Equilibrium vs. optimal city size: evidence from Italian cities

Roberto Camagni, Roberta Capello and Andrea Caragliu*

ABC Department, Politecnico di Milano, Piazza Leonardo 32, 20133 Milan (MI), Italy E-mail: roberto.camagni@polimi.it E-mail: roberta.capello@polimi.it E-mail: andrea.caragliu@polimi.it *Corresponding author

Abstract: In this paper, the stylised assumption that one single 'optimal' city size exists for all cities – achieved when marginal location costs equal marginal location benefits - is abandoned, as well as the opposite view that each city operates on its own cost and production curves, defining a specific optimal size. Instead, this work maintains the comparability among cities and demonstrates that urban specificities in functions performed, quality of life, industrial diversity and social conflicts shift up and down the benefits and costs linked to pure physical size, leading to different 'equilibrium' sizes for cities. A model of equilibrium urban size is set up, and empirically estimated on a sample of 103 Italian cities with data at NUTS3 level. Empirical results verify the empirical model on the analysed sample; results hold both with standard OLS estimates as well as with the use of instrumental variables in order to correct for the possible endogeneity of some of the variables in the model. Differences between predicted and real city size are interpreted with good or bad governance, thereby suggesting future strategies for more efficient urban planning.

Keywords: equilibrium city size; spatial equilibrium; urban functions; city networks.

Reference to this paper should be made as follows: Camagni, R., Capello, R. and Caragliu, A. (2014) 'Equilibrium vs. optimal city size: evidence from Italian cities', *Int. J. Global Environmental Issues*, Vol. 13, Nos. 2/3/4, pp.170–188.

Biographical notes: Roberto Camagni has been the Head of the Department for Urban Affairs at the Presidency of the Council of Ministers, Rome, under the Prodi Government, 1997–98. He is the President of the GREMI, Paris, Sorbonne, since 1987, the President of ERSA, 1989–1992 and the President of AISRE, 1989–1992. He has published a textbook on urban economics: *Economia urbana: principi e modelli teorici* (with translations in French and Spanish). He is the author of many publications on urban and regional development, regional and urban policy. He is the recipient of the 2010 ERSA-EIB prize for his scientific contributions to spatial economics.

Roberta Capello is a Professor of Regional and Urban Economics at Politecnico di Milano. She received her PhD in Economics at the Free University in Amsterdam in 1993 and First Degree in Economics at Bocconi University, Milan in 1986. She is a past President (2009–2010) and has been member of the Organizing Committee (2003–2008) of the Regional Science Association

Copyright © 2014 Inderscience Enterprises Ltd.

International (RSAI). She has been a member of the organising committee and Treasurer of the European Regional Science International (1999–2004). She has been a Visiting Research Fellow in the Centre for Urban and Regional Development Studies of the University of Newcastle (1988–1989). She has published the textbook *Regional Economics* (Routledge, 2007). She is the author of many publications on urban and regional development.

Andrea Caragliu is an Assistant Professor of Regional and Urban Economics at Politecnico di Milano. He obtained his PhD in Management, Economics and Industrial Engineering at Politecnico di Milano in 2010 (Merit Prize of the EU Committee of the Regions and the Diploma d'Onore AISRe for the best doctoral dissertation in Regional Science). He received his Master's (2005) and Bachelor's (2003) degree in Economics at Bocconi University. He has been a visiting student at the Spatial Economics Department of the Free University Amsterdam, 2008–2009. He has published on various international refereed journals (*Annals of Regional Science, Applied Economics, Papers in Regional Science, Regional Studies*, and *TESG*).

This paper is a revised and expanded version of a paper entitled 'One or infinite optimal city sizes? In search of an equilibrium size for cities' presented at International Conference 'The Wealth of Nations, the Wealth of Cities', Naples, 6–7 September 2012.

1 Introduction: the debate on optimal city size

In spite of the recent burgeoning literature on city size, urban performance and structure of urban systems, the present state of theoretical and empirical studies still presents many shades and incomplete research programmes, calling for further efforts in the conceptual interpretation and bridging of separated strands of theoretical models.

The debate on optimal city size has a long-standing tradition. A large consensus exists in the literature on the fact that some net increasing returns exist up to a certain urban size: beyond that size, opposite mechanisms are at work, translating positive externalities into negative ones, and transforming economies into diseconomies. In this case, location costs increase, overcoming location benefits. As in the case of any other resource used in an intensive way, net decreasing returns to scale appear beyond a certain size.

In his seminal paper on 'The economics of urban size', Alonso (1971) made two provocative but opposite propositions that puzzled and challenged scholars for many years. On the one hand, he justified scientific research responding to the questions about the existence of an 'optimal' city size in terms of population, opening a wide strand of subsequent empirical research. On the other hand, he presented the opposite view that optimality – in any sense – *will vary from city to city, from society to society* [Alonso, (1971), p.81], destroying the first research programme – except on a case-by-case basis – and the possibility of intervening normatively on the proper structure of a city system.

Richardson confirmed a 'sceptic's view', suggesting the existence of other determinants influencing urban agglomeration economies, beyond sheer physical size (Richardson, 1972). Consequently, for a long time scientific efforts were redirected outside the problem of searching for an 'optimal' size and mainly dedicated to the identification of urban specificities that affect urban costs and benefits.

The optimal condition for the entire population of the system, urban and not urban, is reached when urban marginal costs equal marginal benefits (to size increases) (Alonso, 1971; Richardson, 1978). As all cities are assumed to operate with similar cost and production functions, the possibility opens up to give the floor to the real world, through empirical econometric analyses; but this implies the search for a single urban size, optimising costs, or incomes or net urban benefits. Yet, this contradicts the stylised fact of the existence of a wide spectrum of city sizes, apparently in (a static or dynamic) equilibrium.

Therefore, over time, many criticisms arose against the optimal city size theory. These include the observations that cities perform different functions, are characterised by different specialisations and, consequently, operate with different production functions (Henderson, 1974, 1996). In the words of Richardson (1972, p.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".

The bizarre fact is that Alonso himself was perfectly aware of this paradox, and, in fact, in his seminal paper (Alonso, 1971) made two provocative but at the same time mutually exclusive and antithetical propositions. On the one hand he justified the theoretical relevance of the question concerning the right size of cities ('how big is too big?'; and 'how big is big enough?'), illustrated by his well-known graphic on urban costs and benefits, opening a wide field of empirical inquiries on 'optimal' city size. On the other hand, in the final paragraph, he took the opposite view that cost and benefit curves (as those presented in the graphic) "will vary from city to city, from society to society" [Alonso, (1971), p.81]. This second proposition definitely destroys the first research programme – except for a case-by-case approach – pointing out that different cities operate on different cost and production curves.

A similar paradox and theoretical challenge – one single vs. infinite urban sizes – was faced by the urban land use theory (the von Thunen-Alonso-Muth-Fujita trajectory), when it came to the issue of the structure of the entire city-system (the so-called 'open city' model). In fact, the abstract equilibrium model, even in its more complex formulations, ended up with an urban landscape made of identical cities, with a single equilibrium – and optimal – size [Fujita, (1989), ch. 5; Camagni, (2011), ch. 3]. The solution proposed by Henderson (1974, 1985), though, leading to infinite city sizes as a consequence of the assumption of perfect specialisation of cities in different sectors, does not present the same theoretical compactness of the main model, and may look as an ad hoc proposal.¹

A more recent synthesis of the same theoretical trajectory was recently provided by Glaeser (2008), where the possibility of justifying the existence of cities of different size was reached through a double path: either including in the utility function of households, in addition to consumption levels and size of apartment, an index capturing urban amenities (or urban distress) for single cities, along the lines of Rosen (1979) and Roback (1982), defining an equilibrium condition in which amenities offset the reduced utility deriving from a wider and more expensive city; or including in the firms' production function exogenous elements as a city-specific productivity level (ch. 3), a term capturing agglomeration economies or other elements leading to efficiency increases like cluster effects, learning processes, skills mix, possibility of shifting employment from low to high productivity sectors (ch. 4). The model maintains the capability of building a micro-funded condition of spatial equilibrium with cities of different sizes, considering

elements of geographical differentiation with a clear empirical relevance, but has to include in a general equilibrium model specificities concerning the single cities.

Besides, traditional neoclassical models failed to consistently incorporate agglomeration economies into 'open-city' models of urban structure along the von Thünen-Alonso-Muth-Fujita paradigm, trapped between a unique size for all cities vs. a differentiated size based on the seemingly ad-hoc assumption of a perfect and diversified urban specialisation [Henderson, 1974; Fujita, 1989; Camagni, (2011), ch. 2]. Furthermore, they lacked a true spatial dimension (Fujita et al., 1999). On the other hand, agglomeration economies models proposed by the New Economic Geography (Krugman, 1991, 1995), given their high simplification of the urban system, reduced to city pairs and one foreign market, seem unable to interpret an urban system in which cities have different sizes (Henderson, 1995, 1996).

Inside the alternative between only one optimal size and infinite sizes, different solutions are provided by central place models and by Zipf's (1949) rank-size rule. In the first case, according to the different functional mix present in each urban rank and the different 'range' and 'threshold' attached to the market area of each function, higher rank cities are expected to show a wider size with respect to lower rank ones, while cities belonging to the same rank share the same size. Of course, this outcome derives from second generation central place models, such as Beckmann and McPherson (1970), incorporating a demographic size for cities and market areas, while the earlier models of Christaller and Lösch had no physical dimension, being defined only by a vector of presence/absence of different functions [Camagni, (2011), ch. 4]. The general forces behind the polycentric structure of the urban system refer to spatial efficiency principles, and have to be found in the interplay between search for production efficiency through increasing returns to scale and agglomeration economies on the one hand and minimisation of spatial interaction costs on the other (Valvanis, 1955; Fujita and Thisse, 1996).

These same forces are called into play in the interpretation of Zipf's law of urban size distribution, following Simon (1955), but the true spatial nature of these and other explanations are rightly questioned (Krugman, 1995; Reggiani and Nijkamp, 2012), as not just geographical space – distances – but also specificities of single cities are not considered in this family of models. Krugman (1995, p.246) provocatively considered Simon's explanation as a 'nihilistic story', but his pretention that the new economic geography actually provides a general equilibrium interpretation of the structure of urban sizes is far from convincing, as Henderson (1995, p.271) clearly stated: "the new economic geography has been unable to model a system or hierarchy of cities in which cities are of different sizes".²

More interesting seems to be a recent effort to link Zipf's law with the entropy model of spatial interaction: maximising an entropy function of the size distribution of centres – governing the diffusion force – subject to the constraint $\Sigma_r P_r r = K$ – which could be interpreted as the cost of non-concentration, or the agglomeration potential of the system – an equation similar to the rank-size rule is derived (Reggiani and Nijkamp, 2012). Still, even if the population stock in each centre could be interpreted as the result of migration flows among centres, the absence of an explicit measure of distance makes the spatial result of the model abstract.

As will be later explained, this paper proposes an intermediate solution to the problem of an equilibrium size of cities, in a different sense with respect to central place and

rank-size models. The model presented in this paper is conceptually similar to Glaeser (2008), as it aims at defining differentiated equilibrium city sizes, but differs methodologically, being based on a meso-approach as in Alonso (1971) considering single cities as optimising units, rather than on a micro-funded approach. Cities are supposed to share the same complex cost and production functions with heterogeneous, substitutable factors linked not just to economic functions but to other context conditions. Therefore, each of them maintains its specificity and consequently its 'equilibrium' size, but comparability (and the possibility to perform cross-sectional analyses) is maintained, along with the possibility to devise policy strategies for urban growth or containment.

Much of the recent literature on Zipf's law (Gabaix, 1999; Reggiani and Nijkamp, 2012), though inspective and relevant in the identification of the basic drivers of urban size distribution and in devising linkages with spatial interaction models, inherently fails to explicitly incorporate space and cities' specificities into the urban hierarchy. Analogously, the huge literature on agglomeration forces and urban performance (Catin, 1991; Glaeser and Mare, 2001; Glaeser and Resseger, 2009) seems unable to answer the two Alonso questions.

This paper contributes to the explanation of equilibrium sizes of cities, taking up Alonso's challenge and confronting with each other the wide array of approaches and models that have dealt with the issue of the structure of urban systems. This implies that an intermediate position is assumed between the idea of a single, 'optimal' size for any city and that of an infinite plurality of 'optimal', but unexplained sizes, one for each city. Cities are assumed to be comparable, sharing common costs and benefits functions, thereby allowing cross-sectional empirical analyses and considering other determinants of urban benefits and costs beyond pure city size. Each city maintains its own specificity and uniqueness, and consequently is attributed its own 'equilibrium' size in an econometric model directly derived equating marginal costs and benefits to urban size.

Both conventional and more innovative determinants of agglomeration costs and benefits are considered. More conventional approaches highlighting elements like amenities and quality of life, urban atmosphere, human capital and skills, and agglomeration economies on the benefits side, and social conflicts/malaise, and costs of the city in general (urban land rent), on the costs side. A more recent and unconventional literature encompasses the role of urban functions (embedded in dynamic urban models), the role of the city within inter-urban cooperation agreements (the so called city-network paradigm) on the benefits side, and the loss of efficiency and sustainability brought in by dispersed urban forms, on the costs side.

The paper is structured as follows. A model for equilibrium city size is presented (Section 2) and then it is subject to empirical test on a sample of 103 Italian cities (Section 3). An 'equilibrium' city size for each single city of our sample is obtained and compared to actual size. Finally, Section 4 concludes; differences with respect to the equilibrium size are interpreted – beyond being the sign of our ignorance – as the result of an efficient (un-efficient) urban governance and can suggest future strategies in urban planning.

2 A model for equilibrium city size

The tangible and intangible elements highlighted by the literature as sources of urban development and size are used in an equilibrium model of urban size, which finds its

roots in the neoclassical stream of location choice models following the von Thünen-Alonso-Fujita trajectory. The theoretical model is based on a previous contribution by the same authors (Camagni et al., 2013), and relates to similar works in the literature on equilibrium city size (see for instance Royuela and Suriñach, 2005). In this class of models, the location choices of single individuals (firms) are driven by utility (profit) maximisation achieved when marginal location costs equal marginal location benefits (Alonso, 1960; Muth, 1969; Fujita, 1989).³

The following implicit total urban cost function is assumed, where total location costs depend on the physical size of the city (*size*), and the intangible aspects highlighted by the literature, namely social costs (*malaise*), costs due to dispersed urban form (*sprawl*) and in general, the costs of the city, captured by urban land rent (*urban rent*):

$$C = f(size, rent, malaise, sprawl)$$
(1)

In turn, total benefits depend on the physical size of the city (*size*), on the intangible aspects highlighted by the conventional literature – namely quality of life (*amenities*), creativity (*diversity*), urban atmosphere (*density*) – and on the unconventional ones – namely quality of economic functions performed (*functions*) and inter-urban networks (*networks*) as in the following implicit function:

$$B = f(size, amenities, diversity, density, functions, networks)$$
(2)

Physical size acts on both costs and advantages, and thus assumes a dual nature (i.e., it represents a joint source of positive as well as negative externalities for city dwellers).

A standard Cobb-Douglas specification for both functions is adopted. This specification is more tractable than most others, and allows to avoid the implausible assumptions about the elasticity of the function's arguments (Uzawa, 1962).

Equations (1) and (2) therefore, become, respectively:

$$C = size^{\alpha} rent^{\beta} malaise^{\delta} sprawl^{\gamma}$$
⁽³⁾

and

$$B = size^{\kappa} amenities^{\zeta} diversity^{\vartheta} density^{\chi} functions^{\mu} networks^{\nu}$$
(4)

In order to increase the tractability of the model and without losing generality, each parameter is assumed to be bounded in the interval (0, 1), but for size exponent in the cost function (α) , which is, following Alonso, assumed to be larger than one, reflecting an exponential cost function with respect to city size. Analytically, these assumptions lead to the following conditions:

$$\frac{\partial C}{\partial size} = \alpha size^{\alpha - 1} rent^{\beta} malaise^{\delta} sprawl^{\gamma} > 0,$$

$$\frac{\partial^{"}C}{\partial size^{"}} = \alpha(\alpha - 1)size^{\alpha - 2} rent^{\beta} malaise^{\delta} sprawl^{\gamma} > 0$$
(5)

and

$$\frac{\partial B}{\partial size} = \kappa size^{\kappa - 1} amenities^{\zeta} diversity^{\vartheta} density^{\chi} functions^{\mu} networks^{\nu} > 0,$$

$$\frac{\partial^{"}B}{\partial size^{"}} = \kappa(\kappa - 1)size^{\kappa - 2} amenities^{\zeta} diversity^{\vartheta} density^{\chi} functions^{\mu} networks^{\nu} < 0$$
(6)

With these assumptions, marginal costs and benefits are well-behaved with respect to the traditional optimal city size theory.

This model is closed by assuming spatial equilibrium across the urban system. As people can freely move across space in order to look for better living conditions (i.e., they can look for cities characterised by higher benefits or lower costs), in equilibrium the city must satisfy the condition in which marginal location costs equal marginal benefits (MLC = MLB), thus maximising utility of people, profits of firms and aggregate national efficiency. The assumption of spatial equilibrium across European countries may attract criticism. For instance, Cheshire and Magrini (2006) find that it may hold at most only within EU countries. However, recent empirical work (e.g., Rappaport, 2004) provides the theoretical rationale for the fact that "even very small frictions to labor and capital mobility along with small changes in local productivity or local quality of life suffice to cause highly persistent population flows" [Rappaport, (2004), p.554].

Analytically, this implies the following condition must hold:

$$\frac{\partial C}{\partial size} = \frac{\partial B}{\partial size} \tag{7}$$

After several intermediate passages,⁴ we obtain an estimable functional form [equation (8)]:

$$\ln(size) = \frac{\ln\binom{\kappa}{\alpha}}{(\alpha - \kappa)} + \frac{\zeta}{(\alpha - \kappa)} \ln(amenities) + \frac{\vartheta}{(\alpha - \kappa)} \ln(diversity) + \frac{\chi}{(\alpha - \kappa)} \ln(density) + \frac{\mu}{(\alpha - \kappa)} \ln(functions) + \frac{\nu}{(\alpha - \kappa)} \ln(networks)$$
(8)
$$- \frac{\beta}{(\alpha - \kappa)} \ln(rent) - \frac{\delta}{(\alpha - \kappa)} \ln(malaise) - \frac{\gamma}{(\alpha - \kappa)} \ln(sprawl)$$

Equation (8) shows that the equilibrium size of the city, and in particular the physical equilibrium size of the city, depends on city-specific characteristics.⁵ Conventional elements like amenities, human capital, industrial diversity, and unconventional elements, such as the presence of high-level functions and urban networking, can act as 'shifters', moving upward the marginal benefit function and achieving, *coeteris paribus*, a higher physical equilibrium size. On the other hand, elements like sprawl, social conflicts and high urban rents can push upward the marginal location costs, reducing the physical equilibrium size.

While for the conventional elements there is a large consensus on the decision whether they impact on costs or benefits, for the unconventional elements guesses are made on the fact that sprawl represents a costs, with a notable exception in Glaeser and Kahn (2004), and high level functions and networking act instead on benefits. Their real effects will be empirically tested by estimating equation (8) using traditional econometric models, and find the elasticity of the equilibrium size to each single urban feature. The

estimated equilibrium sizes for each city can be compared to the actual population, in order to reveal whether in reality each city of the sample exceeds (or is lower than) its equilibrium size. The result – represented by the econometric residual – can be explained by inefficient (efficient) urban governance, and can shed light to strategies of urban planning (see Section 4).

3 Empirical evidence on Italian cities

3.1 The dataset for empirical estimates

This paper presents a novel application of the model described in Section 2. In fact, in Camagni et al. (2013) we tested the validity of the theoretical model on a sample of 59 European Functional Urban Areas (FUAs). In the present paper, the empirical test is carried out on the universe of Italian NUTS3 regions, which, for the sake of this applied analysis, proxy for FUAs in the Italian context. This assumption is also used in the construction of the FUA Urban Audit data base compiled by EUROSTAT; therefore, the estimates shown in this paper can be compared with those presented in Camagni et al. (2013).

The use of a more homogenous spatial context, furthermore, could offer a more insightful perspective on the way the theoretical model holds and is capable of predicting the correct equilibrium urban size of each city. This is also based on the absence of wide cross-country differences, in particular in terms of different urban systems, which characterizes a wider, cross-country, dataset.

Type of variable	Class of variable	Variables	Indicator	Years	Source of raw data
Dependent	Physical size of the city	Dimension	FUA population	2010	ISTAT
Independent	Traditional urban benefits				
	Quality of life	Amenities	Available hotel rooms per 1,000 inhabitants	2002	ISTAT
	Urban creativity	Diversity	Sectoral diversity index obtained as the complement to 1 of a Hirschman- Herfindahl index calculated on the industrial composition of the labour force	2001	Censimento nazionale dell'Industria e dei Servizi (Industrial census) 2001
	Agglomeration economies	Density	Population density	2005	ISTAT

 Table 1
 The variables, the indicators and their sources

Source: Authors' elaboration

Type of variable	Class of variable	Variables	Indicator	Years	Source of raw data
Independent	Traditional urban costs				
	Cost of the city	Land rent	Price of an apartment per square metre	2010	Osservatorio del Mercato Immobiliare
	Social conflict	Malaise	Registered homicides per 1,000 inhabitants	2005	Sistema di Indicatori Territoriali ISTAT
	Unconventional urban benefits				
	Urban networks	Urban networks	Participations of local institutions to FP5 projects over total employees	Total 1998– 2002	CORDIS
	Urban functions	Urban functions	Share of firms in J and K industries over total active firms	2001	Censimento nazionale dell'Industria e dei Servizi (Industrial census) 2001
	Unconventional urban costs				
	Non-compact urban form	Sprawl	Percentage of non-urbanised soil	2006	CORINE Land Cover

Table 1 The variables, the indicators and their sources (continued)

Source: Authors' elaboration

For each variable presented in equation (8), we defined a suitable indicator, which also aims at getting as close as possible to the proxies employed in the previous, EU-wide empirical test. Table 1 provides a summary of the dataset built for the empirical analysis.

The variance in the data is relatively large, the use of a spatially homogeneous data base notwithstanding. Table 2 presents the main descriptive statistics for all the variables used in the empirical test of the model in equation (8).

 Table 2
 Descriptive statistics for the variables included in the empirical analysis

Variable	Obs	Mean	Std. Dev.	Min	Max
Population	103	570,465.40	613,195.20	88,694.23	4,194,060.00
Land rent	103	€1,602.75	€655.26	€622.23	€3,328.66
Malaise	103	0.16	0.09	0.04	0.59
Amenities	103	9,451.80	13,212.81	400.00	80,388.79
Sectoral diversity	103	0.85	0.02	0.80	0.88
Population density	103	254.08	337.61	39.30	2,630.68

Source: Authors' calculations

 Table 2
 Descriptive statistics for the variables included in the empirical analysis (continued)

Variable	Obs	Mean	Std. Dev.	Min	Max
High-level urban functions	103	20.52%	2.92%	14.66%	33.96%
Urban networks	103	0.09	0.12	0.00	0.66
Sprawl	103	0.04	0.05	0.01	0.31

Source: Authors' calculations

3.2 OLS estimates

This paper presents additional evidence on the validity of the model presented in equation (8), with the use of a novel dataset of Italian NUTS3 regions. In this section, we first provide classical OLS evidence of the model (Section 3.2). Section 3.3 shows some robustness checks of such estimates, with the use of Instrumental Variables estimates, in order to rule out endogeneity issues for the variables we believe would be most affected by such issue.

The results of the empirical exercise are presented in Table 3. For each column, 1 through 8, each individual regressor, starting from land rent, is added to the regressions.

Table 3Results of the OLS regressions

Dependent variable: equilibrium city population								
Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Constant	8.93***	8.16***	9.78***	11.50***	9.58***	12.78***	15.34***	12.06***
term	(1.40)	(1.58)	(1.62)	(1.63)	(1.32)	(1.14)	(1.64)	(1.95)
Land rent	0.55***	0.63***	0.05	0.26	-0.15	-0.38**	-0.39**	-0.45***
	(0.19)	(0.20)	(0.22)	(0.24)	(0.22)	(0.16)	(0.17)	(0.16)
Malaise	_	-0.10	0.01	-0.09	-0.17*	-0.21**	-0.25***	-0.26***
		(0.14)	(0.14)	(0.13)	(0.10)	(0.08)	(0.09)	(0.09)
Urban	_	-	0.33***	0.30***	0.32***	0.32***	0.31***	0.31***
amenities			(0.07)	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)
Urban	-	-	_	0.17***	0.07**	0.12***	0.15***	0.09**
diversity				(0.05)	(0.03)	(0.03)	(0.04)	(0.04)
Density	-	-	-	-	0.55***	0.49***	0.43***	0.73***
					(0.05)	(0.05)	(0.07)	(0.12)
City	-	-	-	-	_	0.13***	0.11***	0.10***
networks						(0.04)	(0.04)	(0.03)
Urban	-	_	-	_	-	-	1.01**	1.18***
functions							(0.44)	(0.46)
Sprawl	_	_	_	_	-	_	_	-0.32***
								(0.11)

Notes: *significant at the 90% level; **significant at the 95% level; ***significant at the 99% level; standard errors are shown in brackets.

Source: Authors' elaboration

Dependent variable: equilibrium city population								
Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
R ²	0.09	0.10	0.27	0.35	0.63	0.69	0.71	0.73
Joint F test	8.03***	4.76***	10.79***	12.29***	52.16***	49.29***	42.54***	43.89***
Robust standard errors	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of observations	103	103	103	103	103	103	103	103

 Table 3
 Results of the OLS regressions (continued)

Notes: *significant at the 90% level; **significant at the 95% level; ***significant at the 99% level; standard errors are shown in brackets.

Source: Authors' elaboration

The results show a high degree of robustness, and turn out to be statistically highly significant (with the expected sign for costs and benefits). Overall, the model explains a wide share of total variability ($R^2 = 0.73$ for the most general model, shown in column 8). All parameters maintain throughout the various specifications a high degree of parameter stability, which provides evidence of a relative absence of multicollinearity in the data. The only exception to parameter stability is the land rent parameter, a result which can easily be explained by the fact that when land rent is introduced on its own, it captures all advantages that the city provides; after netting out its relations with other benefit and cost variables, land rent captures urban costs and is the single highest cost for urban population.

Model 8 in Table 3 presents the whole specification of equation (8), since both conventional and unconventional variables, like city-networks and high level urban functions, are added to our estimates. The main conclusions on all relevant variables hold, providing a good fit of the model (73% of total variance explained), and leading to the following results:

- · agglomeration economies, generically measured by urban density, do matter
- traditional views on urban advantages, linked to diversity and amenities, increase the explicative power of the model (R²) from 10% to 35%
- most recent non-conventional views on urban growth, pointing out the relevance of new elements like the presence of economic and power functions and participation to an urban network, are corroborated: these elements allow cities to achieve equilibrium at higher sizes, thus allowing cities to sustain the implied higher urban costs
- land rent, after netting out its relations with other benefit and cost variables, is the single highest cost for urban population, reflected in the highest parameter estimate within the final model (8).

The empirical results also allow us to compare the city population predicted by the model with the actual population for each FUA, and therefore the identification of cities beyond (or below) their theoretically-determined equilibrium size. A comparison between the equilibrium city size predicted by the model and the real population level is presented in Figure 1.

Interestingly, Figure 1 shows that not necessarily smaller FUAs are expected to grow more in the future. On the contrary, the graphical analysis provides further verification of the fact that the model is based on the equilibrium between marginal location costs and marginal location benefits. In other words, large cities not necessarily should decrease in size: they can actually sustain such large dimensions, and even, at the margin, increase, provided they grant local dwellers accessibility to a wide array of urban benefits, or, by the same token, a reduced set of urban costs.⁶

Finally, Figure 2 presents the fit of the rank-size rule on the estimated data. On the x-axis, the log city rank calculated by the model is shown; on the y-axis, the log city population (once again predicted by the model) is instead presented. The functional form is best fitted with a quadratic interpolation,⁷ which provides evidence of the high non-linearities existing in the data used for the empirical analysis.







Figure 2 Rank-size rule for Italian FUAs (see online version for colours)

Source: Authors' calculations

3.3 Instrumental variables regressions

In order to further rule out possible issues of reverse causality (e.g., population increases determining a rise in urban costs and benefits), a solid approach to causality identification is adopted, with the use of instrumental variables (Table 4).

Among all independent variables, malaise (crime rates), the share of high level urban functions, and the intensity of urban networks are likely to be more heavily at risk of reverse causation. In fact:

- crime rates are often found to be higher in large urban agglomerations (Glaeser and Sacerdote, 1999)
- high-level urban functions are typical of large urban agglomerations (Clark, 1945)
- international networks are based on nodes mostly located in large metropolitan areas (Sassen, 2001).

In Table 4, model 8 is replicated (column 1) and the crime, urban functions, and urban networks variables are instrumented in sequence (columns 2–4).

The instruments are chosen as to be correlated with the potentially endogenous regressor, but not with city population. Crime levels are instrumented with time-lagged social capital indicators, and lagged per capita GDP levels as a measure of urban wealth. Recent studies demonstrate in fact, both theoretically as well as empirically, why higher levels of social capital should be correlated with lower crime levels (Akçomak and ter Weel, 2012). Social capital acts as a non-legal constraint to deviations, via informal sanctioning, altering the incentives of agents – costs and benefits – to actually commit offenses (Becker, 1968); moreover, richer and older societies are expected to have lower

crime rates. Social capital characteristics are summarised by social capital infrastructure (social, education, and culture infrastructure as calculated by the Italian *Tagliacarne* Institute).

Urban functions and city networks are stronger in culturally-advanced and rich societies that do not necessary reside in large cities. High-level urban functions and city networks are therefore instrumented with social capital infrastructure and a (time lagged) measure of crime rates. A further instrument is the number (for each FUA) of universities in the top 500 2003 Shanghai ranking;⁸ while several such high-level educational institutions are located outside the largest metropolitan areas, they certainly correlate with the share of workers employed in high-level professions. Finally, city networks are also instrumented with the (time-lagged) level of trade openness, which is expected to correlate with the intercity networks of scientific cooperation, but not with the absolute (physical) size of the city.⁹

Dependent variable: equilibrium city population								
Model	(1)	(2)	(3)	(4)				
Constant term	12.06***	11.45***	15.63***	12.06***				
	(1.95)	(2.12)	(3.04)	(1.95)				
Land rent	-0.45***	-0.33*	-0.48***	-0.45***				
	(0.16)	(0.20)	(0.16)	(0.16)				
Malaise	-0.26***	-0.42**	-0.32***	-0.26***				
	(0.09)	(0.20)	(0.10)	(0.09)				
Urban	0.31***	0.29***	0.29***	0.31***				
amenities	(0.05)	(0.05)	(0.05)	(0.05)				
Urban	0.09**	0.10**	0.12**	0.09**				
diversity	(0.04)	(0.04)	(0.05)	(0.04)				
Density	0.73***	0.74***	0.68***	0.73***				
	(0.12)	(0.13)	(0.14)	(0.12)				
City networks	0.10***	0.10***	0.07*	0.10***				
	(0.03)	(0.03)	(0.04)	(0.03)				
Urban	1.18***	1.34***	2.90**	1.18***				
functions	(0.46)	(0.47)	(1.14)	(0.46)				
Sprawl	-0.32***	-0.33***	-0.38***	-0.32***				
	(0.11)	(0.12)	(0.13)	(0.11)				
Variable instrumented	_	Malaise	Urban functions	City networks				

 Table 4
 Results of the instrumental variables regressions

Notes: *Significant at the 90% level; **significant at the 95% level; ***significant at the 99% level; Standard errors are shown in brackets.

Model	(1)	(2)	(3)	(4)
Instruments used	_	Social capital infrastructure (social, education, and culture infrastructure); time lagged crime rates	Social capital infrastructure, time lagged crime rate, number of universities in the 2003 Shanghai ranking	Trade openness, social capital infrastructure, time lagged crime rate, number of universities in the Top 500 2003 Shanghai ranking
Partial R ² of excluded instruments	_	0.19	0.16	0.14
F test for excluded instruments	-	5.31***	2.38**	1.80*
Anderson canon. corr. likelihood ratio stat.	-	21.62***	17.87**	15.77**
Anderson- Rubin χ^2 test of joint significance of endogenous regressors	_	7.99*	23.95***	22.17***
R ²	0.73	0.72	0.68	0.73
Joint F test	7.61***	43.89***	26.82***	43.89***
Number of observations	103	103	103	103

 Table 4
 Results of the instrumental variables regressions (continued)

Notes: *Significant at the 90% level; **significant at the 95% level; ***significant at the 99% level; Standard errors are shown in brackets.

The results of the instrumental variables regressions confirm the main message from the previous section. The theoretical model holds even after controlling for reverse causation. As Table 4 shows, malaise (the crime variable), urban functions, and urban networks remain significant and with the right signs once instrumented.

As for the instruments' validity, we run the usual battery of tests, whose results are shown in the bottom part of the table. First, the partial R^2 of the instruments is shown. This represents the percentage linear variance due to the selected instruments in the first-stage auxiliary regression. In both cases, a remarkable 14/19% of total variance is explained with our instruments.

In the second block, the Cragg-Donald statistics (Cragg and Donald, 1993) is shown. This statistic tests the null hypothesis of under-identification of the matrix of the regressors/instruments Q_{XZ} . It is distributed as a χ^2 (indeed, it can be reconduced to a Wald test; see for instance Baum et al., 2003). Since in all three cases the null hypothesis is rejected, we can safely infer that the instrumental variables regressions are not under-identified.

The third block in the bottom part of Table 4 shows the Anderson LR statistic, which represents an instrumental variables relevance test. Once again the test demonstrates (at the 99% for malaise, and at the 95% confidence level for functions and networks) the validity of these instruments as tools to exclude reverse causality in these relations.

Finally, the last block of tests in the bottom part of Table 4 shows C statistics for both instrumental variables regressions. This statistic demonstrates that the instruments are also exogenous with respect to the variable being instrumented, which is a crucial condition for the instrument validity.

The use of a time-lagged value for the dependent variable, and instrumental variable techniques, safely allow to conclude that reverse causation does not represent a major issue in our results. Although in the long run exogenous shocks to any urban size determinant may indeed cause circular causation (i.e., a shock in crime rates may reduce equilibrium city population, thereby reducing crime rates), the results in this paper suggest that the theoretical micro-foundations of the model of equilibrium city size hold to the empirical test.

4 Conclusions and policy implications

This paper entered the optimal city size debate by taking on an intermediate position between the advocates of the idiosyncratic approach – each city obeys its own rules, and, therefore, cities cannot be compared – and the pure neoclassical extension of the urban land rent model, at the limit yielding a world with cities with an identical size. While the former approach allows the identification of variables capable of explaining the different city size that can be observed in the real world, it fails in allowing city comparability. The structural elegance of the latter approach, instead, drives to the unrealistic conclusion that, with centrifugal forces that decrease over time (i.e., lower transport costs), cities will eventually converge to a set of identically-sized urban areas.

In order to formulate an alternative explanation, a model of equilibrium city size has been derived, with agents assumed to be free to move across space and set their marginal location costs and benefits equal. The reduced form model has next been tested on sample of 103 Italian FUAs.

The econometric evidence suggests that indeed modern theoretical paradigms, like the city network paradigm, explain most of current city size disparities. While rent, net of the urban benefits it reflects, still represents the single highest cost associated to urban size, cities benefit not only from attracting highly educated professionals and hosting rich and diversified markets, but also from pure amenities and appropriate, compact city form, which are found to be associated with a larger urban size.

Besides, results clearly and consistently show that being connected to an urban network – in this case, a cooperation network in the scientific field – allows cities to achieve a larger equilibrium size. The same role is played by presence of high order functions, even if the empirical validation is less robust.

The use of solid instrumentation techniques allows also to infer that reverse causation and endogeneity do not represent a major problem in our results.

Empirically, this paper is based on a cross section of observations. The model here derived may thus be extended to a dynamic setting, whereby the equilibrium *growth* of cities is explained. Further work on this topic may therefore extend in two directions.

Firstly, the theoretical model may be extended to a dynamic setting, thereby allowing the analysis of the conditions for cities to reach their long-run equilibria. Secondly, the empirical work may benefit from extending the analysis (so far limited by data constraints) to a panel dataset, thus delving into the empirical conditions for reaching city-specific equilibrium size. Besides, additional empirical work may also evaluate the speed of convergence of each city towards its equilibrium, and how this convergence process is influenced by city-specific characteristics, or policy levers.

Our empirical results have in fact allowed the identification of city-specific variability: some cities show an actual population which is slightly different from the equilibrium city size predicted by the model. Differences can be explained by good or bad governance, thereby suggesting future strategies for more efficient urban planning and the construction of sound economic and social 'visions'.

On the one hand, in fact, cities displaying a predicted population level lower than the actual one are expected to (at least potentially) grow to fill this gap. On the other hand, cities with an actual population higher than the population predicted by the model may have been able to reach this level with a prior higher quality of governance.

An alternative explanation for the latter case may be related instead to the existence of different paradigms for the equilibrium city size in different urban systems. While urban systems in the EU or individual EU countries are characterised by a relatively limited internal variability, stylised facts suggest that cities in Eastern or Mediterranean countries, or cities in large countries with respect to cities in small ones, may be obeying partially different laws.

The inclusion of such non-standard (viz. qualitative and governance) elements both in theoretical models as well as in empirical analyses represents an interesting challenge, and even more so, given the positive worldwide momentum of urbanisation trends and the dire logistic and management problems that growing cities will increasingly pose.

References

- Akçomak, I.S. and ter Weel, B. (2012) 'The impact of social capital on crime: evidence from the Netherlands', *Regional Science and Urban Economics*, Vol. 42, Nos. 1–2, pp.323–340.
- Alonso, W. (1960) 'A theory of the urban land market', Papers and Proceedings of the Regional Science Association, Vol. 6, No. 1, pp.149–157.
- Alonso, W. (1971) 'The economics of urban size', Papers and Proceedings of the Regional Science Association, Vol. 26, No. 1, pp.67–83.
- Baum, C.F., Schaffer, M.E. and Stillman, S. (2003) *Enhanced Routines for Instrumental Variables/GMM Estimation and Testing*, Boston College Economics Working Paper No. 667.
- Becker, G. (1968) 'Crime and punishment: an economic approach', *Journal of Political Economy*, Vol. 76, No. 2, pp.169–217.
- Beckmann, M. and McPherson, J.C. (1970) 'City size distribution in a central place hierarchy: an alternative approach', *Journal of Regional Science*, Vol. 10, No. 1, pp.25–33.
- Camagni, R. (2011) Principi di economia urbana e territoriale, Carocci, Rome, IT.
- Camagni, R., Capello, R. and Caragliu, A. (2013) 'One or infinite optimal city sizes? In search of an equilibrium size for cities', *The Annals of Regional Science*, forthcoming, DOI 10.1007/s00168-012-0548-7.
- Catin, M. (1991) 'Économies d'agglomération et gains de productivité', *Revue d'Économie Régionale et Urbaine*, Vol. 5, No. 1, pp.565–598.
- Cheshire, P. and Magrini, S. (2006) 'Population growth in European cities: weather matters but only nationally', *Regional Studies*, Vol. 40, No. 1, pp.23–37.

- Clark, C. (1945) 'The economic functions of a city in relation to its size', *Econometrica*, Vol. 13, No. 2, pp.97–113.
- Cragg, J.G. and Donald, S.G. (1993) 'Testing identifiability and specification in instrumental variables models', *Economic Theory*, Vol. 9, No. 2, pp.222–240.
- Fujita, M. and Thisse, J-F. (1996) 'Economics of agglomeration', Journal of Japanese and International Economies, Vol. 10, No. 4, pp.339–378.
- Fujita, M., Krugman, P. and Mori, T. (1999) 'On the evolution of hierarchical urban systems', *European Economic Review*, Vol. 43, No. 2, pp.209–251.
- Fujita, M. (1989) Urban Economic Theory: Land Use and City Size, Cambridge University Press, Cambridge, UK.
- Gabaix, X. (1999) 'Zipf's law for cities: an explanation', *Quarterly Journal of Economics*, Vol. 114, No. 3, pp.739–767.
- Glaeser, E.L. and Kahn, M.E. (2004) 'Sprawl and urban growth', in Henderson, J.V. and Thisse, J-F. (Eds.): *Handbook of Regional and Urban Economics*, Vol. 4: Cities and Geography, pp.2481–2527, Elsevier, Amsterdam, NL.
- Glaeser, E.L. (2008) Cities, Agglomeration and Spatial Equilibrium, Oxford University Press, Oxford, UK.
- Glaeser, E.L. and Mare, D.C. (2001) 'Cities and skills', Journal of Labor Economics, Vol. 19, No. 2, pp.316–342.
- Glaeser, E.L. and Resseger, M. (2009) 'The complementarity between cities and skills', *Journal of Regional Science*, Vol. 50, No. 1, pp.221–244.
- Glaeser, E.L. and Sacerdote, B. (1999) 'Why is there more crime in cities?', *Journal of Political Economy*, Vol. 107, No. 6, pp.S225–S258.
- Henderson, J.V. (1974) 'The sizes and types of cities', *The American Economic Review*, Vol. 64, No. 4, pp.640–656.
- Henderson, J.V. (1985) Economic Theory and the Cities, Academic Press, Orlando, FL.
- Henderson, J.V. (1995) 'Comment on 'urban concentration: the role of increasing returns and transport costs' by Krugman', *Proceedings of the World Bank Annual Conference on Development Economics*, 1994, pp.270–274.
- Henderson, J.V. (1996) 'Ways to think about urban concentration: neoclassical urban systems vs. the new economic geography', *International Regional Science Review*, Vol. 19, Nos. 1–2, pp.31–36.
- Krugman, P. (1991) Geography and Trade, MIT Press, Cambridge, MA.
- Krugman, P. (1995) 'Urban concentration: the role of increasing returns and transport costs', Proceedings of the World Bank Annual Conference on Development Economics, 1994, pp.241–263.
- Muth, R. (1969) Cities and Housing, The University of Chicago Press, Chicago, IL.
- Rappaport, J. (2004) 'Why are population flows so persistent', *Journal of Urban Economics*, Vol. 56, No. 3, pp.554–580.
- Reggiani, A. and Nijkamp, P. (2012) *Did Zipf Anticipate Socio-economic Spatial Networks?*, Working Paper DSE n. 816, Università di Bologna.
- Richardson, H. (1972) 'Optimality in city size, systems of cities and urban policy: a sceptic's view', Urban Studies, Vol. 9, No. 1, pp.29–47.
- Richardson, H. (1978) Regional and Urban Economics, Penguin Books, Harmondsworth, UK.
- Roback, J. (1982) 'Wages, rents, and the quality of life', *The Journal of Political Economy*, Vol. 90, No. 6, pp.1257–1278.
- Rosen, S. (1979) 'Wage-based indices of urban quality of life', in Mieszkowski, P. and Straszheim, M. (Eds.): *Current Issues in Urban Economics*, pp.74–104, Johns Hopkins University Press, Baltimore, MA.

- Royuela, V. and Suriñach, J. (2005) 'Constituents of quality of life and urban size', Social Indicators Research, Vol. 74, No. 3, pp.549–572.
- Sassen, S. (2001) The Global City : New York, London, Tokyo, Princeton University Press, Princeton, NJ.
- Simon, H. (1955) 'On a class of skew distribution functions', *Biometrica*, Vol. 42, Nos. 3–4, pp.425–440.
- Uzawa, H. (1962) 'Production functions with constant elasticities of substitution', *Review of Economic Studies*, Vol. 29, No. 4, pp.291–299.
- Valvanis, S. (1955) 'Lösch on location', American Economic Review, Vol. 44, No. 4, pp.637-644.
- Zipf, G.K. (1949) Human Behavior and the Principle of Least Effort, Addison-Wesley, Boston, MA.

Notes

- 1 On the other hand, showing urban heterogeneity as a consequence of a casual identity/overlapping of urban production and cost functions inside a given interval, in presence of increasing returns [as in Fujita, (1989), p.163], or as a consequence of the choice of different wage rates by different developers/managers of new towns, instead of the normal national wage rate [as in Fujita, (1989), p.166], means relying on strange, casual cases rather than on theoretically relevant conditions. Introducing a consistent mix of urban externalities, quality public goods and rankings of services and functions into a modelling paradigm-based substantially on the accessibility principle represents still a widely open theoretical challenge. Quite evidently, building on an accessibility principle, it looks difficult to take in full consideration elements linked to an agglomeration principle while maintaining the elegance of the traditional models [Camagni, (2011), ch. 2].
- 2 The complex model addressed to tackle this limit proposed by Fujita, Krugman and Mori (1999) still looks far from providing a convincing solution.
- 3 The controversial issue concerning the capability of a pure market mechanism to lead to an optimal allocation of population in cities evident in the stylised theoretical model of urban land use but not confirmed by more complex formulations with externalities and public goods [Henderson, (1985), p.262; Fujita, (1989), p.284] is not crucial for the model presented here.
- 4 See Camagni et al. (2013) for analytical details of the model.
- 5 A similar result in a different theoretical context was achieved by Fujita (1989, p.151) treating 'the open city model with absentee landowners' inside the land use equilibrium theory. Assuming two cities with two communities maximising their utilities, with similar productivity curves but, in one case a superior level of amenities, the model proves that the city with amenities, being more attractive, reaches a higher equilibrium size (which is also optimal). This result shows that a convergence between our approach and the one of the new urban economics.
- 6 In fact, cities at the right-hand side of the graph in Figure 2, i.e., FUAs which are expected to grow, present on average a larger size, whilst at the same time being affected by lower levels of crime, a lower level of land rent, and a more compact urban form. A full list of statistics is available upon request from the authors.
- 7 The best fit of the data plotted in Figure 2 is provided by the following functional form:

 $\ln(pop) = \alpha + \beta_0 \ln(rank) + \beta_1 \ln(rank)^2$, with $\hat{\alpha} = 14.77$, $\hat{\beta}_0 = -0.26$ and $\beta_1 = -0.06$.

The level of fit (R^2) is equal to 0.91.

- 8 Raw data are available at http://www.arwu.org/ARWU2003.jsp.
- 9 Many examples of small and medium-size cities, but well connected in worldwide networks, include Geneva and Luxembourg.