?

Because of the path dependency of the development of urban structure, dominantly monocentric cities tend to become less monocentric. And because income and mobility has been increasing in most large cities of the world, densities in central areas tend to get lower over time. However, it does not mean that every city is now tending toward a low density extremely polycentric model.

As cities become larger, the CBD itself becomes also larger. But by becoming larger the CBD loose the proximity which made it attractive in the first place. It is therefore inevitable that sub centers would emerge as a city becomes larger and that the degree of monocentricity decreases with size. However, some very large cities like New York, London, Buenos Aires or Shanghai retain a very strong center, which while containing an ever dwindling ratio of total jobs remain a very strong attractor for prestige retail, entertainment and culture. By contrast, some very successful cities manage to grow without any prestigious center, Atlanta or Phoenix are good example of this type of cities.

It is possible that in the future we will see a polarization between 2 types of cities. Both types will be polycentric in terms of jobs distribution. The first type will retain a strong prestigious center with a high level of amenities surrounded by a high density residential area inhabited by mostly high and middle income households. The second type of cities will be a pure labor market without any centrally located amenities; jobs and whatever amenities are provided will be evenly distributed throughout the metropolitan area without any prestigious center.

References

Alonso, William. *Location and land use*. Harvard university Press, 1964

Bertaud, Alain . *Cracow in the Twenty First century: Princes or Merchants? A city's structure under the conflicting influences of Land Markets, Zoning Regulations and a Socialist Past*. ECSIN Working Paper #8, The World Bank 1999

Bertaud, Alain and Bertrand Renaud. *Socialist Cities Without land markets*. Journal of urban Economics, 41, 1997, pp137-51

Bertaud, Alain and Stephen Malpezzi . *"Measuring the Costs and Benefits of Urban Land Use Regulation: A Simple Model with an Application to Malaysia* ".Journal of Housing Economics, 10 (3), pp. 393-418, September 2001

Bertaud, Alain and Stephen Malpezzi, "*The spatial distribution of population in 35 World Cities: the role of markets, planning and topography". The Center for urban land and economic research, The University of Wisconsin, 1999*

Bollinger and Ihlanfeldt, "*Intrametropolitan Locational Patterns of People and Jobs; Which Government Interventions Make a Difference*", 1999

Bruekner, Jan K. *Urban sprawl: Lessons from Urban Economics.* Papers on Urban Affairs 2001, Brookings-Wharton, Gale & Pack editors, 2001

Carlino, Gerald A. *Increasing Returns to Scale in Metropolitan Manufacturing*. Journal of Regional Science, 19, 1979, pp. 343-51.

Cervero, Robert. *Jobs-Housing Balancing and Regional Mobility*. American Planning Association Journal, 55, 1989, ppp. 136-50.

Clark, Colin. *Urban Population densities*. Journal of the Royal Statistical Society, 114, 1951 pp. 375-86

Epple, Dennis, Thomas Romer and Radu Filimon. *Community Development With Endogenous Land Use Controls.* Journal of Public Economics, 35, 1988, pp. 133-62.

Fischel, William A. *Regulatory Takings: Law, Economics and Politics*. Harvard University Press, 1995.

Goldner, William. *Spatial and Locational Aspects of Metropolitan Labour markets*. American Economic Review, 45, 1955,pp.111-28

Hall, Peter, *Cities of Tomor*row, Blackwell, 1988, pp 210-219

Hamilton, Bruce W. *Zoning and the Exercise of Monopoly Power*. Journal of Urban Economics, 5, 1978, pp. 116-30.

Ihlandfeldt, Keith R. *Information on the Spatial Distribution of Job Opportunities Within Metropolitan Areas*. Journal of Urban Economics, 41, 1997, pp. 218-42.

John Holtzclaw, "*Using Residential Patterns and Transit To Decrease Auto Dependence and Costs";* Natural Resources Defense Council, June 1994)

Malpezzi, Stephen and Yeuh-Chuan Kung. *The flight to the suburbs, revisited: the intraurban distribution of population and the price of housing*. January 1997

Mills, Edwin. S. and Byong-Nak Song. *Urbanization and Urban Problems*. Harvard, 1979

Muth, Richard F. *Cities and Housing*. University of Chicago press, 1969.

Prud'homme, Remy, *Managing Megacities*, Le courier du CNRS, n° 82, 1996 pp 174-176

Sveikauskas, Leo. *The Productivity of Cities*. Quarterly Journal of Economics, 89, 1975, pp. 393-413