Article:

Arbabi, H. orcid.org/0000-0001-8518-9022, Mayfield, M. and Dabinett, G. (2019) Urban performance at different boundaries in England and Wales through the settlement scaling theory. Regional Studies, 53 (6). pp. 887-899. ISSN 0034-3404

https://doi.org/10.1080/00343404.2018.1490501

Urban Performance at Different Boundaries in England and Wales through the Settlement Scaling Theory

Hadi Arbabi[a]*, Martin Mayfield[a], Gordon Dabinett[b]

[a]Department of Civil & Structural Engineering, University of Sheffield, Sheffield, UK;

[b]Department of Urban Studies and Planning, University of Sheffield, Sheffield, UK;

*Corresponding Author:

harbabi1@sheffield.ac.uk

Department of Civil & Structural Engineering

Sir Frederick Mappin Building

Sheffield | S1 3JD, UK

-

martin.mayfield@sheffield.ac.uk

Department of Civil & Structural Engineering

Sir Frederick Mappin Building

Sheffield | S1 3JD, UK

-

g.e.dabinett@sheffield.ac.uk

Department of Urban Studies and Planning

Western Bank

Sheffield | S10 2TN

The relationship between transport-led agglomeration and economic performance is evaluated in an English and Welsh context. We examine the effects of scale, i.e. inter-city versus intra-city mobility infrastructure, on urban size-cost performance. An additional contribution of this paper lies in its use of power-law scaling models of urban systems, enabling an assessment of optimality in the trade-off between economic output and mobility costs accounting for ease of access within cities coupled with their built-density. Findings suggest economic underperformance coincides with inadequate mobility at both inter-city and intra-city scales while over-performance is accompanied by over-grown urbanised area and escalating mobility costs.

Keywords: agglomeration; urban scaling; scaling laws; urban performance; infrastructure planning

Subject classification codes: O18, R00, R41

Introduction

Cities and urban cores are the global nuclei of innovation, wealth generation, resource consumption, and energy dissipation. Against a backdrop of expanding urbanised areas and increasing urban populations (Seto, Güneralp, & Hutyra, 2012) with limited resources available to sustain them, the need to design and maintain urban fabric and infrastructures in a manner that enables cities of higher productivity for minimal dissipated resources is evident. In practice, however, the ability to clearly identify, design, and implement infrastructural measures, e.g. transport and mobility, that in fact improve economic performance, for a given definition of performance objective, is dependent on the availability of appropriate models and understanding of the system at appropriate scales of intervention. This seemingly simple dependency could often suffer from the incongruities that may arise between the policy intentions, how they are interpreted into planning interventions, and the adequacy of the theoretical frameworks that are used to assess, inform, and shape them.

These policy needs and the assortment of models and frameworks, abstract or otherwise, used addressing them can be conceptualised along a disconnected spectrum of spatial scales. At the macro-scale end, broad national strategies could be thought of as shaped and informed by economic targets and demand pressure on the existing transport infrastructure. Such strictly rational formulation, however, would admittedly be at the mercy of political agendas and priorities. At the other inter-city micro-scale end, the existing models and insights available are made up of the different variations on land-use and transport interaction (LUTI) models. These require a relatively long list of input parameters for calibration and operation not many of which are routinely and homogeneously collected or easily available (Echenique, Grinevich, Hargreaves, & Zachariadis, 2013). Besides the inherent inability of these models to single-handedly identify strategic infrastructural needs, their deployment at larger scales to inform meso-scale policies is hampered by the aforementioned data intensity and difficulties.

Infrastructural efforts at a meso-scale are thus driven by spatial agglomeration arguments whereby higher regional economic productivities are stimulated through the implementation of inter-city mobility provisions in the image of those in Randstad, the Netherlands, and Germany's Rhine-Ruhr (Burger, Meijers, Hoogerbrugge, & Tresserra, 2015). The theoretical economic models currently used justifying these inter-city mobility-enhancing strategies, however, remain in most parts inherently place non-specific with singular and arbitrary choices of spatial scales (Martin, Pike, Tyler, & Gardiner, 2015). While these frame the agglomeration process as a balance between increases in productivity and accumulating ‘congestion costs’, these size-cost balances remain abstract in such models (Abel, Dey, & Gabe, 2012).

This paper then aims to offer new insights on the effects of spatial scales on city performance balance and the extent to which this is influenced by the provision of transport infrastructure at different scales, i.e. intra-city versus inter-city mobility, using the urban system in England and Wales (EW) as an example. In order to do this, we adopt a scaling formulation of cities based on their population size with an explicit formulation of the balance between economic output, ease of mobility, and its associated costs incurred with reference to the actual physical extent of cities and the larger urban systems to which they belong. To the best of the authors’ knowledge, this is among the first instances of application of such models in exploring explicit considerations of city size-cost balance across spatial scales. Additionally, given that agglomeration frameworks and much of their evidence is based on Asian and North American urban systems, England and Wales offer a particularly unique opportunity for the examination of these spatial effects within an urban system that is currently experiencing unique economic challenges (McCann, 2016, Chapter 3) with an ever-widening divide that exists between the productivity and economic output of the south of England and the other regions in EW (Rowthorn, 2010).

Shaped and framed by the wider policy efforts stemming from the decentralisation and devolution of certain powers to local entities in the form of combined authorities or city regions (Gardiner, Martin, Sunley, & Tyler, 2013), the infrastructure policy debate in EW has been dominated by the attempts to address this performance gap. For national transport infrastructure, these attempts have generally been envisaged as creating and enabling these city regions to act as single economic units by providing inter-city transport infrastructures that reduce journey times encouraging agglomeration economies (National Infrastructure Commission, 2016). The process of devising these transport links, inspired by the inter-city rail connectivity in the Randstad and Rhine-Ruhr, has only been argued at a singular spatial scale (Transport for the North, 2016). Consequently, although these policies have been studied in terms of their implications for infrastructure governance and funding (O’Brien & Pike, 2015), an explicit exploration of the scale effects on the size-cost balances and performance in EW has largely been absent from both the policy papers and the larger academic debate.

Our results signal at a systemic lack of adequate mobility and accessibility for a large portion of city units considered at various spatial scales implying that lack of adequate mobility provisions is at the heart of a less-than-expected economic performance. While such effects are more easily noticeable at larger inter-city scales, the problem is reoccurring at smaller scales and intra-city boundaries. This suggests that although intra-city transport-led agglomeration strategies are fitting, when implemented alone they would only mask mobility shortcomings at smaller scales without addressing underlying causes of under-performance. As such, transport infrastructure planning cannot simply be led by agglomeration theory principles being applied at a single spatial scale and more concurrent consideration of urban scales is needed.

The rest of the article is structured as follows. The next section provides a background to the scaling understanding of cities and introduces the specific model adopted in this study. In Section 3, we first provide a brief account of the methods and data used before establishing a broad empirical agreement with the theoretical scaling model and examining the average city mobility across spatial scales followed by their size-cost balance as compared with their idealised theoretical counterparts at these scales. In light of these comparisons, Section 4 then discusses the implications of inadequate mobility provision and the potential of transport-led agglomeration at different spatial scales. Lastly, an extended account of boundary definitions and some additional results and larger figures are offered in the accompanying appendix online.

Urban Scaling and Settlement Scaling Theory

In the past decade, with growing abilities to collect, share, and analyse larger bodies of data pertaining to urban settlements, an understanding of cities and their properties as population scaling functions, formulated in the vein of similar allometric relations underlying the growth and size of organisms, has gained more traction both analytically and empirically (Bettencourt, Lobo, Helbing, Kühnert, & West, 2007). More importantly, this allometric line of thinking has already made an impression on the planning and economics literature. While part of this influence has been implicit in the form of concurrent observations of city rank-size distributions (P. Cheshire, 1999), others like Glaeser have been more explicit in the use of and reference to this field of literature, its theoretical frameworks, and models it provides, in their own works (Glaeser, Ponzetto, & Zou, 2016). What is important from a policy and planning perspective is the ability of these frameworks to offer not only more tangible articulations of economic/energetic size-cost balances but also formulations that can be implemented and explored across a spectrum of scales without the obstacles faced by traditional LUTI models.

Bettencourt and West (2010) have put forward a notion of ‘universal features’ codifying the regularities present in various urban characteristics as a function of city population. They formalise these correlations of urban properties with city size as

$$F(N) = F_0 N^{\beta} \tag{1}$$

where $F(N)$ denotes the average-aggregate urban characteristic of choice for population size N, e.g. gross value-added (GVA), urbanised land area, employment, etc., F_0, the baseline prevalence of F, and finally, β, the exponent determining the nature of scaling relation.

To explain these observations, a variety of urban models has been developed that yield such scaling behaviours for aggregated average response of urban attributes. These include models rooted in probabilistic conceptualisations of activities taking place in cities and the portion of population contributing towards them (Gomez-Lievano, Patterson-Lomba, & Hausmann, 2016) and those that are based on network realisations of the interactions between inhabitants and/or the geographical embedding of such networks within cities (Yakubo, Saijo, & Korošak, 2014).

Bettencourt (2013) introduces a simplified framework to derive such scaling behaviours which explicitly formulates a balance between the economic output generated in a city, Y, and the energy dissipated within the city, W, through the mobility processes that enable the population to generate this output. The simplest model, i.e. the social reactor model, from this Settlement Scaling theory starts from four simple underlying assumptions:

(1) the aggregate socio-economic outputs are commensurate with the sum total of the number of human interactions locally,

(2) the population is mixing uniformly so that each individual has access and the minimum resources to explore the city (Jones, 2016),

(3) the infrastructure embedded in the city is a hierarchical network that grows incrementally and gradually keeping individuals connected with one another (Samaniego & Moses, 2008), and that

(4) the average baseline human production is not a function of population and remains constant across cities in an urban network (Szüle, Kondor, Dobos, Csabai, & Vattay, 2014).

Before going any further, it should be noted that these four assumptions, and in particular the second one, are to describe and model a highly idealised city. We will provide a discussion of the effects of their violation on the rest of the model behaviour at the end of this section. Nevertheless, based on the first two assumptions, the model formalises an upper bound for average output for a city of population N as

$$Y = \bar{g} a_0 l \frac{N^2}{A_n} \tag{2}$$

where $\bar{g}$ is the average strength of interactions between individuals, $a_0 l$ denotes the average area of coverage of individuals through which they experience the city, and N^2, the total possible number of interactions over the entirety of the city's urbanised area, A_n. The product $\bar{g} a_0 l$, referred to as G hereafter, incorporates the extent of the mobility of individuals and their output. We note that it corresponds to the baseline human production mentioned in the fourth assumption and is presumed to be independent from population size, N. Although objections could be raised regarding a uniformly distributed series of individual interactions over a characteristic path validly portraying the mechanisms controlling encounters in and across all economic sectors and activities, the formulation is to be primarily that of a smoothed-out average of all behaviours rather than individual specific.

Assumptions 2 and 3 similarly enable a formalisation of A_n in terms of population. Describing the geometry of the city and its inhabitants average exploration path through it in terms of fractal dimensions, Bettencourt (2013) derives the scaling of the urbanised area as

$$A_n = A_{n0} N^{1-\delta} \quad (3)$$

With $\delta = \frac{H}{D(D+H)}$ where H and D are fractal dimensions characterising the individual mobility paths and city geometry, respectively, while A_{n0} embodies the baseline of urbanised area and because of the second assumption would itself be a combined function of G and a notional average price of mobility per unit distance. It is worth mentioning here that the real life city geometries would imply a constraint of $2 \leq D \leq 3$ for city dimension. This is to say that cities exist somewhere between a flat surface in space and the full volume of this surface extruded outwards. The individuals path, H, can more intuitively be thought of as the ease of mobility when travelling through the city. The full mixing assumption of Bettencourt's model would mean $H \approx 1$ in terms of the geometry of inhabitants access effectively assuming that even an individual unfamiliar with the city fabric, e.g. its neighbourhoods and amenities, can afford to reach any place in the city. This encompasses both the availability of mobility modes connecting the city and their affordability to an average inhabitant.

The total mobility costs, W, over the urbanised extent can also be estimated by framing it as the energy dissipated through the hierarchy of the infrastructure modelled after parallel-connected resistors with electrical current through resistors replaced with flow of people on roads etc. By substituting equation 3 back in equation 2, the scaling of economic output, its mobility costs, and urbanised area can then be summarised as follows in a form similar to that of equation 1[1]

$$\begin{cases} Y(N) = Y_0 N^{1+\delta} \\ W(N) = W_0 N^{1+\delta} \\ A_n(N) = A_{n0} N^{1-\delta} \end{cases} \quad (4)$$

where Y and W clearly scale super-linearly with population size through the exponent

$\beta = 1 + \delta$ while urbanised area grows sub-linearly with $\beta = 1 - \delta$.

The social reactor model subsequently introduces the subtraction $Y - W$, as an indicator of a size-cost balance of cities. This objective function maximises as a function of G at an optimal value, G^*,. It is evident that the city is only viable when the balance between total output and mobility costs is positive, $G > G_{min} = 0$, reaches a maximum size-cost balance at $G = G^*$, and becomes unstable again for $G_{max} > G > G^*$ as the mobility costs start overwhelming beneficial gains. Also, note that for an idealised urban system where all cities strictly follow the scaling set out in equation 4, substituting the scaling of output and urbanised area back in equation 2 would provide an empirical estimate for the optimal base-line production as

$$G^* \equiv \bar{g} a_0 l = A_{n0} \cdot Y_0 \tag{5}$$

where A_{n0} and Y_0 are the baseline prevalence of urbanised area and economic output corresponding to the scaling of these attributes for idealised city in a distribution.

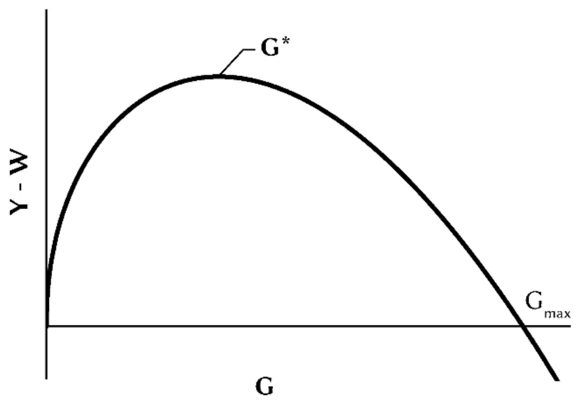

Figure 1. Schematic plot of economic output less the mobility costs.

The particular relevance of this model here, however, is the comparison opportunity it provides for gauging the state of this size-cost balance in cities and potential interventions that are implied to aid sub-optimal performances. For cities where $G < G^*$, their true potential socio-economic capacity is not reached. Looking back at the components of G, i.e. $\bar{g}a_0 l$, this is to say that the inhabitants are not fully exploring the city and hence cannot participate in as many interactions as the city has the potential to offer. As such, the interventions that would help increase the average individual areal coverage experiencing the city, e.g. better transport and accessibility within the city, would help the balance between the output and the energy used in the process of enabling people to generate it. For those with $G > G^*$, however, the socio-economic success of the city is overshadowed by the escalating costs of the transport and mobility requirements embodied by the term W. Consequently, to tilt the balance back and lower mobility costs measures such as compaction of cities built environment, its areal densification, and higher efficiencies of the transport modes could be implemented.

Finally, a few passages earlier, it was acknowledged that the assumption of a city with uniformly mixing population with full and homogenous access to the city is a particularly strong one. For a simple idealised city, this assumption invariably results in presuming the city two-dimensional ($D = 2$, i.e. mobility processes required to bring inhabitants together take place on the ground over city's flat surface rather than through its height and hence volume) and that individuals on average have full linear access to city ($H = 1$). Carrying the assumption forward results in a δ value of $\frac{1}{6}$ and subsequent idealised expectations of the power-law exponent $\beta = \frac{7}{6}, \frac{7}{6}, \frac{5}{6}$ for the scaling of output, mobility costs, and the urbanised area, respectively, in equation 4. Where cities become less fully accessible, H would take values less than unity implying that inhabitants' experience of the city becomes restricted to disparate patches, e.g. neighbouring but virtually disconnected communities, within the city rather than linearly accessible extents. This would in turn shrink the value of δ and hence result in a scaling of Y and

A_n closer to linear in equation 4 without outright violating any other aspect or framing of the overall model and or the size-cost functions.

Scaling and Size-Cost Balance

As discussed, this scaling formulation of cities can be used to offer categorical meso-scale comparisons of cities and urban regions both against an idealised realisation of cities, i.e. $D = 2$ and $H = 1$, and also against any specific performance balance as observed in one particular city, say, London. In this section, we start by demonstrating the extent of the agreement between the underlying assumptions and resulting predictions from Bettencourt's model by estimating the scaling exponents for the GVA and urbanised land area for a number of different city boundary delineations. We then estimate values of G for each city following

$$G = \frac{Y \cdot A_n}{N^2} \tag{6}$$

which is a rearrangement of equation 2. A comparison of these estimates for cities within each boundary definition with the optimal G^* calculated for their idealised fully accessible counterparts is then presented with an examination of the infrastructural needs of cities at different spatial scales.

Boundary Scenarios

We conduct our analysis for 11 different realisations of city boundaries seven of which are based on aggregating smaller cells with a minimum population density cut-off criterion. These are constructed based on the City Clustering Algorithm (CCA) described in (Rozenfeld, Rybski, Gabaix, & Makse, 2011). The algorithm continuously merges neighbouring cells, here on a square grid, based on the population density within cells until the neighbouring cells fall below a set density threshold. We have used the

GEOSTAT Grid (Office for National Statistics, 2016c) which contains population counts from the 2011 census aggregated on a square grid of $1 \times 1\ km^2$. A number of cut-off density values is then used within the range 100-3500 $\frac{N}{km^2}$ with the 1400 $\frac{N}{km^2}$ cut-off specifically adopted from Arcaute et al. (2015) identified as threshold for well-defined cities providing a close match with major cities' urbanised extent.

For the administrative units, the Local Administrative Units Level 1 (LAU1) (Office for National Statistics, 2016a) and the European Nomenclature of Territorial Units for Statistics Level 3 (NUTS3) (Office for National Statistics, 2016b) are used. The LAU1 and NUTS3 units, however, become wholly inappropriate when considering London as they break down connected parts of the city to maintain statistical uniformity. As such for these boundaries, we have merged the smaller constituting units in accordance with the extent of the Greater London Authority. Finally, to include boundaries theoretically consistent with the mixing population assumption in Bettencourt's idealised model which implies functional economies and labour markets (Bettencourt, 2013), we use the Travel-to-Work Area boundaries (TTWA) based on the commuting data from the 2011 census (Coombes & Office for National Statistics, 2015).

Population, urbanised area, and economic output estimation

Urban population for city units in each boundary definition has been calculated by summing the population count from the GEOSTAT grid cells intersecting the city units of different boundaries. This means directly aggregating population counts from the grid in the case of the density-based boundaries and summing area weighted counts for the other four boundaries based on the proportion of the area of the cells intersected by the city unit boundaries under a uniform cell density assumption. Similarly, the urbanised area for each boundary has been estimated by summing the area of the polygon segments from the contiguous built-up area layer intersecting boundary definitions.

The estimation of the economic output at different boundaries involves a few additional steps. We use the OECD’s GIS-based method using area-weighted proportionalities to arrive at the GVA estimates (OECD, 2012, pp. 45–48). Firstly the layer containing the GVA estimates, published by the ONS at NUTS3 levels for the year 2011 (Office for National Statistics, 2014), is intersected with the GEOSTAT population grid and each cell is assigned a portion of the GVA value according to

$$Y_{cell} = \sum \frac{Y_{NUTS3} \times \frac{N_{cell} \times A_i}{A_{cell}}}{N_{NUTS3}} \tag{7}$$

where the Y_{cell} is the total GVA assigned to a cell in the population grid, N_{cell} and A_{cell} the total population and area of the cell, Y_{NUTS3} and N_{NUTS3} the GVA and population of the intersecting NUTS3 areas respectively, and A_i the area of the portion of the cell intersected by the corresponding NUTS3 unit. Subsequently, a similar procedure is performed in reverse to aggregate back the GVA values from the population grid to the desired city boundary definitions discussed previously[2].

Urban performance in England and Wales

In obtaining the baseline prevalence and exponent of the scaling relations for urbanised area and GVA with population in each boundary definition, we use OLS estimators on the linearised log-transformation of Equation 1. The larger numbers of the excessively small units, especially in UA and C100 boundaries due to the smaller density cut-offs and small isolated built-up areas, however, would skew the tail of the power-law and hence result in inappropriate linear fits. As such, arbitrary minimum population limits, e.g. 500000 used by OECD (2012) and 50000 used in Arcaute et al. (2015), are often applied to distinguish urban and metro areas from those that are rural.

Instead of using a single arbitrary limit across boundaries, we apply the statistical method described by Clauset et al. (2009), which estimates a lower bound in the distribution of empirical data above which a power-law distribution, is followed to delineate 'urban' from 'rural' from a purely statistical perspective. This results in limiting the city units in each boundary to those following a rank-size population distribution (Rozenfeld et al., 2011) by finding the minimum population value above which a power law could be assumed to apply, see Appendix. These estimated minimum population cut-offs for each boundary definition and the OLS estimations for the units with populations above them are included in Table 1. It can be seen that while the overall regimes for β_{A_n} and β_Y are in broad agreement with the expectations developed and observed by Bettencourt and Lobo (2013; 2016), the OLS estimates for the majority of the boundaries for both properties fall much closer to unity. This is especially pronounced in the decreasing trend of β_Y estimates at larger scales as population density cut-off decreases.

Table 1. Summary of the boundary definitions used and the estimated scaling exponents β_{A_n} and β_Y.

These deviations from prescribed idealised values of the scaling exponents have previously been noted with exponents estimated for the UK lying much closer to unity rather than the expected values of $\frac{5}{6}$ and $\frac{7}{6}$ for sub- and super-linear scaling, respectively (Arcaute et al., 2015). The larger matter of the comprehensiveness of these particular estimates is part of a broader ongoing debate that also includes issues around the appropriate methods of defining the boundaries of cities (Masucci, Arcaute, Hatna, Stanilov, & Batty, 2015; Arcaute et al., 2016). These, however, do not affect the study presented here since the derivation of the performance balance measure set out previously is independent from the estimated values of the exponents[3]. Within the framework of the social reactor model, however, this prevalent linear scaling can be interpreted as a sign that cities in England and Wales on average exhibit a systemic pattern of impaired accessibility. As mentioned above, the extent of this lack of accessibility and mixing becomes increasingly larger at smaller population density cut-offs, e.g. $100\frac{N}{km^2}$, evident in the shrinking exponent estimates for the economic output. Nevertheless, the TTWA boundary estimates for GVA and urbanised land area scaling exponents show a close match to those prescribed by the model, more or less appearing to uphold the mixing population assumption. Incidentally, these boundaries which can be best described as self-contained functional economies (Coombes, 2010), also present the closest similarity in terms of definition to the idealised units modelled in the Settlement Scaling theory.

Figure 2 shows the estimates of G for individual city units across the boundary definitions against population on logarithmic axes. It can be seen that despite the range of population size that is covered across the boundary definitions the estimates of G remain more or less independent of population size ($\frac{dG}{dN} \approx 0$, $R^2 \in [0.00, 0.06]$) and within the same broad range across the different boundary definitions with an overall median value of around 6.5E+6 $\frac{£\cdot m^2}{N^2}$. The furthermost points to the right in each panel denote the different realisations of London and the Greater London Authority within each boundary definition. This regularity confirms the validity of the fourth assumption in Bettencourt's model and the model's broader relevance in the context of city units in England and Wales.

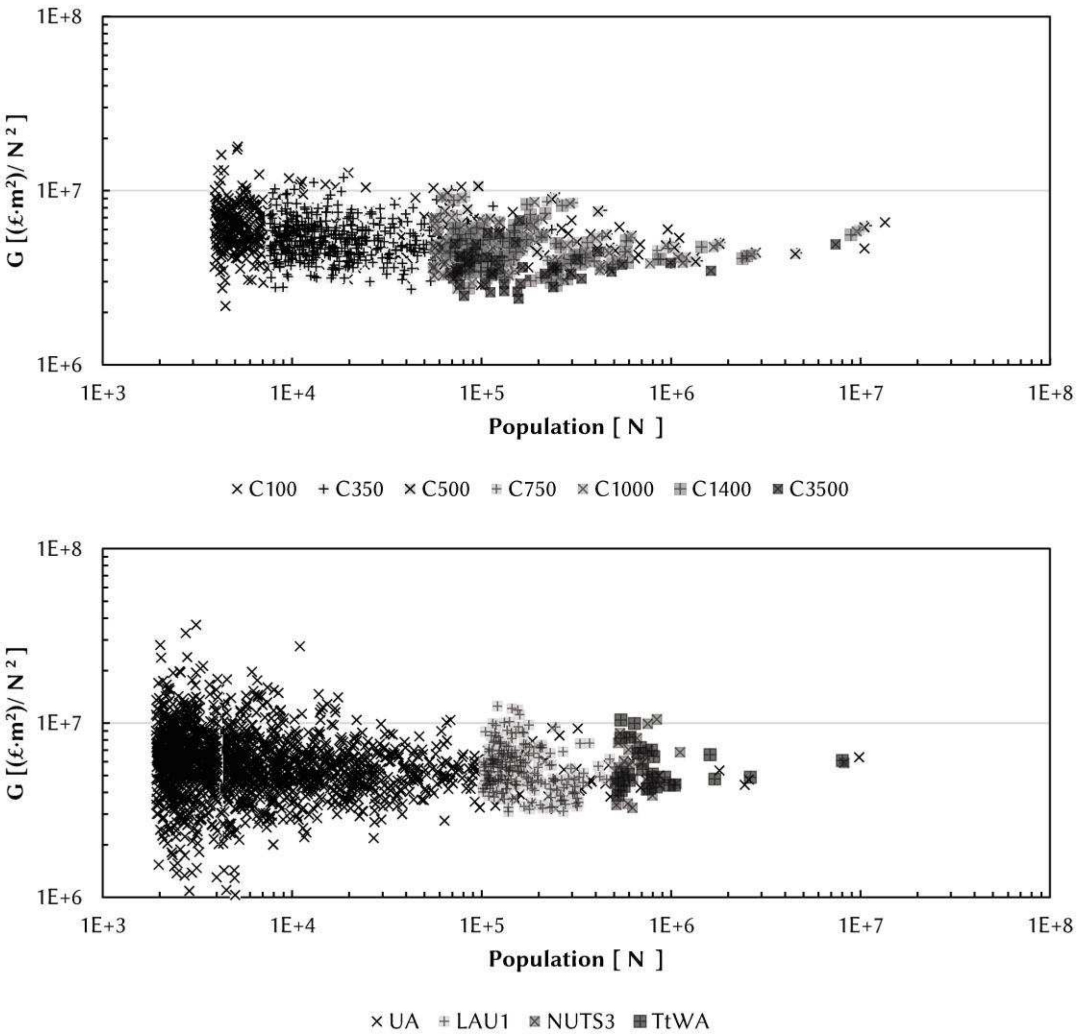

Figure 2. Superimposed plots of G against population for each boundary definition (note logarithmic axes).

Social reactor model theoretical optimum

To find the idealised optimal G^*, which maximises the balance between economic output and the mobility costs of its generation, OLS fits with constant gradients in accordance with those prescribed by the model, i.e. $\beta_{A_n} = \frac{5}{6}$ for urbanised area and $\beta_Y = \frac{7}{6}$ for the GVA, for idealised two-dimensional cities with full inhabitant mobility are implemented. We then take the product of the normalisation coefficients from each fitted line, i.e. the intercepts A_{n0} and Y_0, to represent the theoretical optimal G^* as set out in equation 5.

Figure 3 summarises the distribution of the ratio $\frac{G}{G^*}$, denoted as $\eta = \log\frac{G}{G^*}$, for city units in each boundary. From a first glance, it is clear that estimates of G do indeed tend to cluster close to the optimal value that maximises the urban size-cost balance for idealised cities. A secondary observation can be made regarding the larger portion of the G estimates lying below the optimum highlighting a shortcoming in adequate levels of mobility and access in the city units across the boundary definitions used. This is more easily demonstrated by looking at the percentages of city units at different intervals of η where negative values indicate increasing lack of adequate mobility and mixing compared with the comparable idealised urban unit while positive values indicate higher needs for increased built-density, Figure 4. More than half of the units in density-based boundaries with cut-offs larger than 750 $\frac{N}{km^2}$, the two administrative boundaries, and the Travel-to-work Areas show ratios below the optimum.

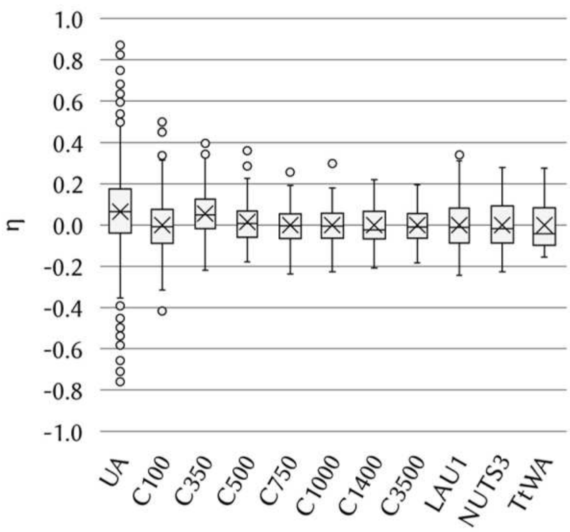

Figure 3. Box-chart showing the distribution of $\eta (\equiv \log\frac{G}{G^*})$ within each boundary definition.

A cursory inspection of the units in the UA and C350 boundaries, that exhibit larger portion of cities with $\eta > 0$, indicates this larger portion consists of city units often of a small population that are near larger units or in close proximity of a number of other similarly small units where the economic output is effectively not a product of the interactions within single individual units and would involve interactions and commutes between units or to larger nearby conurbations. This can be verified by estimating η for city units discarded previously with populations below the minimum cut-offs indicated in Table 1. For this comparison, we do not re-estimate the theoretical point of optimum anew rather we use the theoretical optimum obtained for the larger 'urban' units to quantify the notional performance balance of all city units compared with that of the average urban ideal, right panel Figure 4[4]. This extension results in increases in the portion of units with larger than optimal η ratios especially in boundaries that would include large numbers of small city units on the periphery of larger ones, i.e. UA, C100, and C350. The move from the smaller density cut-offs in C100 to those in C3500 in essence eliminates the satellite commuter suburbs where as mentioned gains in GVA are not achieved over their own urbanised area.

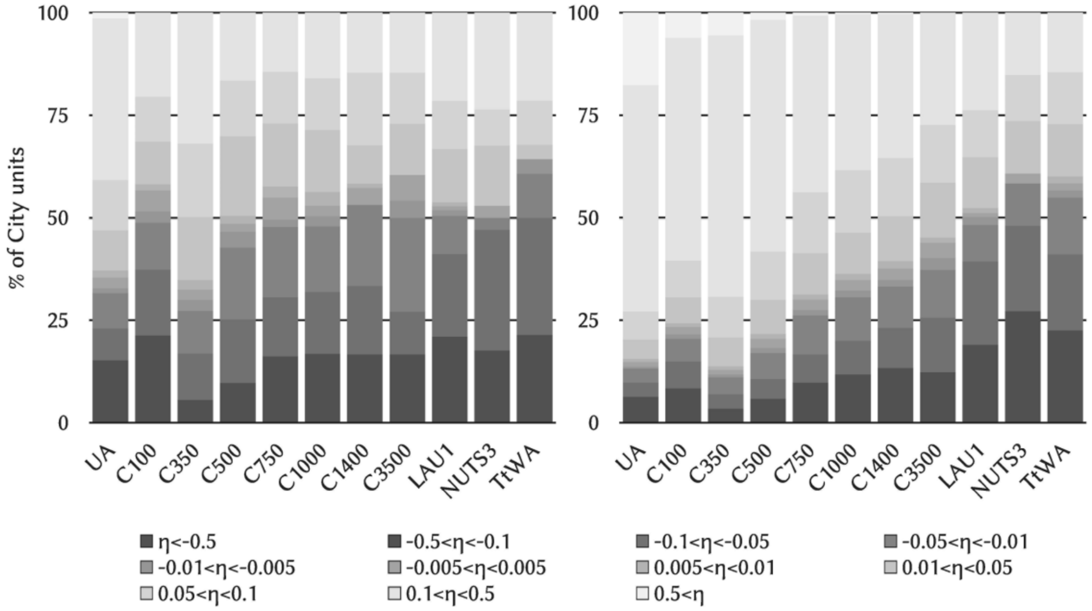

Figure 4. Bar charts showing the percentage of city units within the indicated range of η – left: city units above the population cut-offs in Table 1, right: all city units.

Finally, we geographically contextualise this optimality comparison by mapping each boundary definition and the corresponding ratio estimates. Figures 5 and 6 illustrate these for C100, C500, C1000, C1400, and the other four non density-based boundaries[5]. Note that the maps show estimated η for all city units within each boundary definition and not just those above population cut-offs indicated in Table 1. The first visual pattern to be immediately evident, especially in the density-based boundaries, is the change from below optimal ratios to those over the optimum crudely separating the south-east from the rest of EW. Another notable observation is that the units corresponding to the Leeds and/or its greater city region are the only major urban centres in the north exhibiting $G > G^*$ and as such, the only northern urban core indicating a need for densification to improve its size-cost balance rather than improvements to the intra-city transport similar to the rest. Additionally, subsequent disaggregation of the larger city unit of the north, note that in C100 the combined areas of Liverpool through Leeds and then downwards through Nottingham are identified as a single city unit, in the density-based boundaries as the density cut-off is continually raised from 100 to 3500 $\frac{N}{km^2}$, does not appear to affect the identified size-cost balance where a need for better mobility persists despite the changing scales. These remain largely stable even when comparing the corresponding boundaries in the non-density-based boundary definitions in Figure 6.

Figure 5. Maps of density-based boundaries colour-coded based on the range of η. From left to right C100 and C500 at the top and C1000 and C1400 at the bottom – Contains National Statistics and OS data © Crown copyright and database right 2017.

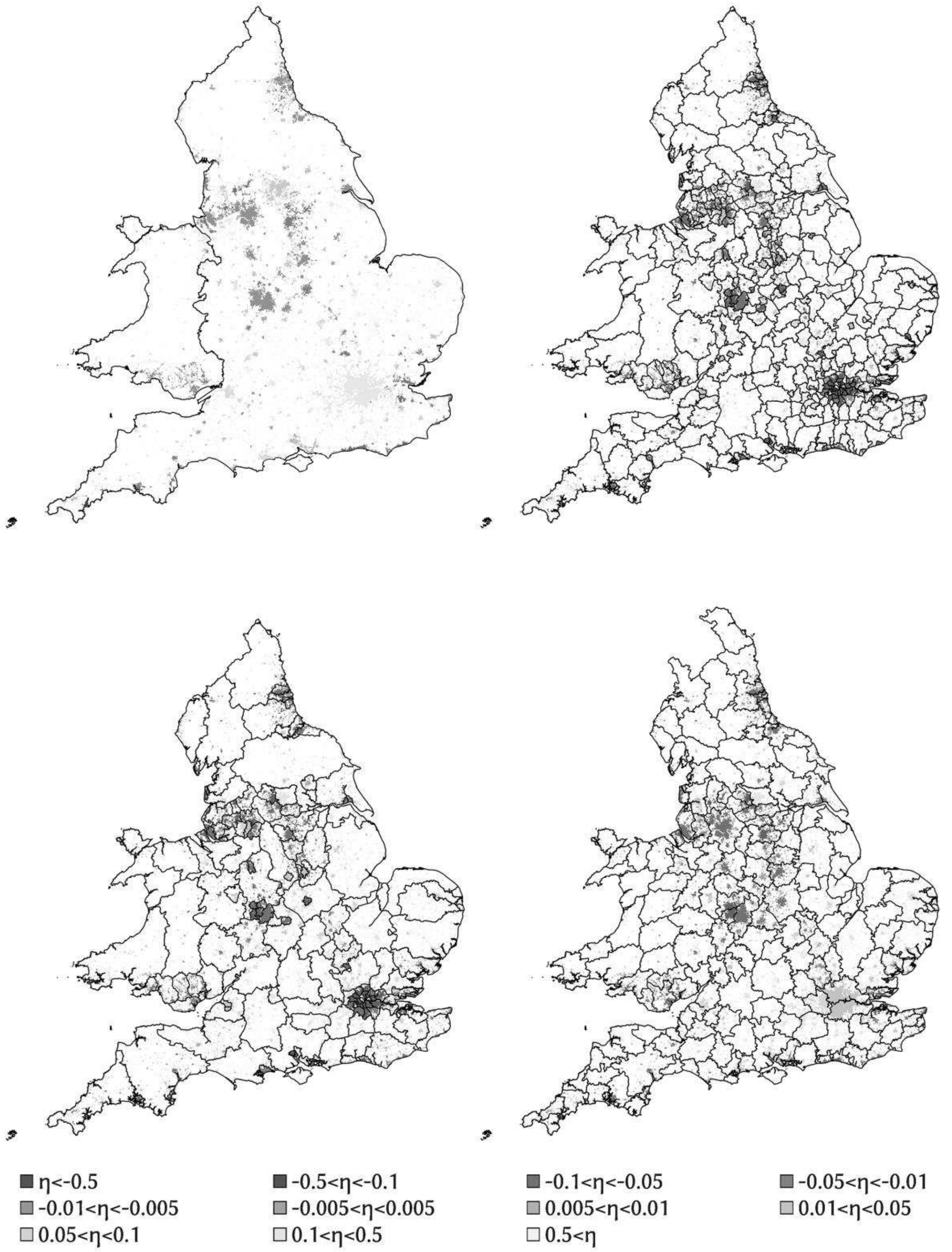

Figure 6. Maps of city units colour-coded based on the range of η. From left to right UA and LAU1 at the top and NUTS3 and TTWA at the bottom – Contains National Statistics and OS data © Crown copyright and database right 2017.

Discussion

The planning policy in EW is being driven with the emphasis on connecting the under-performing cities through improved transport infrastructure. As mentioned, this is seen as fundamental in enabling these regions to perform as a single functional economy and as such contributing towards the rebalancing of the national economy (National Infrastructure Commission, 2016). These have precipitated in transport, specifically inter-city connectivity, building up the largest portion of the infrastructure pipeline 2017 onwards with project prioritisation focused on reducing current travel time and reacting to the existing capacity demand while identifying the city regions with the highest economic opportunity associated with their inter-city connection (Infrastructure and Projects Authority, 2015).

The concluding observations from the previous section, however, noted a persistent lack of adequate mixing, or in other words a need for an improvement in the extent of the mobility provisions, in a majority of these regions regardless of the scale at which city regions are considered from LAU1s to larger TTWAs or density-based units. This is important when considering the generic recommendations borrowed from agglomeration theory regarding inter-city transport policy. The overall transport and connectivity focus of such insights appears in agreement and supported by the social reactor model's interpretation of the current size-cost balance in EW across spatial scales. The inter-city focus of stylised agglomeration principles, however, overlooks the overall performance balance, as formulated by Bettencourt (2013), and as such infrastructural needs across smaller scale boundaries, Figures 5 and 6.

As an illustration, considering the C100 or C500 boundaries from Figure 5, centre-to-centre inter-city transport links connecting Liverpool, Manchester, Sheffield, and Leeds can be seen as beneficial. They would improve the performance balance as these regions appear as a single metro region with an apparent lack of appropriate connectivity plaguing their size-cost performance that an inter-city mobility scheme could potentially remedy. All the while, the individual incarnations of cities building up these areas in other boundaries, for the exception of Leeds, also show the same requirement for better mobility across smaller areal extents. This is indicative of a lack of accessibility at different levels starting from within the high-density core areas, e.g. those in C1400, and persisting at larger scales, e.g. those in C100 or TTWA.

With this in mind, transport-led agglomeration, as is often articulated as facilitating connectivity between major city centres, involves mobilising populations into city units that may not individually have the transport capability to provide for the efficient mobility that is implicit in agglomeration theory and conducive to the improved size-cost performance of the overall aggregated regions. Although such single-scale interventions could perhaps increase economic output nominally, the size-cost analysis suggests that they would do so to the detriment of the overall comparative balance at other scales. In EW where the city centres have had the largest population growth in the last decade and accommodate the bulk of employment opportunities as the suburbs and rural areas provide the residential housing (Thomas, Serwicka, & Swinney, 2015), multi-scale, i.e. intra-city and inter-city, infrastructural interventions provided concurrently would seem more coherent. Considering practicalities such as a limited funding bandwidth, prioritising policy interventions to start from smaller scales and moving on to larger ones would adjust the size-cost performance more effectively since improved mobility at an intra-city scale facilitates inter-city access while inter-city access would only increase demand on existing intra-city infrastructure.

This is not to say that the inter-city infrastructure is not needed or to imply that it constitutes an entirely wrong strategy. In fact for any pair of cities where one exhibits

$G > G^*$ and the other $G < G^*$, the model used in this study would project an estimate of G closer to G^* for the hypothetical and idealised city region which would have the sum of the pair's population, urban extent, and economic output. Thus, assuming Leeds and Manchester areas comprise a well-connected metro region, the model would project a better-balanced size-cost performance for this hypothetical city. There is, however, an implicit assumption in the model used here that the resulting aggregated metro region is in itself a uniform urban conurbation providing for an ideal population mixing meaning for real-life examples balancing the performance of individual units would be required prior to a natural merging of the regions. In a sense, the agglomeration economies principle would perform as intended when the units are themselves performing well at smaller scales prior to connection at larger scales so that the aggregation helps to introduce the efficiencies and productivities of higher populations. These effects will not inexplicably overcome under-performance of the contributing cities if they are not previously addressed. This vital issue and importance of mobility at an intra-city scale is only often acknowledged in passing (National Infrastructure Commission, 2016).

On a slightly different note, one might question the degree to which approaching G^* is desirable and practical. Despite its more tangible formulation of what essentially are 'congestion costs', Bettencourt's model aggregates all costs associated with population mobility over the infrastructure network. The energy dissipated and the overall size-cost balance, G, then have to include, for instance, fuel/energy source, type, and cost bundled together. In a context where most mobility solutions are fossil-fuel intensive and concerns for the effects of climate change exist, maximising the economic output for the transport energy lost becomes an imperative. In such cases, G^* embodies this maximisation point and target. However, for the same targets, policy could focus on decoupling modes of mobility and transport from their fuel sources instead. As an extreme illustration, if similar levels of mobility could be provided through freely available public transport run by renewable energy sources the relevance of a $Y - W$ balance becomes diminished significantly. This would mean the optimum point of G^* may not be a practical target in situations where cities are indicating estimates more than the optimum since escalating mobility costs will not have the same tangibly negative implications.

In a broader context, our study also points to a different facet of the north-south divide in EW. This division of economic output can be reformulated with a size-cost perspective and seen as long-term planning needs. Despite higher economic outputs, economic success in a majority of the south-east appears to be achieved through mounting mobility costs as compared with idealised urban cities of the same population that would have exhibited smaller urbanised areas. Our examination of Bettencourt's explicit formulation of these balances shows that the cities in the southeast and London particularly have in fact grown too large and require built-up area densification. This is in contrast with Cheshire's (2013) recent criticism of densification and urban containment strategies labelling them theoretically grounded but without empirical grounds. Moreover, Arcaute et al. (2016) use percolation at different distances on the UK road network to obtain a hierarchal classification of the road transport network and by extension the clustering of the geographical regions as represented by their road connectivity where the EW network initially collapses into one radiating out from London connecting the southern regions and the other connecting the North, Wales, and Cornwall. A similar classification pattern can be observed in Figure 5 and 6 to some extent where the southeast exhibits an overwhelming need for more compactness, especially along the radiating motorways, while the majority of the north suffers from poor intra-city connectivity and mobility.

We would be remiss, however, if we did not point out the remaining shortcomings and potential direction. The majority of models from the same family of the one used here start from the assumption that the units under study are in fact uniformly urban and functional economic catchments (Bettencourt, 2013; Yakubo et al., 2014; Gomez-Lievano et al., 2016). Unfortunately, this leaves them highly sensitive to the urban population count at each spatial scale and hence the choice of boundary used in that scale (Arcaute et al., 2015). Although it should be noted that while such fluctuations were observed in the G estimates from the model used in this study when considering slightly different boundaries at similar scales, the determination of the planning needs relative to the idealised city remained consistent. Louf and Barthelemy (2014) argue that until a comprehensive, universal, and mechanistic understanding of how cities are formed, evolve, and function is developed such models should not be used in shaping policy advice. However, given that planning policy will be formed one way or the other and that the current economic agglomeration models informing policy not only suffer from the same fundamental lack of universality but are also as previously mentioned placeless and single-scale in nature, providing and considering alternative pictures of city performance and infrastructural needs would benefit the overall policy and planning debate.

Conclusions

We have characterised the potential influence of spatial scale on overall city performance especially when considering the provision of intra-city or inter-city transport solutions. There are a number of insights that emerge from our study. Improvements of inter-city mobility and access based on agglomeration principles are often argued for boosting productivity and economic performance. We have, however, demonstrated that the part played by inadequate mobility in regional economic under-performance appears to permeate through the change of spatial scales starting at smaller intra-city scales. In light of this, we strongly argue that planning transport infrastructure solutions at regional levels cannot only rely on the consideration of larger spatial scales alone and policy efforts should require an examination of planning needs across scales particularly when larger-scale solutions could hide root causes of unbalanced performance at scales below them.

Finally, we have demonstrated the use of Bettencourt's social reactor model, one amongst many within the data driven science of cities, in identifying such regional infrastructural requirements that would maintain size-cost performance balance over multiple spatial scales in the context of the urban system in England and Wales. Referring back to the disconnected conceptual spectrum of the infrastructural planning scales, we have shown the model's potential to provide a way of bringing together the macro-scale agendas with the micro-scale planning models at a meso-scale where the consistency of the planning strategies promoted by larger-scale concerns can be scrutinised across a variety of intermediate levels rather than a limited grouping of spatial data.

Acknowledgements

The authors wish to thank Philip McCann for helpful discussions and his comments on the manuscript. Population data are from Census 2011, Office for National Statistics, and maps contain OS Map data © Crown copyright and database rights 2017.

Notes

[1] Interested readers are encouraged to view the detailed derivation of the model based on a hierarchical urban infrastructure network and the optimisation of output-energy balance in the supplementary materials of (Bettencourt, 2013).

[2] Although not of importance to the focus of this paper, it should be noted that the simplicity of the OECD method might cause problems when aggregating back up to units not so larger than the base population layer resulting in linear scalings or noise recordings (Smith,

2014), due to the nature of the simple population proportionality and the uniform density distribution assumption in equation 7, see Appendix for an expanded discussion.

[3] It should be noted that the numerical value of the theoretically optimal G^* is not independent of the exponents observed for economic output and urbanised area. It is the overall maximisation of $Y - W$ that does not depend on specific values of the exponents.

[4] OLS regression estimates of scaling exponents using all units in each boundary are available in the appendix for those interested.

[5] Larger Figures are available in the online appendix.

References

Abel, J. R., Dey, I., & Gabe, T. M. (2012). Productivity and the Density of Human Capital. Journal of Regional Science, 52(4), 562–586.

Arcaute, E., Hatna, E., Ferguson, P., Youn, H., Johansson, A., & Batty, M. (2015). Constructing cities, deconstructing scaling laws. Journal of The Royal Society Interface, 12(102), 20140745.

Arcaute, E., Molinero, C., Hatna, E., Murcio, R., Vargas-Ruiz, C., Masucci, A. P., & Batty, M. (2016). Cities and regions in Britain through hierarchical percolation. Royal Society Open Science, 3(4), 150691.

Bettencourt, L. (2013). The Origins of Scaling in Cities. Science, 340(6139), 1438–1441.

Bettencourt, L., & Lobo, J. (2016). Urban scaling in Europe. Journal of The Royal Society Interface, 13(116), 20160005.

Bettencourt, L., Lobo, J., Helbing, D., Kühnert, C., & West, G. B. (2007). Growth, innovation, scaling, and the pace of life in cities. Proceedings of the National Academy of Sciences, 104(17), 7301–7306.

Bettencourt, L., & West, G. (2010). A unified theory of urban living. Nature, 467(7318), 912–913.

Burger, M. J., Meijers, E. J., Hoogerbrugge, M. M., & Tresserra, J. M. (2015). Borrowed Size, Agglomeration Shadows and Cultural Amenities in North-West Europe. European Planning Studies, 23(6), 1090–1109.

Cheshire, P. (1999). Chapter 35 Trends in sizes and structures of urban areas. Handbook of Regional and Urban Economics, 3, 1339–1373.

Cheshire, P. C. (2013). Land market regulation: market versus policy failures. Journal of Property Research, 30(3), 170–188.

Clauset, A., Shalizi, C. R., & Newman, M. E. J. (2009). Power-Law Distributions in Empirical Data. SIAM Review, 51(4), 661–703.

Coombes, M. (2010). Defining Labour Market Areas by Analysing Commuting Data: Innovative Methods in the 2007 Review of Travel-to-Work Areas. In J. Stillwell, O. Duke-Williams, & A. Dennett (Eds.), Technologies for Migration and Commuting Analysis: Spatial Interaction Data Applications (pp. 227–241). Hershey, PA: IGI Global. Retrieved from http://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/978-1-61520-755-8

Coombes, M., & Office for National Statistics. (2015). Travel-to-Work Areas (Report No. RR2015/05) (p. 20). Newcastle, UK: Centre for Urban & Regional Development Studies (CURDS).

Echenique, M. H., Grinevich, V., Hargreaves, A. J., & Zachariadis, V. (2013). LUISA: A Land-Use Interaction with Social Accounting Model; Presentation and Enhanced Calibration Method. Environment and Planning B: Planning and Design, 40(6), 1003–1026.

Gardiner, B., Martin, R., Sunley, P., & Tyler, P. (2013). Spatially unbalanced growth in the British economy. Journal of Economic Geography, 13(6), 889–928.

Glaeser, E. L., Ponzetto, G. A. M., & Zou, Y. (2016). Urban networks: Connecting markets, people, and ideas. Papers in Regional Science, 95(1), 17–59.

Gomez-Lievano, A., Patterson-Lomba, O., & Hausmann, R. (2016). Explaining the prevalence, scaling and variance of urban phenomena. Nature Human Behaviour, 1, 0012.

Infrastructure and Projects Authority. (2015). National Infrastructure Pipeline factsheet. London, United Kingdom: HM Treasury. Retrieved from https://www.gov.uk/government/publications/national-infrastructure-pipeline-july-2015/national-infrastructure-pipeline-factsheet-july-2015

Jones, C. (2016). Spatial economy and the geography of functional economic areas. Environment and Planning B: Planning and Design, 0265813516642226.

Louf, R., & Barthelemy, M. (2014). Scaling: Lost in the Smog. Environment and Planning B: Planning and Design, 41(5), 767–769.

Martin, R., Pike, A., Tyler, P., & Gardiner, B. (2015). Spatially rebalancing the UK economy: The need for a new policy model. In Regional Studies Association Pamphlet. Retrieved from http://www.regionalstudies.org/uploads/documents/SRTUKE_v16_PRINT.pdf

Masucci, A. P., Arcaute, E., Hatna, E., Stanilov, K., & Batty, M. (2015). On the problem of boundaries and scaling for urban street networks. Journal of The Royal Society Interface, 12(111), 20150763.

McCann, P. (2016). The UK Regional–National Economic Problem: Geography, globalisation and governance. Oxon, UK: Routledge.

National Infrastructure Commission. (2016). High Speed North: A National Infrastructure Commission Report. National Infrastructure Commission. Retrieved from https://www.gov.uk/government/publications/high-speed-north-a-national-infrastructure-commission-report

O'Brien, P., & Pike, A. (2015). 'The governance of local infrastructure funding and financing'. Infrastructure Complexity, 2(1), 3.

OECD (Ed.). (2012). *Redefining 'Urban': A New Way to Measure Metropolitan Areas*. Paris: OECD.

Office for National Statistics. (2014, December 10). Regional GVA NUTS3. Retrieved 25 April 2016, from https://www.ons.gov.uk/economy/grossvalueaddedgva/datasets/regionalgvanuts3

Office for National Statistics. (2016a, August 24). Local Administrative Units Level 1 Boundaries in England and Wales. Retrieved 24 October 2016, from http://geoportal.statistics.gov.uk/datasets/29124941855a42d689cc22859a15e581_0

Office for National Statistics. (2016b, August 30). NUTS Level 3 in England and Wales. Retrieved 24 October 2016, from http://geoportal.statistics.gov.uk/datasets/a75f6de7006d4472a72800bbf8aada1d_0

Office for National Statistics. (2016c, October 3). GEOSTAT Grid Boundaries. Retrieved 24 October 2016, from http://geoportal.statistics.gov.uk/datasets/506b400adfaa4e8d9df59da5d0064702_0

Rowthorn, R. (2010). Combined and Uneven Development: Reflections on the North–South Divide. Spatial Economic Analysis, 5(4), 363–388.

Rozenfeld, H. D., Rybski, D., Gabaix, X., & Makse, H. A. (2011). The Area and Population of Cities: New Insights from a Different Perspective on Cities. American Economic Review, 101(5), 2205–2225.

Samaniego, H., & Moses, M. E. (2008). Cities as Organisms: Allometric Scaling of Urban Road Networks. Journal of Transport and Land Use, 1(1). Retrieved from https://www.jtlu.org/index.php/jtlu/article/view/29

Seto, K. C., Güneralp, B., & Hutyra, L. R. (2012). Global forecasts of urban expansion to 2030 and direct impacts on biodiversity and carbon pools. Proceedings of the National Academy of Sciences, 109(40), 16083–16088.

Smith, D. A. (2014). Domestic energy use in England and Wales: a 3D density grid approach. Regional Studies, Regional Science, 1(1), 347–349.

Szüle, J., Kondor, D., Dobos, L., Csabai, I., & Vattay, G. (2014). Lost in the City: Revisiting Milgram's Experiment in the Age of Social Networks. PLOS ONE, 9(11), e111973.

Thomas, E., Serwicka, I., & Swinney, P. (2015). Urban Demographics: Where People Live and Work (p. 30). London, UK: Centre for Cities. Retrieved from http://www.centreforcities.org/wp-content/uploads/2015/07/15-07-20-Urban-Demographics.pdf

Transport for the North. (2016). The Northern Transport Strategy: Spring 2016 Report (p. 63). London, United Kingdom: Department for Transport, HM Government. Retrieved from https://www.gov.uk/government/publications/northern-transport-strategy-spring-2016

Yakubo, K., Saijo, Y., & Korošak, D. (2014). Superlinear and sublinear urban scaling in geographical networks modeling cities. Physical Review E, 90(2).

Table 1.

Boundary		No. of units (N_{min})*	β_{An} – 95%CI	β_{Y} – 95%CI	R-squared β_{An}, β_{Y}
C100	Population density-based contiguous grid	586 (3895)	0.94 [0.92 0.95]	1.01 [1.00 1.02]	0.97, 0.98
C350		480 (7627)	0.94 [0.93 0.96]	1.02 [1.01 1.03]	0.98, 0.97
C500		103 (59698)	0.96 [0.94 0.98]	1.02 [0.97 1.06]	0.99, 0.96
C750		111 (57698)	0.96 [0.94 0.97]	1.02 [0.97 1.06]	0.99, 0.95
C1000		119 (55031)	0.95 [0.93 0.97]	1.03 [0.98 1.07]	0.99, 0.95
C1400		96 (67495)	0.96 [0.93 0.98]	1.03 [0.98 1.09]	0.99, 0.94
C3500		48 (66671)	0.95 [0.92 0.99]	1.07 [1.00 1.14]	0.98, 0.95
UA[1]	Contiguous built-up area	1787 (1913)	0.94 [0.93 0.95]	1.00 [1.00 1.01]	0.91, 0.97
LAU1[2,+]	Administrative	214 (101355)	0.86 [0.82 0.91]	1.02 [0.96 1.08]	0.88 0.84
NUTS3[3,+]		34 (499766)	0.79 [0.70 0.88]	1.29 [1.13 1.45]	0.90, 0.89
TTWA[4]	Functional economy	28 (510149)	0.84 [0.76 0.91]	1.14 [0.99 1.29]	0.95, 0.91

* Values in parentheses denote the population cut-off for the smallest unit within each boundary definition when used for estimating the exponents

1 Urbanised Area – based on the built-up area boundaries from December 2011

2 Local Administrative Units Level 1

3 Nomenclature of Units for Territorial Statistics Level 3

4 Travel-to-Work Areas

+ Constituting boundary units for the Greater London Authority have been aggregated and treated as one data point in these boundaries instead of treating boroughs or NUTS3 boundaries as separate cities.