Urban Studies, Vol. 37, No. 9, 1479–1496, 2000

Beyond Optimal City Size: An Evaluation of Alternative Urban Growth Patterns

Roberta Capello and Roberto Camagni

[Paper first received, January 1999; in final form, August 1999]

Summary. The aim of the paper is to present a critical view of theoretical works on city size. We begin with the consideration that, during the 1960s and 1970s, the question of optimal city size tended to be expressed in a misleading way. The real issue is not 'optimal city size' but 'efficient size', which depends on the functional characteristics of the city and on the spatial organisation within the urban system. Economies of scale exist up to a certain city size. However, urban development generates conditions leading to structural readjustments which may create new economic advantages. These structural adjustments may either be sectoral trasformations towards higher-order functions, or increases in external linkages with other cities. The paper provides empirical evidence of these processes, and contains an econometric evaluation of urban location benefits and cost functions with respect to different levels of network integration, size and urban function. The model is applied to 58 Italian cities.

1. Introduction

In the real world, the number of people living in cities is growing in all countries and continents. Urbanisation is a phenomenon which, in the past decade, has become increasingly intense in developing countries. The share of urban population in the more developed continents, such as Europe and North America, is extremely high and at the world scale is nearly 50 per cent. This percentage, according to the forecasts of the World Resources Institute (1994), is expected to rise yet further in future decades. As a consequence of increasing population, cities physically expand through processes which have been labelled as 'ville éclatée', 'ville éparpillée' and 'ubiquitous city'. The population of large cities is continuing to grow, though sometimes more slowly than previously (Camagni, 1996).

The constantly increasing size of cities encountered in the real world is in contrast with the famous 'optimal city size' theory, which envisages a size above which an increase in physical dimension decreases the advantages of agglomeration. The declining rate of urban population growth recorded in the past decade in most developing countries appears to be common to all cities, independently of physical size, and represents a general slowing down, rather than a specific crisis in the larger cities. Indeed, during the 1970s, there were negative population

Roberto Camagni is in the Dipartimento di Economia e Produzione, Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milano, Italy. Fax: 2 2399 2710. E-mail: Roberto.Camagni@polimi.it. Roberta Capello is in the Department of Economics, University of Molise and at the Dipartimento di Economia e Produzione, Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milano, Italy. Fax: 2 2399 2710. E-mail: Roberta.Capello@polimi.it. Although the paper is the result of a common research effort, R. Capello is responsible for writing the paper.

0042-0980 Print/1360-063X On-line/00/091479-18 © 2000 The Editors of Urban Studies

growth rates in the urban system of the Po Valley in northern Italy, not only in the major cities, but also in a number of secondary centres of 75 000 to 150 000 inhabitants (8 out of 19) and even in some smaller towns of 20 000 to 75 000 inhabitants (27 out of 113) (Camagni *et al.*, 1985, 1986). According to the theory, however, medium-sized towns are expected to increase their size, since the advantages associated with the physical dimension are still higher than location costs.

This seemingly mistaken interpretation of the real world by the optimal city size theory has already been pointed out by various authors. Richardson (1972) was the first to present a 'sceptic's view', by emphasising that an apparent paradox existed between the theoretical acceptance of an 'optimal city size' and the contradictory development patterns of urban systems in the real world. According to Richardson, this paradox could be explained by the existence of other determinants influencing urban agglomeration economies, not merely physical size. Since Richardson's paper, other interpretations have been given to this apparent paradox, through the 'urban life cycle' theory (see, for example, van den Berg *et al.*, 1983; Camagni *et al.*, 1985) and through the integration of dynamic elements, such as innovation, continuous information and knowledge acquisition, into the static framework of optimal city size theory.

In this paper, we accept the basic criticism of the optimal city size theory and the idea that there are determinants of urban location advantages other than urban size. Such determinants have already been identified in the literature; these include the kind of economic functions developed by the urban centre, the spatial organisation in which the centre operates and the efficiency of each centre's internal structure. The aims of the present paper are the following:

—to identify the main theories which supersede the limits of the neoclassical approach (section 2);
—to extend the definition of urban location benefits and costs to aspects constituting the city other than the mere economic elements (section 3);
—to define a methodology to estimate these effects and their variations with respect to the main determinants emerging from the literature (section 4); and
—to apply this methodology to a sample of 58 Italian cities (section 5).

2. New Paradigms for an Old Problem

The starting-point of our reflection is optimal city size theory, which claims that urban size is the fundamental determinant of urban location costs and benefits. The theory states that the well-known indivisibility and synergy mechanisms, which are at the basis of economies of scale in cities, apply up to a certain urban size, after which diseconomies of scale due to congestion effects take place and decrease the average revenues of an urban location. The optimal city size is calculated as the result of the maximum difference between a location cost curve, defined by Alonso as the land rent costs associated with urban size, and the aggregate agglomeration advantage curve.[1]

Although demonstrated by a large number of empirical estimations, many criticisms have been made of the neoclassical approach to optimal city size theory. These include the observations that:

—Cities are different from one another. They are characterised by different functions and perform different specialisations (Henderson, 1985, 1996). The use of the same urban production function for all cities in econometric analyses estimating optimal city size is extremely restrictive. In the words of Richardson (1972, pp. 30): "We may expect the efficient range of city sizes to vary, possibly dramatically, according to the functions and the structure of the cities in question".
—If cities are different from one another, the optimal city size may be different, depending on the specific characteristics. Richardson elegantly compares the opti-

mal city size theory with the theory of the behaviour of firms. We would never expect the optimal position for each and every firm to occur at the same level of output, so why should we expect the optimal point in each city to be located at the same population level?

—Cities exist in an interurban environment. The optimal city size theory, on the contrary, does not consider the spatial context in which cities operate.

—Cities generate a large variety of externalities as a result of the qualitative characteristics of the urban production environment. As long ago as 1961, Chinitz expressed some doubts about the fact that urban factor productivity depends mainly on the physical size of cities. He emphasised, on the contrary, the importance of a diversified and competitive urban production system as a source of urban productivity. Such a system is able to provide a far larger variety of externalities for small firms than an oligopolistic and specialised urban structure. Chinitz supported his thesis with an empirical analysis of New York, a large and diversified urban area, and Pittsburgh, a highly specialised city.[2]

The theories which have superseded the above limitations of the neoclassical theory on city size can be grouped into two different conceptual paradigms.[3] We refer to these two paradigms as the 'neoclassical city interpreted within a logic based on the Christaller model' and the 'network city paradigm' (Table 1).

The first paradigm deals with some of the limitations of the optimal city size theory by stating that: urban size is defined as the equilibrium between production benefits and location costs; and, cities are not all the same, but produce different goods according to their size. In the neoclassical city, location benefits and costs are by definition equal. This is true in an intraurban equilibrium logic, according to which, in the Alonso–Fujita model, the residential and production location equilibrium—for example, for a suburban location—is achieved via a compensation mechanism between accessibility and urban rent (Alonso, 1964; Fujita, 1985). The result of the model is an indifferent location choice among all possible locations—i.e. lower accessibility to the centre is compensated for by lower rents and higher environmental quality. The same holds for an interurban equilibrium: in an equilibrium solution, the same profits and utility levels have to be guaranteed by each city. If this is not the case, *ceteris paribus*, a city offering higher rents but lower agglomeration benefits (with the hypothesis of non-existent transport costs) would lose both residents and firms. Urban size is in this case the result of market forces, pushing towards the maximisation of utility levels for residents and profits for firms.

In a neoclassical approach, the use of the same production function for all cities inevitably generates cities of the same size (Camagni, 1992). This evident paradox can be overcome either through the hypothesis of different production functions for each city (and thus a single production function for each city) as suggested by Henderson (1985), or by expressing the neoclassical logic through the use of the Christaller model. In this second case, neoclassical logic leads to the definition of a hierarchical urban system—by definition in equilibrium, thanks to market forces—where differences in city size can be interpreted as the compensation between agglomeration advantages, on the one hand, and higher urban rents and diseconomies of congestion, on the other.

This reasoning, though elegant and fascinating in its theoretical interpretation, overcomes perhaps rather too simply the problem of optimal city size. The problem simply does not exist—thanks to the ability of an urban system of any size to equalise costs and benefits, and to find by definition an equilibrium solution. An indifferent location choice emerges over the whole of geographical space, since the agglomeration advantages are perfectly capitalised into urban rents.

While optimal city size theory gives the

Table 1. A comparison of three theoretical paradigms

Paradigms/ Elements	Optimal city size	The neoclassical and Christallerian city	The network city
Characteristics of the approach	Empirical	Theoretical	Theoretical and empirical
Characteristics of the city	Undefined city (aggregated)	Despecialised city	Specialised city linked with a large urban system
Characteristics of the urban system	Not considered	Hierarchical	Networked
Characterising element	Urban size	Urban size interpreted through the urban functions	Distinction between size and urban function. Analysis developed in a spatial context
Urban efficiency	Agglomeration economies	Functional upgrading of economies	Co-existence of network externalities, economies of agglomeration and functional upgrading
Result of the analysis	An intraurban equilibrium exists which has to be reached	An intraurban and interurban equilibrium exists by definition	There exists an intraurban equilibrium which can be reached through interurban system relationships
Urban policy aims	Achievement of an intraurban equilibrium between costs and benefits obtainable through the urban dimension	None: the system is in equilibrium by definition	Achievement of a cost–benefit equilibrium through specialisation policies and/or network integration

Source: Capello (1998a).

impression that it is an empirical exercise without an underlying theory,[4] we could argue that the 'neoclassical city interpreted within the logic of Christaller' is a theory without empirical application. The result achieved by the model—a general equilibrium of cities, all of the same size—is undoubtedly unrealistic. This approach leaves no space for normative interventions.

An interesting paradigm that has emerged recently, which overcomes some of the limits of both the optimal city size and the neoclassical/Christallerian approach (Table 1), is the so-called network city paradigm. The most important theoretical novelty provided by this paradigm is the break of the link between urban size and urban function imposed by the Christallerian logic. With Christaller's approach, it is in fact impossible to explain why a city like Zurich, with only 300 000 inhabitants, is specialised in international finance in the same way as the city of New York or Tokyo. In the real world, urban size is not always characteristic of function.

The break between the relationship between urban size and function is one of the main characteristics of the SOUDY (supply-oriented dynamic approach) model (Camagni *et al.*, 1986), which argues that

—Higher-order functions are characterised by higher thresholds for the level of appearance in the city (in terms of urban population) (d_1, d_2, d_3 …).
—The average (aggregate) benefit–cost curve increases for higher-order functions, due to: growing entry barriers; decreasing elasticity of demand which allows extra profits to be gained in all market conditions; and, increasing possibility of obtaining monopolistic revenues due to the use of scarce, qualified factors.
—The location cost curve has the traditional form suggested by Alonso.

When the average benefit function is compared with Alonso's traditional location cost curve, the following results are achieved. For each economic function and each associated urban rank, it is possible to define a minimum and a maximum efficient city size, which would increase with the level of the urban function and rank (Figure 1).

The interest of this model is that it over-

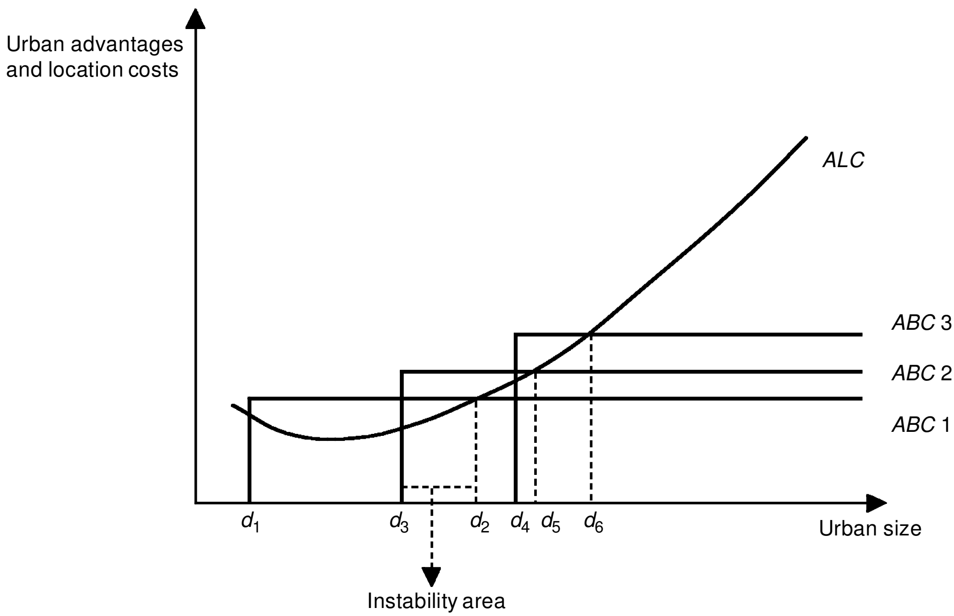

Figure 1. Efficient urban size for different urban functions. *Source:* Camagni *et al.* (1986).

comes some of the limits of the optimal city size theory, by suggesting:

—the need to replace optimal city size by an interval within which the city size is efficient—i.e. where average benefits exceed average location costs;[5]
—that the interval of efficient city size corresponds to greater urban size, the higher the economic functions developed in the city;
—that, as a result of the previous statement, the economic functions characterising the city are an important determinant of the efficient city size.

The interpretation of the SOUDY model becomes more interesting if analysed jointly with another theoretical paradigm, that of the network city (Camagni, 1993; Camagni and de Blasio, 1993).[6] The logic underlying the paradigm is that the spatial organisation in which cities operate is fundamental to understanding their efficiency, growth, factor productivity and sometimes their specialisation. While the organisational logic underlying Christaller's central place model is a territorial logic, emphasising a gravity-type control over market areas, in the network model another logic prevails. This refers to long-distance competition and co-operation regardless of the distance barrier (Camagni, 1993). While, in the more traditional analysis, transport costs and economies of scale were the principal forces shaping the spatial organisation of functions and cities, in the new logic other kinds of economies come to the fore—economies of vertical and horizontal integration, and network externalities similar to those emerging from 'club goods'.

The joint application of the SOUDY model and the network city theory implies something very important for the definition of economies of agglomeration: that size is not the only determinant of factor productivity and economies of agglomeration in large centres. The presence of higher urban functions and integration in the network of urban systems are also extremely important in explaining the size of the city. Both of these elements may permit the achievement of economies of scale, even in small cities. We now attempt to prove this statement from an empirical point of view.

3. The Indicators of City Effect and Urban Overload

Our empirical analysis is based on the consideration that in an urban area three environments exist—the physical (natural and built) environment, the economic environment and the social environment—each of them explaining in part or in combination the existence and persistence of a city. All three environments generate advantages and disadvantages—i.e. user benefits and costs for a city. All three have to be considered together, because they interact with one another and represent, or express, goals, means and constraints to human action in the city (Table 2).

The interactions between the economic and physical environments in a city are usually characterised by negative externalities. The negative effects generated by economic activity on the natural/built environment are well known and can be seen in various environmental disasters—the depletion of natural resources, noise, water and air pollution, loss of green areas, traffic congestion and intensive energy use are all negative external effects caused by excessive economic activities in cities. Given the role of cities as vehicles for economic growth, they are recognised as the places where the production of environmentally damaging gases, such as emissions of carbon dioxide, nitrous oxide or ozone, is concentrated.

On the other hand, we have to recognise that many of these negative effects are also highly visible as a consequence of the mass and high-density effect. If the same amount of economic activity were to take place in a more diffused territorial pattern, the spatial concentration of emissions would be reduced, but the absolute consumption of natural resources such as energy and land would be much greater. In other words, the concentration of activities and proximity are not only a precondition for social interaction and economic efficiency, but also are the source, up to certain levels, of increasing returns in

Table 2. City effect and urban overload

	Interaction between the economic and physical environments	Interaction between the economic and social environments	Interaction between the social and physical environments
City effect	Efficient energy use Efficient use of non-renewable natural resources Economies of scale in the use of urban environmental amenities	Accessibility to: good housing facilities; skilled jobs; social amenities; social contacts; education facilities; health services	Green areas for social amenities Residential facilities in green areas Accessibility to urban environmental amenities
Urban overload	Depletion of natural resources Intensive energy use Water, air pollution Depletion of green areas Traffic congestion noise	Suburbanisation forced by high urban rents Social friction in the labour market New poverty	Urban health problems Depletion of historical buildings Loss of cultural heritage

Source: Capello (1998b)

the use of scarce and non-renewable resources. An interesting example in this respect is the Milan metropolitan area. Although it represents 44 per cent of the population of Lombardy, Milan accounts for 33 per cent of the region's energy consumption for public lighting, 38 per cent of domestic electricity and 31.8 per cent of electricity for all uses. As a result of proximity and indivisibilities in energy consumption, the city may be an efficient user of natural resources.

The paradox is, however, that because of the concentration of environmental externalities in city areas, urban inhabitants are tempted to move out to the surrounding area. The inevitable consequence is that, although the individual level of well-being may rise, at an aggregate level in a wider territorial setting, the volume of environmental pollution will rise due to reduced advantages of scale and increased transport needs. This is a well-recognised modern social dilemma.

The interaction between the economic and social environments gives rise to specific positive and negative external effects. The positive effects stem from accessibility to social services, such as education, health, social amenities (theatres and cinemas) and highly paid jobs. Conversely, agglomeration diseconomies may cause negative external effects on the social environment through, for example, suburbanisation due to high urban rents, class segregation, new forms of poverty and inertia in social class division. Negative social externalities may negatively influence the economic sphere by generating various frictions on the labour markets, urban conflict, repulsion of potentially incoming firms, etc.

The last form of interaction concerns the advantages and disadvantages stemming from the physical and social environments. To give some examples: green areas for social amenities are environmental resources, which have a positive impact on social welfare, whereas the decay of historical buildings, loss of cultural heritage or urban health problems can create negative effects on the social environment (Camagni *et al.*, 1998).

The 'city effect' describes a situation in which agglomeration economies should be associated with positive environmental externalities and social network externalities. The most important difference from the traditional neoclassical view on optimal city size is that the advantages and disadvantages referred to relate not only to the economic environment, but also to the interaction of the three environments constituting the city. The city effect is the result of a pure economic efficiency goal, but also responds to wider policy objectives, defined as environmental equity, long-term allocation efficiency and distributive efficiency (Camagni, 1996, 1997).

According to this logic, an overload takes place when economic location costs are associated with negative social and environmental externalities. On the basis of the traditional view, we incorporate in the cost curve all negative externalities stemming from an urban location. In other words, in the case of costs, we widen the definition to all negative externalities which are encountered at the urban level and which have not been explicitly mentioned by Alonso.[7] The city effect and the urban overload resulting can be called the social city effect and social urban overload, since they incorporate all positive and negative cross-externalities which can stem from an urban location.

This approach is important since it influences the choice of the indicators used for measuring the city effect and the urban overload, shown in Table 3. These are:

—positive externalities stemming from the interaction between the economic and the physical environments. The indicators chosen here are the per capita use of energy, petrol and water.
—positive externalities stemming from the interaction between the economic and the social environments. In this area, the indicators chosen are: the share of people holding a university degree; the number of schools, of bank branches, and the supply of urban services with respect to urban population;[8] and the price of new houses per square metre. In microeconomic terms, this last factor is generally regarded as a cost, like in the Alonso type of location cost curve. In our macro-urban approach, urban rent is assumed as a proxy for urban economics and the well-being of the inhabitants, as it reflects the income and economic wealth of the city.
—positive externalities stemming from the interaction between the physical and the social environments. The indicator chosen is the square metres per capita of green areas in cities.

Each indicator has been divided by its maximum value, in order to standardise the different values and thus sum the different indices.[9] The general city effect indicator is in fact calculated as the unweighted sum of the different indices obtained; the indices refer to cross-effects between the different environments, and therefore the choice of a weighted sum would imply an arbitrary choice of weights. The first group of indices, relating to the interaction between the economic and the natural environments enters the sum with their 'complement to one' value, reflecting their negative correlation with city size.[10]

In the same way, the urban overload indicator takes into account the negative aspects of the interaction between the three environments (see Table 3)—namely:

—negative externalities stemming from the interaction between the economic and the physical environments. Here, all social costs for the natural environments have to be taken into account: per capita NO_x emission; per capita quantity of urban waste; number of vehicles per sq km.
—negative externalities stemming from the interaction between the economic and the social environments. In this area, the indicator chosen is the percentage of unemployment in the total urban population.
—negative externalities stemming from the interaction between the physical and the social environments. The indicator chosen is the number of crimes relative to the size of the urban population.

In this case also, the overall indicator is the unweighted sum of the different indicators, each divided by its maximum value before being aggregated (see Appendix).

Table 3. Statistical definition of the city effect and the urban overload indicators

	Interaction between the economic and physical environments	Interaction between the economic and social environments	Interaction between the social and physical environments
City effect indicator (*ALB*)	Energy use per capita Petrol use per capita Water use per capita	Number of graduates/ population Number of schools/ population Number of banks/ population Supply of public services/population Urban rent per square metre	Green areas in city (square metres per capita)
Urban overload indicator (*ALC*)	NO_x emissions per capita Urban waste (kg per capita) Number of vehicles per square kilometre	Unemployment/ population	Number of murders/ population

Other two indicators are necessary for the analysis, since they act as independent variables. The first concerns the types of high-order economic function developed in the city. For this, the share of private tertiary value-added produced by the city is used.[11] The second relates to the level of network integration of the city with the rest of the world: the lack of statistical information on the flows of interaction between our sample cities (duration of phone calls or number of phone calls) for these groups of cities has obliged us to choose a variable representing the number of telephone subscribers. However, the share of flows of international phone calls (both duration and number of phone calls) and the number of telephone subscribers available for a different group of cities (municipalities) in the metropolitan areas of Milan have shown a correlation equal to 0.8.

The database on which these indicators have been built contains 58 Italian cities and refers to the year 1991. The geographical area of reference is the urban agglomeration area.

4. The Measurement of the City Effect and the Urban Overload

The methodology used for the measurement of urban agglomeration economies and diseconomies is a cross-section analysis of 58 Italian cities, through an econometric model able to estimate aggregate city effects and urban overload. In particular, two functions have been estimated, an aggregate location benefit function, and a location cost function, represented respectively by the following translog functions:[12]

$$\begin{aligned}\ln ALB = {} & \ln \eta + \alpha_1 \ln D + \alpha_2 \ln FUN \\ & + \alpha_3 \ln NET + \beta_1 \frac{1}{2}(\ln D)^2 \\ & + \beta_2 \frac{1}{2}(\ln FUN)^2 + \beta_3 \frac{1}{2}(\ln NET)^2 \\ & + \delta_1 \ln D \ln FUN \\ & + \delta_2 \ln D \ln NET \\ & + \delta_3 \ln FUN \ln NET \end{aligned} \quad (1)$$

where, ALB represents the average location benefits (or average city effect); D the absolute size of the city; FUN the type of urban functions developed; and NET the network integration level achieved in the city.

For the average location costs:

$$\begin{aligned}\ln ALC = {} & \ln \eta + \alpha_1 \ln D + \alpha_2 \ln FUN \\ & + \alpha_3 \ln NET + \beta_1 \frac{1}{2}(\ln D)^2 \\ & + \beta_2 \frac{1}{2}(\ln FUN)^2 + \beta_3 \frac{1}{2}(\ln NET)^2 \\ & + \delta_1 \ln D \ln FUN + \delta_2 \ln D \ln NET \\ & + \delta_3 \ln FUN \ln NET \end{aligned} \quad (2)$$

where, ALC represents the average urban location costs, (or urban overload effect); D the absolute size of the city; FUN the type of urban functions developed; and NET the network integration level achieved in the city.

The translog function allows us to estimate the elasticity of benefits (or costs) directly with respect to any of the right-hand-side variables—that is, the percentage cost (benefits) change due to a 1 per cent change of a specific determinant, other things being equal. If D is the absolute size of the city, in order to test whether the size reduces the benefits, it is enough to estimate from (1) the following expression, and to test the sign of e_D:

$$e_D = \alpha_1 + \beta_1 \ln D + \delta_1 \ln FUN + \delta_2 \ln NET \quad (3)$$

where, e_D is the size elasticity of the urban city effect ($SECE$). Based upon equation (3), the interaction of the elasticity of benefits with the other two determinants can be studied. In particular, from (3) one can see how the size elasticity of the city effect ($SECE$) changes in relation to different urban levels, different economic functions and different levels of network integration, by estimating the following parameters:

$$e_{DD} = \beta_1 \quad (4)$$
$$e_{DFUN} = \delta_1 \quad (5)$$
$$e_{DNET} = \delta_2 \quad (6)$$

The same logic applies for the cost function, so the size elasticity of the overload effect ($SEOE$) with the other two determinants can be studied with the same methodology applied to the benefit curve.

5. The Evidence

In this part of the paper, we present the findings of the empirical analysis of 58 Italian cities. The results obtained were satisfactory from both the statistical and the interpretative point of view. From the statistical point of view, the model concerning the city effect as the dependent variable has an R^2 value of 0.48, showing an acceptable fit of the model. In relation to the urban overload as a dependent variable, the model has an R^2 value of 0.56.

The results have regard to the size elasticity of the city effect and the urban overload variables (*SECE* and *SEOE*), calculated in three specific circumstances:

—for different levels of urban size;
—for different types of function;
—for different levels of network integration.

In mathematical terms, the analysis has allowed us to calculate the slope of the average city effect function and of the average overload function with respect to urban size, for different levels of the three independent variables above. The results are summarised in Figure 2.

5.1 Urban Size: Optimal City Size Validated Once Again

The first results relate to the variable traditionally interpreted in the literature as the most important source of city effect and urban overload: urban size. Figure 2 (graphs a and b) shows the estimated city effect and urban overload functions for different levels of urban size. In economic terms, the calculated parameters reflect the elasticity of the city effect with respect to size—i.e. how the city effect and urban overload change with an increase in size of 1 per cent, for different urban sizes. The results obtained are in line with the abstract interpretation of the optimal urban size theory. In fact, the curves are 'well-behaved', showing a city effect which increases with urban size up to a certain point (approximately 361 000 inhabitants) and then decreases (parameter estimations are presented in Table A1 in the Appendix).

As far as the city effect is concerned, the results in fact show the possibility for a city to exploit:

—Economies of scale: our analysis shows that economies of scale exist for public services (like schools, public transport and banks), and also for environmental resources, like water, petrol and energy use;
—Indivisibilities of public services in general, since the larger the city, the greater the possibility of exploiting a critical mass of users.

The city effect, however, is exploited up to a certain urban size, after which its slope becomes negative. The expected congestion effects and diseconomies of scale prevail in large cities (Figure 2a).

As far as the urban overload effect is concerned, our results show a decreasing trend up to a certain urban size (approximately 55 500 inhabitants) and an increasing trend afterwards, once again in line with the traditional expectations. Two elements seem to generate this trend:

—In small cities, an economic and territorial effect: for very small cities, the results show that an increase in the physical dimension decreases urban overload, in terms of unemployment rates and all the social diseases associated with a peripheral local economy, dependent upon larger surrounding centres.
—In large cities, a negative environmental effect: for larger urban areas the results are the opposite—an increase in urban size increases the level of overload. The explanation is related to the natural environment indices contained in the general overload indicator. Large cities pollute more and generate more environmental damage than medium-sized ones; higher levels of production, linked to increasing physical urban size, are likely to mean a higher pollution density.

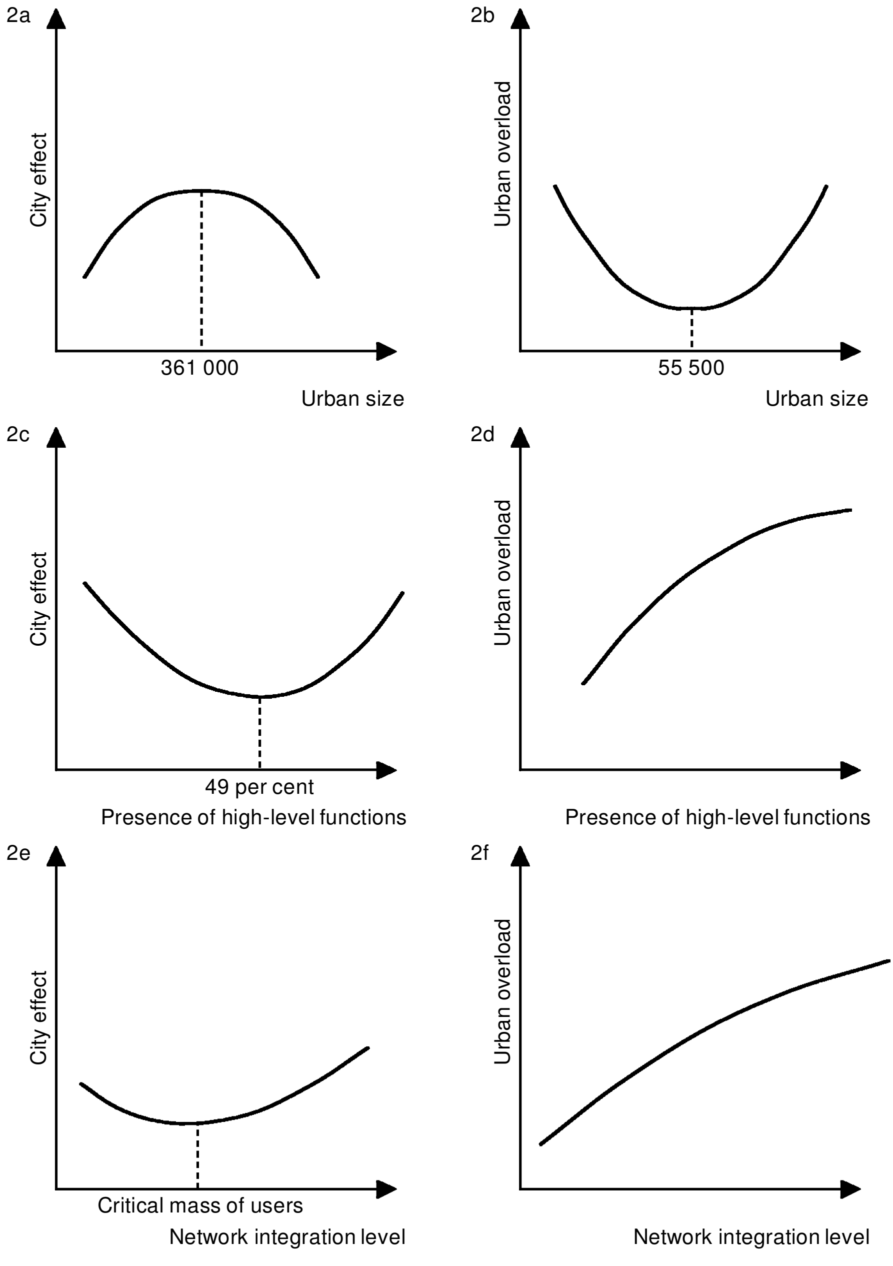

Figure 2. Estimated city effect and urban overload.

Medium-sized cities appear to have a greater endogenous capacity to keep social, economic and environmental costs under control.[13] Interestingly enough, we can see from Figure 2a that the urban overload effect reaches a minimum value at a lower urban size than the maximum city effect. This result confirms the outcome of other econometric studies, which show a lower optimal size of the city when cost-efficiency factors are utilised than when advantages are taken into consideration directly.

Another interesting element is the significance of the estimated parameters

(Table A1 in the Appendix). While positive city effects are guaranteed by a Student's *t*-value, diseconomies of scale for large urban sizes are not so evident. In the same way, urban overload is seen for large urban sizes, while for small sizes, the urban overload is decreasing. This means that urban size is important for explaining economies of scale and the considerable 'city effects' of large cities. On the contrary, other determinants are necessary to explain fully the diseconomies of scale and decreasing overload effects of small cities.

The results presented above hold for cities of different physical size, but characterised by the same type of economic function and the same level of network integration. All things being equal, we can say that the city effect and the overload curves have the slope presented in Figure 2. But as we mentioned in the conceptual part of this paper, other things are not equal—in the sense that both the type of economic activity developed in cities and the level of network integration are different in the different cities examined and thus influence the results, as we shall see in the following.

5.2 The Type of Economic Functions: The Validity of the SOUDY Model

Our main conclusion up to now is that, *ceteris paribus*, for the same type of economic function and level of network integration with other cities, the city effect can be exploited up to a certain level of urban size. By the same token, urban overload can be kept under control up to a certain level.

The picture changes when the analysis is made on the basis of the different types of economic function which can characterise a city. The results are quite interesting. As far as the city effect is concerned, these results are in line with the conclusions suggested by the SOUDY model. The estimated curve confirms the theoretical hypotheses of the SOUDY model (Figures 2c and 2d): higher-order functions guarantee a greater city effect, due to the positive returns generated.

Another result seems to emerge from the analysis, although from the statistical point of view some doubts remain (see Table A2 in the Appendix). This is that a minimum threshold of high-order tertiary functions has to be achieved before increasing returns to urban scale manifest themselves. Only if the city achieves a substantial share of tertiary activities (49 per cent of its total activities), can it exploit the advantages of urban size.

The urban overload effects increase at a decreasing rate when there is a strong presence of high-level functions. This means that the increase in tertiary activities tends to entail congestion and location costs, but that this negative aspect does not occur in a disruptive and uncontrollable way, as in the case of increasing urban size (Figure 2d). The urban overload effect increases at a decreasing rate, which indicates that higher-order functions produce economic development and also local congestion costs, but with a decreasing marginal productivity, and thus in a more controlled way. The decreasing order of magnitude with which overload is generated in the presence of higher-order functions may be explained by the following:

—From the point of view of environmental indices involved in the overload indicator: tertiary activities are by definition less-polluting activities than industrial ones; higher-income communities (stemming from economies based on higher-order functions and higher profit levels) treat the environment as a luxury good, due to the emergence of new social values with respect to the environment (Camagni, 1996). These results are in line with the apparent paradox described by Baldwin (1995) in his provocative and very important statement: "sustainability requires growth".

—From the point of view of the economic and social costs involved in the overload indicator, tertiary activities have in the past decade been characterised by high employment rates, and thus a higher percentage of these activities in a city guarantees a lower level of social disease resulting from the lack of jobs.

—From the economic point of view, the increasing overload is the consequence of the broad economic development that higher-order functions generate. Higher-order functions stimulate strong economic development because of their capacity to generate greater multiplicative effects than more traditional functions. This is a mechanism which has been widely overestimated in the empirical analysis.

5.3 The Level of Network Integration: The Existence of Network Externalities

The results of the size elasticity of the city effect and urban overload for different levels of network integration produce an interesting picture (Figures 2e and 2f).

As far as the city effect is concerned, the city effect decreases up to a certain level of network integration, when it starts to increase (see Table A3 in the Appendix). These results are stimulating, since they suggest that

—For low levels of network integration, advantages of autarchy and independence take place, although these results seem to be statistically weak.
—When the network integration process starts, cities are vulnerable and are weak partners, risking in general being exploited by the network, rather than exploiting the advantages of a network. This result is in line with the general idea that being part of a network does not necessary mean obtaining advantages from it (Camagni, 1993; Capello, 1994). As expected, this is true up to a certain level of network integration.
—After a certain threshold level, the city is able to exploit the advantages associated with the interconnected economy and network externality advantages are in full operation. (See the vast literature existing on network externalities, including Hayashi, 1992; and Rohlfs, 1974.) Through the network, the city is able to exploit more dispersed information collection, the acquisition of more know-how and more qualified input factors, as well as a wider market for final goods.

For the urban overload effect, the picture which emerges is similar to that for different levels of high-order functions (Figure 2f). When the level of network integration increases, urban overload increases, too. This is what would be expected: higher levels of network integration stimulate more economic activities and generate higher city effects, but with the negative counterpart of an increasing overload.

What is rather interesting is that urban overload has decreasing growth rates. Again, this result is different from the exploding situation which occurs when the city size is taken into consideration. As the level of network integration increases, positive mechanisms come into effect, decreasing the size elasticity of the overload:

—From the economic point of view, as the city increases its ability to exploit network externality advantages, unemployment and social diseases related to a stagnating economy decrease.
—From an environmental point of view, the city economy increases via the network, by keeping under control the local pressure in terms of environmental costs. The networked city can reorganise its production system by decentralising the most-polluting and less-attractive functions, while specialising in higher-order functions like control and decision-making processes. In this way, it benefits from the advantages of an expanding economy, while keeping environmental costs and local pressures under control.

6. Conclusions

A critical approach to the theory of optimal city size has produced the following findings. The influence of urban size exists and is important, but cannot be efficiently assessed without overcoming some of the limitations imposed by the theory. It is not a problem of *optimal* city size, but of *efficient* size, which largely depends on what the city produces, how it produces and the way in which it co-operates within the urban system. Urban size inevitably influences location costs and benefits; however, the same also holds for its level of specialisation and integration with the urban system. Economies of scale exist, *ceteris paribus*, but turn into diseconomies after a certain urban dimension. However, with increasing size, the preconditions increase for developing structural changes allowing a greater mix of higher urban functions.

The present work has confirmed these hypotheses from an empirical point of view. In particular, the type of economic function and the spatial organisation within which the city is integrated appear to be strategic elements for the definition of location benefits and costs, analysed in relation to all aspects constituting the city—i.e the social, environmental and economic aspects.

Our analysis has important normative consequences. Since it is difficult to envisage a large city having a strong city effect without facing high overload costs, local urban policies are absolutely vital and play a significant role in the definition of the growth potential of our cities. These policies should focus, among other things, on upgrading the economic functions within the city, as well as the development of linkages outside the city, such as alliances, co-operation agreements, advanced international transport and telecommunications infrastructure. All of these elements are undoubtedly important for guaranteeing the survival of a modern city.

Notes

1. Alonso (1971) stressed the mistaken tendency of many authors to look for optimal city size only by minimising the location cost function. As he argued, this would be sensible only if output per capita were constant (p. 70).
2. Carlino (1980) provides a criticism of Chinitz' analysis, and demonstrates on a sample of 65 US towns that economies of scale, both internal and external to the firm, play a role in the definition of urban productivity.
3. We mean 'superseded' in a conceptual way, rather than chronologically, in that the Christaller model predates the optimal city size theory.
4. Mills (1993) argues that in this field econometric analyses are more advanced than the theoretical framework.
5. Richardson (1972) has already suggested replacing the concept of optimal city size with an efficient interval of urban size in which urban benefits are greater than location costs.
6. Camagni (1993) theorised the concept and applied it to urban systems. The same concept has already been applied to many fields, such as the behaviour of the firm and macroeconomic organisational behaviour. For a review of the concept, see Capello (1998a); for the policy implications, see Capello and Rietveld (1998).
7. Richardson (1972, p. 31) was the first to deal explicitly with the environmental costs associated with city size, and with the problem of valuing the negative aspects of city size not registered in market prices.
8. In this case, the information used is in fact the number of people using public services—i.e. the demand, which is used as a proxy for the supply, as the latter data are unavailable.
9. Many methods exist for standardising the variables. The one chosen has been applied by Biehl (1986), where an aggregate physical infrastructure index was obtained as the sum of different indices of different physical infrastructures.
10. The negative correlation between the consumption of energy, petrol and water and city size has been estimated on the same sample in Capello (1998b).
11. Value-added is measured at the provincial level, since an estimate at the urban level is not available.
12. Considerable discussion exists around the need to impose restrictions on parameters of production functions, so that they are able to approximate the stylised facts of economic behaviour that neoclassical economists generally agree characterise the real world (Chambers, 1988). On the production function, these restrictions relate to: monotonicity (and strict monotonicity) of the production function; quasi-concavity (and concavity) of the production function; weak essentiality (strict essentiality); closed and non-empty input requirement set for all outputs; the production function is finite, non-negative, real valued for all non-negative and finite inputs; and, the production function is everywhere twice-continuously differentiable.

   It has been emphasised, however, that these restrictions have generally been rejected in econometric analyses (Evans and

Heckman, 1984). The decision to impose these restrictions even though rejected by empirical data, as is often the case, is not necessary in any direct comparison between our analysis and other studies where the restrictions have been imposed. Moreover, in our particular case, the economic reasoning behind these analytical restrictions is difficult to accept *a priori*. Our production function is in fact a quasi-production function (where inputs are more than the conventional capital and labour inputs), representing an aggregate economic behaviour which may not follow the same economic rules as those imposed by the neoclassical individual firm's behaviour.

13. The trend of our location cost curve differs from the traditional location cost curve of Alonso. Also, our more macro type of cost function, where the social cost to the environment and the social disamenities associated with urban size are contained, has an increasing shape, but only from a certain urban size. Before that level, other mechanisms, which are not considered in Alonso's microeconomic type of location cost curve, take place and allow small cities to increase their size without paying in terms of the economic, environmental and social diseases that physical growth may imply.

References

ALONSO, W. (1964) *Location and Land Use: Towards a General Theory of Land Rents*. Cambridge, MA: Harvard University Press.

ALONSO, W. (1971) The economics of urban size, *Papers and Proceedings of the Regional Science Association*, 26, pp. 67–83.

BALDWIN, R. (1995) Does sustainability require growth?, in: I. GOLDIN and A. L. WINTERS (Eds) *The Economics of Sustainable Development*, pp. 51–82. Cambridge: Cambridge University Press.

BERG, L. VAN DEN, DREWETT, R. and KLAASSEN, L. *ET AL*. (1983) *Urban Europe: A Study of Growth and Decline*. Oxford: Pergamon.

BIEHL, D. (1986) *The Contribution of Infrastructure to Regional Development*. Regional Policy Division. Brussels: European Community.

CAMAGNI, R. (1992) *Economia Urbana: Principi e modelli*. Rome: NIS.

CAMAGNI, R. (1993) From city hierarchy to city network: reflections about an emerging paradigm, in: T. LASCHMANAN and P. NIJKAMP (Eds) *Structure and Change in the Space Economy*, pp. 66–90. Berlin: Springer Verlag.

CAMAGNI, R. (1996) Lo sviluppo urbano sostenibile: le ragioni e i fondamenti di un programma di ricerca, in: R. CAMAGNI (Ed.) *Economia e pianificazione della città sostenibile*, pp. 13–51. Bologna: Il Mulino.

CAMAGNI, R. (1997) *Cities in Europe: Globalisation, Sustainability, Cohesion*. Report prepared for the Italian Presidency of the European Union.

CAMAGNI, R. and BLASIO, G. DE (Eds) (1993) *Le reti di città*. Milan: Franco Angeli.

CAMAGNI, R., CAPELLO, R. and NIJKAMP, P. (1998) Towards sustainable city policy: an economy-environment technology nexus, *Ecological Economics*, 24, pp. 103–118.

CAMAGNI, R., CURTI, F. and GIBELLI, M. C. (1985) Ciclo urbano: le città tra sviluppo e declino, in: G. BIANCHI and I. MAGNANI (Eds) *Sviluppo multiregionale: teorie, problemi, metodi*, pp. 197–234. Milan: Franco Angeli.

CAMAGNI, R., DIAPPI, L. and LEONARDI, G. (1986) Urban growth and decline in a hierarchical system: a supply-oriented dynamic approach, *Regional Science and Urban Economics*, 16, pp. 145–160.

CAPELLO, R. (1994) *Spatial Economic Analysis of Telecommunications Network Externalities*. Aldershot: Avebury.

CAPELLO, R. (1998a) Economies d'echelle et taille urbaine: théorie et etudes empiriques révisités, *Révue d'Economie Régionale et Urbaine*, 1, pp. 43–62.

CAPELLO, R. (1998b) Urban return to scale and environmental resources: an estimate of environmental externalities in an urban production function, *International Journal of Environment and Pollution*, 10, pp. 28–46.

CAPELLO, R. and RIETVELD, P. (1998) The concept of network synergy in economic theory: policy implications, in: K. BUTTON, P. NIJKAMP and H. PRIEMUS, (Eds) *Transport Networks in Europe*, pp. 57–83. Cheltenham: Edward Elgar.

CARLINO, G. (1980) Constrast in agglomeration: New York and Pittsburgh reconsidered, *Urban Studies*, 17, pp. 343–351.

CHAMBERS, R. G. (1988) *Applied Production Analysis*. Cambridge: Cambridge University Press.

CHINITZ, B. (1961) Constrast in agglomeration: New York and Pittsburgh, *American Economic Review, Papers*, 51, pp. 279–289.

CHRISTALLER, W. (1933) *Die Zentralen Orte in Suddeuschland*. Jena: Gustav Fischer Verlag.

CHRISTENSEN, L., JORGENSON, D. and LAU, L. (1973) Transcendental logarithmic production frontiers, *Review of Economics and Statistics*, 55, pp. 28–43.

EVANS, D. and HECKMAN, J. (1984) A test of suadditivity of the cost function with application to the Bell system, *American Economic Review*, 74, pp. 615–623.

FUJITA M. (1985) *Urban Economic Theory: Land Use and City Size*. Cambridge, MA: Cambridge University Press.

HAYASHI, K. (1992) From network externalities to interconnection: the changing nature of networks and economy, in: C. ANTONELLI (Ed.) *The Economcs of Information Networks*, pp. 195–216. Amsterdam: North-Holland.

HENDERSON, J. (1985) *Economic Theory and the Cities*. Orlando, FL: Academic Press.

HENDERSON, J. (1996) Ways to think about urban concentration: neoclassical urban systems vs. the new economic geography, *International Regional Science Review*, 19, pp. 31–36.

MILLS, E. (1993) What makes metropolitan areas grow?, in A. SUMMERS, P. CHESHIRE and L. SENN (Eds) *Urban Change in the United States and Western Europe*, pp. 193–216. Washington, DC: The Urban Institute.

RICHARDSON, H. W. (1972) Optimality in city size, systems of cities and urban policy: a sceptic's view, *Urban Studies*, pp. 29–48.

ROHLFS, J. (1974) A theory of interdependent demand for a communication service, *The Bell Journal of Economics and Management Science*, 5, pp. 16–37.

WORLD RESOURCES INSTITUTE (1994) *People and the Environment*. Oxford: Oxford University Press.

Appendix. Results of the Estimated Parameters

Table A1. Size elasticity of city effect and urban overload for different levels of urban sizes

	Coefficient	Student's *t*
City effect		
Average city	0.043	2.61
Largest city	−0.014	−0.21
Smallest city	0.102	1.91
Urban overload		
Average city	0.10	4.29
Largest city	0.31	3.40
Smallest city	−0.10	−1.42

Table A2. Size elasticity of city effect and urban overload for different levels of high-order economic function

	Coefficient	Student's *t*
City effect		
Average city	0.024	1.13
City with the highest percentage of high-order functions	0.074	2.85
City with the lowest percentage of high-order functions	−0.039	−0.70
Urban overload		
Average city	0.14	6.35
City with the highest percentage of high-order functions	0.06	1.97
City with the lowest percentage of high-order functions	0.24	4.80

Table A3. Size elasticity of city effect and urban overload for different levels of network integration

	Coefficient	Student's *t*
City effect		
Average city	0.04	2.61
City with the highest level of network integration	0.13	4.37
City with the lowest level of network integration	−0.04	−1.27
Urban overload		
Average city	0.13	4.64
City with the highest level of network integration	0.12	2.60
City with the lowest level of network integration	0.14	2.92