CENTRE *for* ECONOMIC PERFORMANCE

Urban Capacity and Economic Output

Stephen Gibbons and Daniel J. Graham

Special Paper No. 36

August 2018

Centre for Economic Performance Special Paper
Stephen Gibbons is a Professor of Economic Geography at the LSE and Director of the Urban Programme at the Centre for Economic Performance, LSE. Daniel J. Graham is Professor of Statistical Modelling, Director of the Statistical Modelling and Economics Research group, and joint Director of the Railway and Transport Strategy Centre within the Centre for Transport Studies at Imperial College London.

This report was commissioned by the National Infrastructure Commission as part of the evidence base for the National Infrastructure Assessment. The views expressed and recommendations set out in this report are the authors' own and do not necessarily reflect the position of the National Infrastructure Commission.

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

LSE THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

1. Background

One of the most globally pervasive empirical regularities describing the spatial economy is that, on average, firms and workers in cities are economically more productive than those in rural areas, and firms and workers in big cities are more productive than those in smaller ones. These productive benefits of concentrating activity in large places are known as 'agglomeration economies' and are thought to occur through a number of channels related to the sharing of infrastructure, markets and information (sharing, matching and learning in the theoretical terminology of Duranton and Puga 2004). Large cities can also generate benefits for consumers in terms of more product market competition and provision of amenities (Glaeser at al 2001).

The productive benefits of cities may depend not just on their size, but on their composition in terms of industrial structure and mix of skills in the labour force. The distribution of individual skills and industries across cities could affect the productivity of workers in each city and so the output of each city as a whole. The share of educated workers in a city could affect individual productivity and the magnitude of this effect could differ by a workers' own skill level (human capital externalities, Moretti 2004). Concentration of industries into specialised cities might boost aggregate productivity if firms benefit from co-location with other firms (localisation economies), though for some industries location in cities with a diversified industrial mix might be more beneficial (urbanisation economies). Concentration of firms and workers into particular places, increasing local density within cities – usually the central business district – is also a potential driver of higher productivity.

Despite the apparent economic benefits of large cities, size is limited by natural costs that arise as city sizes increase. These costs include those involved in transporting people and goods within the city, building houses and commercial premises at higher density and extending energy and water infrastructure at the margins of cities. There are also many negative externalities from crowding activity in confined spaces, such as congestion, crime and air pollution.

Some of the costs are partly controllable by government policy, through the provision of infrastructure and the planning regulations that permit or restrict housing and commercial development. If industrial composition matters, the agglomeration benefits could also be affected, in principle, by appropriate zoning policies. Infrastructure and planning can therefore potentially influence the size and composition of cities, the productivity of workers in these cities, and output for the economy as a whole (Cheshire, Nathan and Overman 2014).

2. Outline of work

The ideas in Section 1 raise a question over whether there is an 'optimal' distribution of populations (or employment) across cities and within cities. It is difficult to answer this question in a general sense – taking into account all the costs and benefits of city sizes, and the heterogeneity in these benefits and costs across different types of person and firm. An extensive theoretical literature and empirical literature exists. Henderson (1988) provides an early discussion of the issues. However, the detail of the underlying relationships, while the subject of a large body of evidence, is just not understood in enough detail to make meaningful statements that would be relevant to policy. Answering it would require a fully specified spatial general equilibrium model of the economy, and the policy relevance of this would be limited by the lack of knowledge about the underlying relationships, making the predictions very model dependent.

However, one aspect that is better understood empirically is the magnitude of the effect of city size (or other measures of access to economic mass) on firm or individual productivity, typically measured by Total Factor Productivity, labour productivity or wages (Combes and Gobillon 2015). It is therefore straightforward to describe how the distribution of city sizes translates into the output per worker in each city, the total output of each city and the total output of the economy – holding other factors such as worker skills and city industrial composition constant. This provides a very broad-brush indication of the gains and losses from relaxing or increasing constraints on the size of cities. As noted above, and in Glaeser and Gottlieb (2008) this is not necessarily informative about the policy implications for economic welfare, but does tell us something about the potential policy implications for production and nominal income.

In a world with agglomeration economies and no other constraints the trivial answer to the most productive allocation of people and firms across space in the presence of agglomeration economies, is that everyone should be concentrated in one mega city. This follows because, if average output of a worker in a city of size n is described by a process $y = \mathrm{f}(n)$ with y increasing monotonically with n, then the output of a city is $ny=n\mathrm{f}(n)$. City total output exhibits increasing returns to city size, which implies that the total production of two unequal cities is always higher than the total production of two equal sized cities. In a two city economy with total population N, total output is $Y = n\mathrm{f}(n) + (N\text{-}n)\mathrm{f}(N\text{-}n)$. It is straightforward to demonstrate that output is maximised with all production in one place (i.e. $n= N$ or $n=0$). Of course, this is not a very realistic proposition, and we assume that government policy makers will be unwilling move everyone in the UK to London or some other arbitrary city to maximise the benefits of agglomeration in production, even if the costs associated with supporting a city of this size could be overcome.

An additional question is to what extent the distribution of workers within a city affects productivity. The evidence on the existence of agglomeration economies at a city level, and the existence of dense clusters of productive economic activity in specific places within cities – particularly city centres – has led to the speculation that agglomeration economies exist at localised level and that firms and workers benefit from working in close proximity within cities. Therefore, policy that facilitates the reallocation of jobs across space within cities could also influence productivity of workers, firms, cities and the national economy. For example, infrastructure policy that reduces costs of travelling to a city centre might encourage firms and workers to relocate to the centre (although in the absence of other planning policy, it might also encourage firms to relocate further out, and use the transport system to connect with other firms in the city centre). The empirical literature on these within-city agglomeration economies is limited, despite the idea being frequently used to justify industrial cluster policy and to justify transport schemes that move jobs into central city locations.

Our methodology builds on these ideas by using simple simulations to demonstrate how different distributions of city sizes across Britain and different distributions of jobs within a city, leads to different levels of economic output at city level and at the national level. The functional forms and parameters needed to inform these simulations are derived from the existing literature on agglomeration economies. The inputs in terms of city sizes can be taken from a) any number of hypothetical scenarios based on redistributing employment around geographical areas across the country, or b) existing or modified projections of future employment in different city areas. The outputs are graphs and tables of the distribution and level of output under different population scenarios.

While the outputs of such a simulation are simplistic and highly stylised, they can be useful in terms of demonstrating the magnitude of the gains and losses in economic output associated with different distributions of city sizes. They will help lay to rest some misconceptions about the potential role of agglomeration economies, and help with visualisation of the implicit equity-efficiency trade off.

Although straightforward in the simple case where productivity depends only on city size, the picture is much more nuanced and complicated when other city characteristics come into play. To incorporate these industrial and skill structure aspects into these scenarios is a much bigger challenge. This exercise requires estimates of the parameters characterising the relationship between the city skill mix and the productivity of workers, and this relationship is likely to vary according to individual skills. The exercise would also require estimates of how a city's industrial mix affects the productivity of firms, and how this varies according to the firm's own industry. Unfortunately the literature on these issues is not as clear cut as that on city size and there is no consensus on the magnitudes or existence of these more subtle forms of externality. We do not incorporate these aspects into the methodology within this project.

3. Methods

The specification use in this modelling exercise assumes that the productivity of a worker working in a sub area ij within a city j is determined by a combination of city wide 'agglomeration' effects at the city level j, and more localised, within-city density effects working at the sub-city level ij. City level agglomeration is represented by the total number of workers N_j in the city (e.g. a Travel to Work Area). Sub-city density effects are represented by the areal density of workers in sub-area i in city j, n_{ij}/a_{ij} where n_{ij} is the number of workers in the sub-city area and aij is is land area. We allow for two sub-city areas per city, corresponding to the 'city centre' (i=1) and the 'outer city' (j=1). By construction, the numbers of workers in the centre and outer zones within each city add up to the total number of workers in the city, N_j (and the land areas also add up). There are J total cities in the study area (e.g. England and Wales).

This modelling assumption allows us to think about the effects of redistributing workers across cities, and changing the distribution of workers within cities. Following the agglomeration literature, the underlying assumption used in the prediction is that the productivity of the average worker in a geographical area i within a city j is determined by a standard Cobb-Douglas production function of the form:

$$\bar{y}_{ij} = C_{ij} N_j^{\beta} \left(\frac{n_{ij}}{a_{ij}} \right)^{\gamma}$$

where $\overline{y}_{ij}$ is the average worker output in sub-area i in city j, N_j is the city total workforce, and n_{ij}/a_{ij} is the sub-city area worker density. C_{ij} represents other factors at city and sub-city area level (including worker average characteristics) which we will assume to stay constant under alternative modelled scenarios.

The parameter beta is the standard 'agglomeration elasticity', estimates of which are available from a wide range of literature (see the literature review and Combes and Gobillon 2015). We refer to this as the 'size' elasticity as it relates city size to productivity. Parameter gamma is the elasticity of worker productivity with respect to sub-area density, conditional on city size. We refer to this as the 'density' elasticity. Fewer estimates of this parameter are available from the literature. These parameters can be set when generating the simulations, to any value desired.

Given there are n_{ij} workers in each area ij , the total output of sub-area ij is:

Output $y_{ij} = C_{ij} N_j^{\beta} \left(\frac{n_{ij}}{a_{ij}} \right)^{\gamma} n_{ij}$

To obtain the city output, we simply sum up across sub-areas i within each city j:

City output $Y_j = \sum_{i=0}^{1} y_{ij}$

To obtain total national output (or total output in a subset of cities) we add this up across cities

Total output $\hat{Y} = \sum_{j=1}^{J} Y_j$

Simulating city outputs requires only that we have inputs in terms of N_j, n_{ij} and a_{ij}, plus parameters beta and gamma. Note, it is not necessary to have estimates of C_{ij}, because the aim is to model the distribution of city and aggregate outputs relative to some baseline scenario. Assuming C_{ij} remains constant under alternative scenarios, it drops out from the predictions of relative output. For example, if the area outputs in a baseline and alternative scenarios are:

$$y_{ij0} = C_{ij0} N_{j0}^{\beta} \left(\frac{n_{ij0}}{a_{ij0}} \right)^{\gamma} n_{ij0}$$

$$y_{ij1} = C_{ij0} N_{j1}^{\beta} \left(\frac{n_{ij1}}{a_{ij1}} \right)^{\gamma} n_{ij1}$$

then the relative outputs are

$$y_{ij1} / y_{ij0} = \left(N_{j1} / N_{j0} \right)^{\beta} \left(\frac{n_{ij1} a_{ij0}}{n_{ij0} a_{ij1}} \right)^{\gamma} n_{ij1} / n_{ij0}$$

The necessary inputs can be obtained from external sources for a range of alternative scenarios. Alternatively, outputs can be simulated for any hypothetical distribution of workers across cities, or between the centre and outer starting from the existing baseline distribution of workers across cities, we can reallocate workers across cities, and redistribute workers within cities.

An important point to note here is that this exercise is modelling what are sometimes called 'dynamic agglomeration economies' – i.e. the effects on productivity from policy which causes workers to move between or within cities, resulting in a change in the distribution of city sizes and density. The transport economics and appraisal literature also recognises that transport potential generates externalities – the so called 'wider benefits' of transport – by bringing workers and firms closer together in journey time, without the need for firms or workers to relocate to be physically closer together (see Venables, Laird and Overman 2014, or Graham and Gibbons 2018, for recent discussions). These 'static' agglomeration benefits from improved transport infrastructure are not incorporated in this analysis, because there is no explicit specification of any transport infrastructure improvements or connectivity changes.

4. Data sources and generation

4.1. Simulated TTWA employment distribution

Our first illustration of the potential scale of the agglomeration and density effects from redistributing the UK workforce between cities is based on simulated distributions. We define cities in this simulation as 2011 Travel to Work Areas (TTWAs). TTWAs are, in effect, commuting zones. There are 228 TTWAs in Britain, but we restrict attention to the largest 79 (with at least 100,000 workers, based on location of workplace), to represent major urban areas in Britain.

To simulate different workplace employment distributions across cities (distribution of N_j in Section 3), we start with 2011 Travel to Work Area (TTWA) workplace employment and land areas. We then redistribute total employment across TTWAs under different assumptions. We will show three scenarios: the current workforce distribution, and all workers in one TTWA (one mega city), an equal distribution across TTWAs, and a log normal distribution (the distribution which lies behind Zipf's law for city sizes, that characterises many city size distributions around the world).

To model within-city density, we generate two zones per TTWA, inner and outer. The shares of workers in these two areas are controllable using parameters that set the share of employment in the inner city zone, and the share of land area in the inner city zone. Increasing the share of workers in the inner zone, relative to its area, increases inner density relative to outer zone density. Simulations are then made using the equations provided in Section 3 above. Output is normalised relative to the predicted output under the current baseline employment distribution across TTWAs, assuming a uniform density within them (i.e. as if the TTWA is just one area without an inner and outer zone, or equivalently, as if the density elasticity is zero). Total number of workers in the country is the same under all scenarios.

Note these simulations assume that the land area of each TTWA is fixed at its 2011 value. So when workers are redistributed between TTWAs (either uniformly, into a single city or log normally), we assume that they fit into the existing TTWA commuting footprint. If we want to explore the effect of moving populations around while allowing these TTWA commuting boundaries to expand geographically, we need only switch off the density parameter (γ=0) such that the simulations provided by the equations in Section 3 ignore any effects from changing density.

4.2. Primary Urban Area employment scenarios under inner city capacity constraints

The second set of outputs related to a set of future working population distribution scenarios, which were generated by the Leeds Institute of Data Analytics (LIDA) in conjunction with the National Infrastructure Commission (NIC). These scenarios are based on ONS working population predictions for Local Authority Districts (LADs) in decades up to 2050. These ONS population predictions are based, fundamentally, on extrapolating from past trends so implicitly represent a 'business-as-usual' scenario in which policy, the supply of infrastructure and housing, and population growth and migration evolves in much the same way as it has in the past.

LIDA defined 55 bespoke core inner city centres associated with these LAD units – based on Census 2011 Output Area boundaries. LADs were then mapped to Primary Urban Areas (PUA) in England, and aggregated to give a dataset of LAD working population predictions for inner and outer Primary Urban Areas, plus a residual group made from non-PUA LADs in England, Scotland, Wales and Northern Ireland. This gives 111 geographical units (55 PUAs, inner and outer, plus one residual group).

Various assumptions were then made in forming a number of population distribution scenarios from present to 2050. The scenarios fall into two groups, one assuming the ONS projections as baseline, the second assuming the same general projection but with more urbanisation as a baseline (i.e. more working population in the urban LADs and PUAs). These baselines are then modified to give four scenarios, assuming greater constraints on within-city transport and inner district capacity than implied by the business-as-usual ONS projections. These constraints affect the distribution of working populations in urban centres versus outer areas, and also have some effect on the distribution between urban areas (and between the residual non-urban, Scotland, Wales and NI group). The total UK population is constrained to the same as the ONS baseline projection in all scenarios.

The scenarios are summarised and labelled below follows:

- *Business-as-usual baseline:* ONS working population projections.
- *Business-as-usual constrained:* ONS working population projections, but with moderate constraints on central city transportation and workforce capacity leading to less central city employment relative to employment in the outer urban areas in bigger cities.
- *Business-as-usual more constrained:* ONS working population projections, but with more severe constraints on central city transportation and workforce capacity leading to less central city employment relative to employment in the outer urban areas in most cities.

- *High-urbanisation baseline:* ONS working population projections modified to give a higher share of the population in urban areas.
- *High-urbanisation constrained:* ONS working population projections modified to give a higher share of the population in urban areas, and with moderate constraints on central city transportation and workforce capacity leading to less central city employment relative to employment in the outer urban areas in bigger cities.
- *High-urbanisation more constrained*: ONS working population projections modified to give a higher share of the population in urban areas, and with more severe constraints on central city transportation and workforce capacity leading to less central city employment relative to employment in the outer urban areas in most cities.

The exact details of how these scenarios were constructed are available from LIDA and NIC.

4.3. Parameter values

Many estimates of the agglomeration elasticity beta are available. Typically, modern figures are in the range of 0.01-0.05. A comprehensive survey is provided in Combes and Gobillon 2015 and our Appendix tabulates some of these estimates. Fewer estimates are available for the effects of within-city density, conditional on city size. One point of comparison is the report SERC (2009) which provides estimates for agglomeration effects based on the densities implied by inter-ward road travel times, conditional on inter-Local Authority rail journey times. This study provides estimates of 0.03-0.05 for the city level agglomeration elasticity, and 0.007 (excluding London from the sample) to 0.07 (with London) for within-city density effects. Other studies for Stockholm find intra-city density effects of around 0.01-0.02 (Andersson et al 2016). Some studies estimate bigger implicit effects from within-city density, usually when looking at house prices or rents. Liu Rosenthal and Strange (2017) find office rents respond to zip code employment with an elasticity of around 0.10 in US Metropolitan Statistical Areas. Dericks and Koster (2018) get even bigger effects on office rents from density in London (estimated from the long run effects of World War II bombings on density), with an elasticity of around 0.3. However they calculate that this translates into a productivity elasticity of around 0.07 (their assumptions would imply that the rent elasticity of Liu, Rosenthal and Strange 2017 also translates into a lower density-productivity elasticity of only 0.025). Ahlfeldt at al (2015) come up with an agglomeration elasticity for density within Berlin of around 0.07, using a structural model of the urban economy.

For this study, we also carried our own analysis using aggregate data on wages and employment in Britain at Parliamentary Constituency level, controlling for employment at Travel to Work Area Level. This analysis suggests figures of 0.025-0.03 for city-level agglomeration, and 0.02-0.025 for within city density (the higher figures corresponding to the case when London is included in the sample). Unfortunately this is the only estimate available for both size effects and density effects simultaneously, as this comparison is not presented in any other existing work. This lack of evidence means we cannot be confident on which values are appropriate when we want to model both size and density effects, because the size elasticities in the literature could be partly attributable to density effects within cities. We therefore experiment with values within the 0.01-0.05 range, which covers most of the estimates available.

5. Results

5.1. Illustrative results for simulated city population distributions

Table 1 shows the summary statistics for the 79 largest TTWAs forming the sample for the analysis. Each TTWA is divided into inner and outer areas so there are 158 units of analysis in total. The table shows the mean, standard deviation, minimum and maximum for the 158 TTWA sub areas, the 79 TTWAs, and for the total output from all these TTWAs as a whole (i.e. the implied economic output of the primary UK cities in total). The table is subdivided into three parts a, b and c.

5.1.1.Redistributing workers across TTWAs – no density effects

To start, Table 1a shows the distributions and totals with the density elasticity set to zero, and the size elasticity set to a high value of 0.05. The inner and outer area and worker shares are arbitrarily set to 0.5 (i.e. equal size areas with equal numbers of workers), but here the distribution of workers within TTWAs is irrelevant because density effects are switched off. Density is discussed later in Table 1b and 1c. The top panel shows the baseline predicted output, with workers distributed across TTWAs as they were in 2011. The values are normalised such that total output is 100 in this baseline. Each TTWA sub area contributes 0.633% to total output on average (158 x 0.633 = 100) and each TTWA contributes 1.266% to total output (79 x 1.266 = 100). The range of outputs across TTWAs is wide, from 0.4% of total output to 20.6%. Unsurprisingly, the highest figures relate to the London area. This predicted share is close to the actual current figure – the Greater London area contributed 22.7% to UK gross value added in 2016 (ONS regional GVA figures).

The lower panels show what happens to output when we redistribute workers across these 79 TTWAs. To reiterate, these numbers related to predicted output under the assumed functional form and scale of agglomeration economies and take no account any other welfare benefits or costs. When we allocate all workers to one TTWA in the next panel, total output increases in total by around 21% and is self-evidently confined to one TTWA and output is zero in TTWAs with no workers. For individual workers, this would imply an equitable distribution, since everyone lives in the same city so no one is disadvantaged by where they live. Obviously this is an infeasible practical example, but illustrates the basic point that concentration of economic activity implies aggregate gains in output under the assumption of city level economies of scale.

The next panel redistributes workers equally across TTWAs, again an infeasible but illustrative example. By construction, all TTWAs have equal output, but total output falls by 2.6% relative to the baseline. Here there is an equity-efficiency trade-off due to the increasing returns to city size. The final panel redistributes workers according to a log-normal distribution, basically reducing the role of London's agglomeration in national output and bringing the city size distribution more in line with the US. This is approximately the distribution that characterises city sizes in other developed countries and has been proposed as a potential target when the aim is to spatially rebalance the UK economy by boosting its second-tier cities (Overman and Rice 2008). This distribution reduces the size of the largest TTWA relative to the others and implies a more equal distribution across TTWAs. The standard deviation across TTWAs is 1.1% compared to 3.2% in the baseline and the maximum TTWA output (London) is 8.7% compared to 21%. Because the distribution is generated randomly, the exact figures will be dependent on the particular distribution generated in any instance. Predicted total output falls, by 1.4% relative to the baseline, again illustrating spatial equity-efficiency trade off inherent in a world characterised by agglomeration economies.

Figure 1a illustrates the difference between the current distribution and the lognormal distribution across TTWAs in this example.

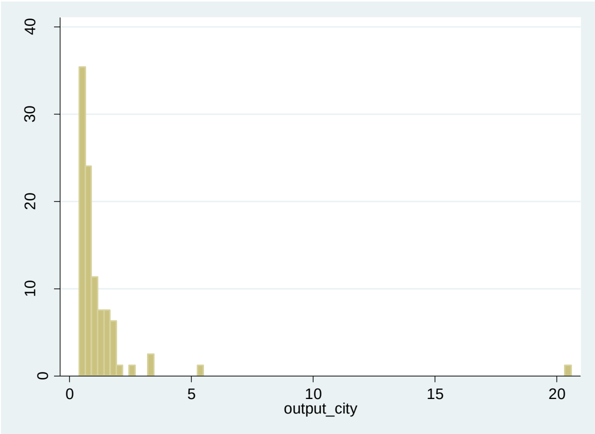
Table 1a: Illustrative simulations of distributions across TTWAs, descriptive statistics:
size=0.05, density=0.00, inner workers =0.5, inner area =0.5

Actual TTWA distribution	mean	sd	min	max
TTWA-sub area output	0.633	1.163	0.196	10.293
TTWA output	1.266	2.327	0.391	20.587
Total output (79 TTWAs)	100	0	100	100
One city				
TTWA-sub area output	0.767	6.795	0	60.591
TTWA output	1.534	13.59	0	121.181
Total output (79 TTWAs)	121.181	0	121.181	121.181
Equal sizes				
TTWA-sub area output	0.616	0	0.616	0.616
TTWA output	1.233	0	1.233	1.233
Total output (79 TTWAs)	97.399	0	97.399	97.399
Log normal				
TTWA-sub area output	0.624	0.559	0.109	4.343
TTWA output	1.248	1.117	0.217	8.687
Total output (79 TTWAs)	98.616	0	98.616	98.616

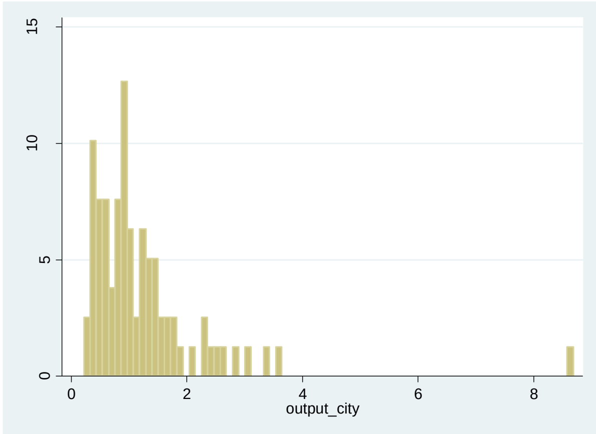
Output is normalised such that predicted total output under the actual between-TTWA distribution with a uniform distribution of workers within the city is 100.

Figure 1a: Simulated distribution of output across TTWAs under current and log-normal scenarios

Current, assuming b=0.05, c=0.0



Log normal assuming b=0.05, c=0.0



Output is normalised such that predicted total output under the actual between-TTWA distribution with a uniform distribution of workers within the city is 100. Note log normal distributions are randomly generated, so the exact distribution should not be compared across cases.

The numbers in Table 1a will depend on the value of the parameter that characterises the elasticity of productivity with respect to city size. Table 1b shows what happens to total output when we reduce the elasticity to 0.03 and to 0.01. As expected, the potential gains from concentration are highly dependent on how productivity responds to city size. With an elasticity of 0.01 the aggregate output gain from concentrating activity in one city is only 4% (compared to 21% when the elasticity was 0.05) and the output cost from equal size cities is only 0.5% (compared to 2.6% when the elasticity was 0.05). Clearly, evidence on the magnitude of these elasticities is important for understanding the implications of policy to promote or impede the growth of cities in order to achieve efficiency gains.

Table 1b: Illustrative total output from TTWAs for different distributions and size elasticities, descriptive statistics

	size = 0.05	size = 0.03	size =0.01
Actual TTWA distribution	100	100	100
Once city	121.181	112.267	103.947
Equal sizes	97.399	98.474	99.503
Log normal	98.616	99.083	99.687

Output is normalised such that predicted total output under the actual between-TTWA distribution with a uniform distribution of workers within the city is 100.

5.1.2.Redistributing workers within TTWAs

In this section, we extend the results of section 5.1.1 by redistributing workers within TTWAs, and switching on the density elasticities, so that there are gains from concentrating workers in particular places within cities (notionally, the central business district although the geographical location is irrelevant in these simulations). The format and interpretation of the tables is identical to Table 1a.

Many combinations of the various parameters and distributions are possible, so we present just a few examples based on the current distribution of workers across TTWAs and with the size elasticity set to a mid-range value of 0.03. Table 1c starts with inner TTWA land area set to 0.3% of the total, and the inner worker share set to 0.15 which approximates the current situation (based on the NIC/LIDA data used in section 5.2) and with an assumed (high) density elasticity of 0.05. The next panel reduces the elasticity of productivity with respect to density down to 0.01. The rows below show the case when we increase the share of workers in the inner area to 30%, for the same two density elasticities (i.e. doubling the share in the central city). As in Tables 1a and 1b, output is normalised such that predicted total output under the actual between-TTWA distribution with a uniform distribution of workers within the city (or with the density elasticity equal to zero) is 100.

The key take-out from this table is that the gains from increasing density are potentially quite large. Doubling central city density, when the elasticity is 0.05 towards the upper end of estimates in the literature, implies gains of 3.9% relative to the case where the inner-outer shares are roughly what they are now (106.541/102.565 = 1.039). This comes without changing relative between-TTWA inequality: all TTWA outputs are simply scaled up by 3.9%. The gains from concentration within cities are much more muted when the elasticity is at the lower end of the range at 0.01: around 0.7%. Clearly, given the relatively limited empirical evidence on the effects of changes in within city density, obtaining more and better estimates of this elasticity is an important direction for future work.

In the next section we go on to look at the effects of changing inner city workforce density, based on a more realistic range of future scenarios.

Table 1c: Illustrative simulations, descriptive statistics:
Actual between TTWA distribution 2011, size elasticity = 0.03, inner areas share = 0.3%

Inner share 0.15, density = 0.05	mean	sd	min	max
TTWA-sub area output	0.649	1.525	0.07	17.939
TTWA output	1.298	2.466	0.393	21.819
Total output (79 TTWAs)	102.565	0	102.565	102.565
Inner share 0.15, density = 0.01				
TTWA-sub area output	0.636	1.45	0.063	17
TTWA output	1.272	2.275	0.403	20.124
Total output (79 TTWAs)	100.463	0	100.463	100.463
Inner share 0.30, density = 0.05				
TTWA-sub area output	0.674	1.348	0.145	14.63
TTWA output	1.349	2.561	0.408	22.665
Total output (79 TTWAs)	106.541	0	106.541	106.541
Inner share 0.30, density = 0.01				
TTWA-sub area output	0.64	1.249	0.126	13.972
TTWA output	1.281	2.291	0.406	20.265
Total output (79 TTWAs)	101.167	0	101.167	101.167

Output is normalised such that predicted total output under the actual between-TTWA distribution with a uniform distribution of workers within the city is 100.

5.2. Results for NIC scenarios for Primary Urban Areas

The following table shows some headline examples of relative UK total output implied by the various NIC-provided future population distribution scenarios as discussed in Section 4.1. The population and transport scenarios are labelled as *Business-as-usual constrained*, *Business-as-usual more constrained*, *High-urbanisation constrained*, *High-urbanisation more constrained.* In each case, estimates are presented for various assumptions about the agglomeration and elasticity parameters that determine output. The analysis relates to working population data aggregated to Primary Urban Areas, divided into inner core areas and outer areas (as defined by NIC and ITA Leeds). Output is presented as relative to the Business-as-usual and High Urbanisation baselines and as a GDP loss figure, derived by multiplying the relative output loss by the projected UK GDP figures (in 2018 prices, based on OECD projections).

Table 2a: Illustrative predicted UK outputs from various city transport constraint scenarios

	Relative to baseline			Estimated GDP loss (£million 2018 prices)		
	2030	2040	2050	2030	2040	2050
Business-as-usual constrained						
Size = 0.01, Dens = 0.01	99.999	99.998	99.998	26	57	82
Size = 0.03, Dens = 0.03	99.996	99.993	99.991	118	247	349
Size = 0.05, Dens = 0.05	99.990	99.984	99.981	263	539	755
Size = 0.01, Dens = 0.05	99.989	99.983	99.980	295	584	805
Size = 0.05, Dens = 0.01	100.003	100.003	100.003	-68	-105	-126
Business-as-usual more constrained						
Size = 0.01, Dens = 0.01	99.998	99.997	99.996	48	103	149
Size = 0.03, Dens = 0.03	99.993	99.988	99.985	195	411	588
Size = 0.05, Dens = 0.05	99.985	99.975	99.970	411	850	1,209
Size = 0.01, Dens = 0.05	99.984	99.975	99.970	424	848	1,192
Size = 0.05, Dens = 0.01	100.002	100.001	100.001	-42	-43	-37
High-urbanisation constrained						
Size = 0.01, Dens = 0.01	99.999	99.998	99.997	33	81	138
Size = 0.03, Dens = 0.03	99.994	99.990	99.986	149	349	574
Size = 0.05, Dens = 0.05	99.988	99.977	99.970	332	755	1,215
Size = 0.01, Dens = 0.05	99.986	99.976	99.969	374	812	1,267
Size = 0.05, Dens = 0.01	100.003	100.004	100.004	-86	-133	-144
High-urbanisation more constrained						
Size = 0.01, Dens = 0.01	99.998	99.995	99.993	62	151	273
Size = 0.03, Dens = 0.03	99.991	99.982	99.974	253	597	1,037
Size = 0.05, Dens = 0.05	99.980	99.963	99.948	532	1,227	2,077
Size = 0.01, Dens = 0.05	99.979	99.964	99.951	547	1,214	1,978
Size = 0.05, Dens = 0.01	100.002	100.001	99.998	-53	-39	70

The stand-out result from Table 2a, is that the effects of these transport constraints on output – working through the effects on these on the population distribution, and in turn city size effects and density effects on productivity – are not large when measured in terms of % of total GDP. The worst case scenario is one in which size and density elasticities are high (0.05), population becomes more urbanised than in the ONS projections, but within city transport and central capacity is heavily constrained so that the number of inner city workers is lower than what would currently be projected. This yields UK output that is 0.052% lower in 2050 than if inner city employment capacity evolves in line with past trends (column 3, *High-urbanisation more constrained*, *Size = 0.05, Dens*

= 0.05). However, this does imply reasonably large monetary costs given this is a percentage of total UK GDP, implying a cost to the economy of £2 billion each year (or £57 billion present value over an infinite time horizon, with a discount rate of 3.5%).

The results are evidently more sensitive to the assumed density elasticity than they are to the size elasticity. This is unsurprising, given the main difference between the *constrained* and *more constrained* scenarios is to shift working population between the PUA inner outer areas. This can be seen by comparing the cases *Size = 0.05, Dens = 0.05* and *Size = 0.01, Dens = 0.05* in Table 2: the total UK output is only marginally different in each case. We can also switch off the size elasticity completely and change the density elasticity. Taking the *Business-as-usual more constrained* case as an example, the 2050 UK output assuming density elasticities of 0.01, 0.03 and 0.05 are 99.994, 99.983 and 99.970 respectively. These are only marginally different from the corresponding figures in Table 2, column 3 when the size elasticity is changed to the same value as the density elasticity (99.996, 99.985, 99.970 respectively).

Note, that the constrained scenarios produce slightly higher predicted GDP than the unconstrained baseline in the case when the size elasticity is set high and the density elasticity set low (as in the *Size = 0.05, Dens = 0.01* cases in Table 2*)*. The reasons for this is that the assumed constraints redistribute working populations amongst PUAs and from the residual area to England PUAs, which turns out to lead to small output improvement relative to the baseline business as usual scenario. Not much should be read into this finding since it is in part an artefact of having a large residual group in which the size and density effects are set to zero.

As well as affecting total UK output, these city workforce scenarios affect the distribution of output between Primary Urban Areas, and between the central cores and outer areas of these urban areas. In practice, these differences in the distribution are too small to visualise graphically through histograms. The figures are also difficult to interpret in terms of magnitudes when comparing across predictions based on different agglomeration parameter assumptions, because these assumptions change the share of output produced by England's Primary Urban Areas, relative to the residual (non-urban England, Scotland, Wales and NI).

We therefore present just one example, comparing the *Business-as-usual, Business-as-usual constrained*, and *Business-as-usual more constrained* scenarios, for 2050, with the size and density elasticities set to a mid-range plausible value (0.03). These results are shown in Table 2b. As in Table 2a, the output figures are normalised relative to total output in the *Business-as-usual* case (i.e. unmodified ONS projections). Working population density figures (per hectare) are given by the various working population scenarios and are shown for the Primary Urban Area overall, and for its sub-areas (inner core and outer areas).

To understand these figures, consider an example. From the top row, in the Business-as-usual case based on the ONS projections, the Primary Urban Areas sub-areas in England will each contribute 0.61% on average to total output of the UK in 2050. Each Primary Urban Area contributes 1.22% to total output, since there are two sub-areas (inner and outer) in each Primary Urban Area. Note that the contribution of PUAs as whole to UK output is 110 x 0.61 = 67% (or 55 x 1.22). The remaining 29.9% is attributable to the residual group, where all size and density effects are set to zero. A point of comparison with actual data here is that Scotland, Wales and Northern

Ireland as a whole (our residual group, but excluding England's non-PUA areas) contributed about 15% to UK GVA in 2016 (ONS regional GVA figures).

The range of predicted sub-area outputs is wide under this baseline scenario, ranging from 0.01% to 17.53% of total UK output, or 0.19% to 28.35% of total output for the PUA as a whole. As we would expect, the highest figures for the sub-PUA and PUAs relate to inner London (Westminster and the City) and London as a whole. These predicted shares are higher than the actual current figures – the Greater London area contributed 22.7% to UK gross value added in 2016 (ONS regional GVA figures). There are three interrelated reasons for this discrepancy: a) the figures in Table 3 are based on future population projections; b) the size and density elasticities have been assumed based on past literature and have not been calibrated to the current distribution of output across PUAs and sub-areas; c) the assumption of zero-density and size effects outside England PUAs could mean that the relative contribution of the this residual group is understated.

Imposing constraints on the evolution of central city employment makes relatively little difference to these distributions. The standard deviation of output across BUAs is reduced by only 0.001% when moving from the *Business-as-usual* to *Business-as-usual more constrained* scenarios.

Table 2b: Illustrative predicted output distribution from various city transport constraint scenarios. England Primary Urban Areas

	Mean	*Standard Deviation*	*Minimum*	*Maximum*
Business-as-usual				
PUA sub area output	0.609	1.994	0.009	17.529
PUA sub-area density	96.16	133.45	2.04	1066.48
PUA output	1.218	3.800	0.191	28.356
PUA density	11.54	8.00	2.27	30.42
UK total output	100	0	100	100
Business-as-usual constrained				
PUA sub area output	0.609	1.995	0.009	17.515
PUA sub-area density	95.32	132.12	2.035	1065.61
PUA output	1.218	3.800	0.191	28.332
PUA density	11.54	7.99	2.27	30.40
UK total output	99.991	0	99.991	99.991
Business-as-usual more constrained				
PUA sub area output	0.609	1.994	0.009	17.511
PUA sub-area density	93.76	130.10	2.03	1065.30
PUA output	1.218	3.796	0.191	28.325
PUA density	11.55	8.00	2.27	30.40
UK total output	99.985	0	99.985	99.985

Analysis relates to 110 units: 55 Primary Urban Areas split into two sub units inner and outer. UK total output based on these 110 units plus residual group. Values normalised relative to UK total output in baseline *Business-as-usual scenario*.

Evidently, the impact of redistributing workers in these examples – based on actual population projections, and credible within-city changes in employment distribution – is much smaller than in the extreme benchmark cases discussed in the simulations in Section 5.1. So density and agglomeration effects clearly matter, but the redistribution of workers has to be quite large for these effects to really have a dramatic impact on aggregate output. From a cost benefit perspective, infrastructure improvements to avoid constraints such as those in Table 2 may be worthwhile. From the perspective of making strategic decisions about the national economy – either in terms of aggregate output or geographic 'rebalancing', they seem less important.

6. Conclusions

This report provides some simulations of the effects on productivity and economic output achievable by policy that redistributes workers within and between cities within the UK. The analysis is based on theories and empirical evidence on economies of city size and density that are prevalent in the urban economics and transport economics literature. The results are generated by simulating various city size and density distributions, and then predicting output at the sub-city, city and UK level.

Qualitatively, the results follow mechanically from the assumptions underlying the modelling exercise. City total output exhibits increasing returns to city size, because individual worker productivity increases with city size and there are more workers in bigger cities. This assumption is in line with a large body of empirical evidence. The main contribution is to demonstrate quantitatively how big the potential output gains are, from concentrating workers in some cities and from concentrating workers in central areas of cities in a UK context. We do this using a range of scenarios in which workers are redistributed between cities, and between the inner and outer areas of cities. The results say nothing about welfare or wellbeing or the feasibility of achieving these distributions. We have not considered anything on the costs side (infrastructure costs, congestion, crime and other negative externalities), nor anything about benefits other than economic output (consumption amenities, economies of scale in provision of energy infrastructure, positive externalities).

We first looked at some ad-hoc simulated distributions of workers across the biggest 79 Travel to Work Areas in Britain. The extreme but totally infeasible scenario of concentrating this urban population in one city produces large gains in output (+21%). This option of course also achieves complete equality amongst urban workers because everyone lives in the same city so no one is disadvantaged in terms of productivity by where they live. An alternative scenario which equalizes the sizes of thee 79 cities implies a loss in total output or around 2.6%. Bringing the distribution of city sizes closer to that found in many other developed economies (i.e. making London less dominant) implies greater equality in worker and city productivity, but an aggregate loss of around 1.4% relative to the status quo. Simulations involving moving employment towards the inner core of cities, to exploit economies of density, also generates substantial gains, between 0.7% and 3.8% depending on the elasticity assumed from the range of estimates available. More empirical evidence is needed on these elasticities, given the scarcity of estimates in the literature.

We also looked at range of scenarios using data supplied by NIC, based on ONS workforce projections up to 2050, aggregate to Primary Urban Areas (PUA), divided into inner cores and outer areas. Alternative scenarios reduced the inner PUA workforce relative to the outer area workforce, assuming constraints on capacity for workers to commute to the inner city areas. The effects of these constraints are small scaled relative to GDP in total implying a loss of 0.05% of GDP or less relative to the 'business as usual' baseline. Expressed as monetary quantities, these costs appear more substantial, up to £2 billion per year.

References

Andersson, Martin, Johan Klaesson and Johan P. Larsson How Local are Spatial Density Externalities? Neighbourhood Effects in Agglomeration Economies, Regional Studies, Vol. 50, No. 6, 1082–1095

Cheshire, Paul, Max Nathan, Henry. G. Overman (2014) Urban Economics and Urban Policy, Edward Elgar

Combes, Pierre.-P. and Laurent. Gobillon (2015). The empirics of agglomeration economies. In J. V. H. Gilles Duranton and W. C. Strange (Eds.), Handbook of Regional and Urban Economics, Volume 5 of Handbook of Regional and Urban Economics, pp. 247 -348. Elsevier.

Dericks, Gerard. H. and Hans R. A. Koster (2018) The Billion Pound Drop: The Blitz and Agglomeration Economies in London, Centre for Economic Performance Discussion Paper CEPDP1542, London School of Economics

Duranton, Gilles. and Puga Diego. (2004). "Micro-foundations of Urban Agglomeration Economies" in J. F. Thisse and J. V. Henderson (eds.) Handbook of Urban and Regional Economics, Volume 4. North Holland.

Glaeser , Edward L. Jed Kolko Albert Saiz (2001) Consumer city, Journal of Economic Geography, Volume 1, Issue 1, 1 January 2001, Pages 27–50

Glaeser, Edward L. and Gottlieb, Joshua D. (2008) The Economics of Place-Making Policies, Brookings Papers on Economic Activity, Spring, pages 155-253

Graham, D.J and Stephen Gibbons (2018), Quantifying Wider Economic Impacts of Agglomeration for Transport Appraisal: Existing Evidence and Future Directions, Report for Transport for London

Henderson, J.Vernon. (1988) Urban Development, Oxford University Press, New York

Liu, Crocker H., Stuart S. Rosenthal, William C. Strange (2017) The Vertical City: Rent Gradients, Spatial Structure, and Agglomeration Economies, unpublished manuscript.

Moretti, Enrico. (2004) Human capital externalities in cities. In J. V. Henderson, J.F. Thisse (eds.), Handbook of Regional and Urban Economics. Amsterdam: Elsevier.

Overman, Henry G. and Patricia Rice (2008) Resurgent Cities and Regional Economic Performance, Spatial Economics Research Centre Policy Paper 1, London School of Economics

SERC (2009) Strengthening Economic Linkages between Leeds and Manchester: Feasibility and Implications, Report for the Northern Way

Venables, Anthony. J., James. Laird, and Henry G. Overman (2014). Transport investment and economic performance: Implications for project appraisal. London, DfT.

Appendix A: Literature on agglomeration elasticities and within-city density effects on productivity

International empirical evidence on urban agglomeration economies

One of the most globally pervasive empirical regularities of the spatial economy is the existence of a positive relationship between city size and productivity. On average, firms and workers in cities are economically more productive than those in rural areas, and firms and workers in big cities are more productive than those in smaller ones. These productive benefits of concentrating activity in large places are known as 'agglomeration economies' and are thought to occur through a number of channels related to the sharing of infrastructure, markets and information (sharing, matching and learning in the theoretical terminology of Duranton and Puga 2004). Large cities can also generate benefits for consumers in terms of more product market competition and provision of amenities (Glaeser at al 2001).

Estimates of the productivity benefits of agglomeration are typically obtained from the statistical association of measures of productivity with a measure of urban size or density. The estimates from these studies are known as urbanisation or agglomeration 'elasticities'. Spatial variance in productivity is usually represented via total factor productivity (TFP) within a production function framework or labour productivity via a model for wages, assuming that workers are paid the value of their marginal product (for details see Combes and Gobillon 2015. The majority of studies look at the effects of city size as measured by the population or employment of urban areas. A smaller number investigate the effects of density (following from Ciccone and Hall 1996) including Cheshire and Magini 200#. Others measure the degree of access to economic mass, such as an effective density of the form

$$\rho_i = \frac{1}{n}\sum_{j=1}^{n} m_j f(d_{ij}),$$

where for n spatial units (zones) indexed by i, i = (1, ...n), or j, j = (1, ...n), m_j is a measure of economic mass at zone j and f(d_{ij}), often referred to as the impedance function, is a decreasing function of the cost of travelling from origin i to destination j. Denoting productivity for each industrial sector s, s = (1, ...S), by ω_s, an econometric model for

$$\omega_s = f(\rho_i, Z_{si})$$

is specified, where Z_{si} represents other relevant effects on productivity. From this model elasticities of productivity with respect to urban agglomeration for each defined sector of the economy, i.e.

$$\delta_s = \frac{\partial \log \omega_s}{\partial \log \rho},$$

are estimated.

The standard approach to estimation is to use a linear regression of firm output or wages, on measure of city size or density.

Very few studies consider both density and size together, or the effects of the density of city centres for given city size. This distinction is theoretically important since cities of a given population can have different densities, and the effects of some sources of agglomeration such as inter-firm interaction, may depend on localised business concentration rather than the overall city size. On the other hand some sources of agglomeration will be more dispersed, such access to a large shared pool of labour. We discuss the more limited evidence on desnity within cities in the next section.

Table A1 below reports results from 47 international empirical studies that have estimated over 1000 urban agglomeration elasticities using TFP or wage models to represent productivity.

Table A1: International estimates of urban agglomeration elasticities

Study	*country*	*period*	*data*	*aggregation*	*Obs.*	*Mean*
Aberg (1973)	Sweden	1965-68	CS	reg	4	0.017
Au and Henderson (2006)	China	1997	CS	reg	2	0.013
Baldwin et al (2007)	Canada	1999	CS	plant	8	0.061
Baldwin et al (2008)	Canada	1989-1999	PD	plant	6	-0.088
Brulhart and Mathys (2008)	Europe	1980-2003	PD	reg	14	-0.080
Ciccone (2002)	Europe	1992	CS	reg	7	0.047
Ciccone and Hall (1996)	US	1988	CS	reg	8	0.053
Cingano and Shivardi (2004)	Italy	1992	CS	reg	13	0.054
Combes et al (2010)	France	1988	PD	worker	43	0.035
Combes et al (2008)	France	1988	PD	zone	11	0.052
Davis and Weinstein (2001)	Japan	1985	CS	reg	11	0.027
Di Addario and Patacchini (2008)	Italy	1995-2002	PD	worker	1	0.010
Fingleton (2003)	UK/GB	1999-2000	CS	reg	3	0.017
Fingleton (2006)	UK/GB	2000	CS	reg	7	0.025
Graham (2000)	UK/GB	1984	CS	reg	22	-0.006
Graham (2005)	UK/GB	1998-2002	PD	firm	36	0.193
Graham (2007a)	UK/GB	1995-2004	PD	firm	28	0.110
Graham (2007b)	UK/GB	1995-2004	PD	firm	18	0.194
Graham (2009)	UK/GB	1995-2004	PD	firm	108	0.097
Graham and Kim (2008)	UK/GB	1995-2004	PD	firm	18	0.079
Graham et al (2009)	UK/GB	2000-2006	PD	plant	5	0.041
Henderson (1986)	Brazil	1970-72	CS	reg	52	0.010
Henderson (2003)	US	1982	PD	firm	4	0.024
Hensher et al (2012)	Australia	2006	CS	zone	39	0.071
Kanemoto et al (1996)	Japan	1985	CS	reg	9	0.089
Lall et al (2004)	India	1991	CS	plant	18	0.017
Maré and Graham (2010)	NZ	1999-2007	PD	plant	114	0.043
Melo and Graham (2009)	UK/GB	2002-2006	PD	worker	64	0.029
Mion and Naticchioni (2005)	Italy	1995	PD	worker	30	0.034
Moomaw (1981)	US	1967	CS	reg	18	0.060
Moomaw (1983a)	US	1977	CS	reg	26	0.038
Moomaw (1985)	US	1972	PD	reg	36	0.040
Nakamura (1985)	Japan	1979	CS	cities	38	0.026
Rice et al (2006)	UK/GB	1998-2000	CS	reg	14	0.026
Rosenthal and Strange (2008)	US	2000	CS	worker	9	0.042

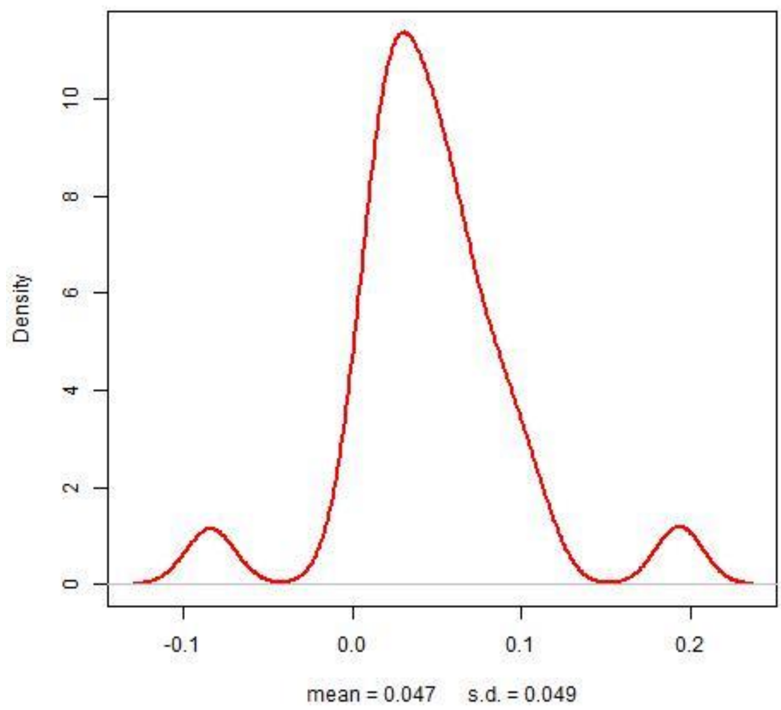
Sveikaukas et al (1988)	US	1977	CS	reg	6	0.013
Sveikauskas (1975)	US	1967	CS	reg	42	0.057
Tabuchi (1986)	Japan	1980	CS	reg	57	0.060
Wheeler (2001)	US	1980	CS	worker	3	0.017
Morikawa (2011)	Japan	2002-2005	PD	firm	4	0.110
Ahlfeldt et al. (2015)	Germany	1936-1986-2006	PD	reg	3	0.062
Combes et al (2012)	France	1994-2002	PD	plant	17	0.090
Marrocu (2013)	Europe	1996-2007	CS	reg	5	0.036
Holl (2011)	Spain	1991-2005	PD	firm	23	0.089
Mare (2016)	NZ	2001-2012	PD	plant	31	0.075
Martin et al. (2010)	France	1996-2004	PD	plant	8	0.011
Total					1043	0.046

Key: CS – crossectional, PD – panel data, TS – time series, reg – regional data, plant – plant data, firm data.

The table shows information about the country and data / methods used for estimation as well as the number of elasticity estimates collected from each study and the mean elasticity value. Estimates vary between -0.800 and 0.658, and have unweighted mean equal to 0.046 (and median equal to 0.043). The bootstrapped confidence interval for the mean of this sample, from 5000 replactions, is (0.033, 0.061) with standard error 0.007.

Figure 1 shows a histogram of the values presented in the table.

Figure 1: Histogram of urban agglomeration elasticities.



Combes and Gobillon (2015) provide a comprehensive review of the empirical literature on urban agglomeration. They argue that failure to adjust adequately for endogeneity, in the form of spatial selection and unobserved individual level heterogeneity, leads to larger urbanisation elasticity values in the range 0.04 to 0.07. Melo et al (2009), in their meta-analysis of urban agglomeration elasticities, also report such an effect. The likely magnitude of the agglomeration elasticity; net of sorting, skills and reverse causality effects; is, according to Combes and Gobillon (2015), around 0.02, or more generally, around half the size of the estimate from an unadjusted model. However, along with Melo et al (2009), they also note that there is considerable 'natural' variation across different empirical contexts due to contextual characteristics, including country specific effects, industrial coverage, the metric used to represent agglomeration, and the nature of controls for unobserved factor quality and spatial selection.

For Britain, the agglomeration elasticity values used for appraisal of transport schemes were estimated by Graham et al (2009). Estimation made use of extensive firm level panel data for four broad sectors of the economy: manufacturing, construction, consumer services and business services. The analysis was based on a production function model using a control function approach to address potential sources of endogeneity and to allow for unobserved firm level heterogeneity.

The panel data control function approach addresses endogeneity by proxing for unobserved productivity effects. It proceeds as follows: find a function to proxy for unobserved productivity and include it as an additional model component to obtain consistent parameter estimates. Use an AR(1) specification to draw on variation over time, rather than across firms, thus adjusting for unobserved time-invariant confounding.

The key empirical results of their research are summarised in the table below.

Table A2: Summary of UK agglomeration elasticities.

	sic	agglomeration elasticity
Manufacturing	15-40	0.024
		(0.002)
Construction	45	0.034
		(0.003)
Consumer services	50-64	0.024
		(0.003)
Business services	65-75	0.083
		(0.007)
Economy (weight aver.)	15-75	0.044

The table shows positive agglomerations effects for the four broad sectors under study, with substantially larger effects for business service industries.

Evidence on within-city density

Estimates of the effects that within city density have on productivity – holding overall city size constant – are less prevalent in the literature. The idea that density is important, at least for some industries is quite pervasive, and is the standard theoretical justification for the existence of high density central business districts in all major cities. Estimates are usually obtained using similar methods to those discussed above, using a regression of wages, firm output or commercial rents on a measure of local density. Overall city size is typically controlled for by including city level fixed effects (city specific dummy variables, so that estimation is based only on within-city variation in density and productivity), by directly controlling for city size, or in some cases by only considering one city in isolation. The existing estimates are discussed in the main body of this paper in Section 4.3.

Evidence on the costs of agglomeration

From theory we believe that the spatial economy is organised in relation to what can broadly be classed as centripetal and centrifugal forces (e.g. Henderson 1988, Fujita and Thisse 2002). Centripetal forces are mainy those of agglomeration, which lead to the concentration of economic activity; centrifugal forces refer to the corresponding 'dis-benefits' of concentration arising from higher factor prices, congestion and other costs. The co-existence of these forces means that we do not expect to find economic activity either organised into a single large concentration, or distributed randomly, but rather into a number of concentrations of various sizes that are geographically separated.

There is considerably less empirical evidence on the costs of agglomeration than there is on the economies. A recent useful paper is by Combes et al (2012) which estimates the costs of urban agglomeration using data on French cities. They estimate an elasticity of urban costs with respect to population of 0.033. They note that this estimate is slightly larger than, but close to, previously estimated elasticities of urban agglomeration economies for France by Combes et al (2010) (0.015 to 0.030), and argue that this indicates that cities operate near aggregate constant returns to scale. If true, then the fundamental trade-off between urban benefits and costs is unlikely to have much influence on city sizes, thus providing a possible explanation for why we find such diversity in city sizes.

Appendix References

Cheshire, Paul, Max Nathan, Henry. G. Overman (2014) Urban Economics and Urban Policy, Edward Elgar

Combes,Pierre-P, Duranton G.,Gobillon L.,and Roux S. (2010) Estimating agglomeration economies with history, geology, and worker effects. In Edward L. Glaeser (ed.) The Economics of Agglomeration. Cambridge (mass.): National Bureau of Economic Research, 15– 65.

Combes, Pierre.-P. and Laurent. Gobillon (2015). The empirics of agglomeration economies. In J. V. H. Gilles Duranton and W. C. Strange (Eds.), Handbook of Regional and Urban Economics, Volume 5 of Handbook of Regional and Urban Economics, pp. 247 -348. Elsevier.

Combes, Pierre-P and Duranton, G and Gobillon, L (2012) The Costs of Agglomeration: Land Prices in French Cities. IZA Discussion Paper No. 7027. Available at SSRN: https://ssrn.com/abstract=2186805

Duranton, Gilles. and Puga Diego. (2004). "Micro-foundations of Urban Agglomeration Economies" in J. F. Thisse and J. V. Henderson (eds.) Handbook of Urban and Regional Economics, Volume 4. North Holland.

Fujita, M. and Thisse J.F. (2002) Economics of agglomeration: Cities, industrial location, and regional growth. Cambridge University Press, Cambridge..

Glaeser , Edward L. Jed Kolko Albert Saiz (2001) Consumer city, Journal of Economic Geography, Volume 1, Issue 1, 1 January 2001, Pages 27–50

Glaeser, Edward L. and Gottlieb, Joshua D. (2008) The Economics of Place-Making Policies, Brookings Papers on Economic Activity, Spring, pages 155-253

Graham D. J., Gibbons S., Martin R. 2009. Transport investments and the distance decay of agglomeration benefits. Working paper, Imperial College of London.

Henderson, J.Vernon. (1988) Urban Development, Oxford University Press, New York

Melo, P., D. J. Graham, and R. B. Noland (2009). A meta-analysis of estimates of urban agglomeration economies. Regional Science and Urban Economics 39, 332{342.

Moretti, Enrico. (2004) Human capital externalities in cities. In J. V. Henderson, J.F. Thisse (eds.), Handbook of Regional and Urban Economics. Amsterdam: Elsevier.