Spatial Patterns of Development and the British Housing Market

Tim Leunig[*]
London School of Economics

Henry Overman[†]
London School of Economics

15th January 2008[‡]

[*] Department of Economic History, London School of Economics Houghton Street, London WC2A 2AE, United Kingdom. T.leunig@lse.ac.uk, www.leunig.net. Also affiliated with the Centre for Economic Performance at the London School of Economics.
[†] Department of Geography and Environment, London School of Economics, Houghton Street, London WC2A 2AE, United Kingdom, h.g.overman@lse.ac.uk, http://personal.lse.ac.uk/overman. Also affiliated with the Centre for Economic Policy Research and the Centre for Economic Performance at the London School of Economics.
[‡] We thank Jonathan Wadsworth for preparing the Labour Force Survey estimates. We also thank John Muellbauer, Pat Rice and participants at the OxREP Housing Markets seminar for comments.

I. INTRODUCTION

Britain needs to build many more houses. As Meen and Andrew show in this volume, if it does not, affordability will continue to worsen in the medium term. With the publication of the Barker Reviews into Housing and then Land Use Planning, and now of the UK Government Green Paper "Houses for the Future" (Barker, 2004, 2006; CLG, 2007) the proposal to build 3 million more homes by 2020 has become the subject of heated public debate (CPRE, 2004, 2007; Centre for Cities, 2007). Much of it has revolved around whether these new homes are needed and whether the current planning system will deliver them. This paper is concerned with a third crucial question: where should these new homes be built? Insofar as the debate has moved beyond "Not in my backyard", it has been framed largely in terms of "Brownfield" versus "Greenfield" sites, with some discussion of the optimal density of new developments, in terms of environmental and social sustainability. These are valid issues, but they are only part of a more fundamental question: where is the optimal location for these new houses in terms of maximising living standards?

We know that living standards vary dramatically across Britain: some areas are affluent, others are depressed. Economists use "Gross Value Added" (GVA) to capture this variation.[1] In 2004 GVA per person (adjusted for commuting) in Inner London and Berkshire, Buckinghamshire and Oxfordshire was over £23,000, about 40% above the UK average. In contrast Cornwall, West Wales and the Valleys and the area around Durham each had GVA per person of less than £13,000, 24% below the UK average. This variation cannot be explained by inherent differences in physical geography such as climate or natural resources. Instead, the way the economy functions determines what is produced where, and by whom, and it is the resulting differences that determine geographic variations in living standards.

The strand of economics that explains these processes is known as spatial or urban economics. Research in this area provides theoretical justification and empirical evidence for two propositions. First: in terms of living standards there exists an optimal city size which changes over time.[2] Second, again in terms of living standards there is an optimal location for cities which, once more, changes over time. In practice, if our cities are larger or smaller than the optimum, or in the wrong place, productivity, wages and employment will be lower than they could be.

Houses last a long time and they are expensive to demolish and rebuild in another location. It is correspondingly important to build them in the right place initially. The Government's commitment to 3 million new homes by 2020 will add more than 10% to the housing stock and thus gives us a significant opportunity to "relocate Britain". We can locate those new houses to make city sizes closer to the economically optimal, and locations closer to the optimal locations for the modern economy. Building those houses will lead to people being better housed wherever we build them. But building them in line with insights from spatial economics means that not only will people be better housed, but wages, employment and national income will rise.

[1] Regional GVA is essentially the sub-national equivalent of GDP. It measures the output produced in a region. When corrected for commuting, it comes closer to national income: the amount of output produced using the factors of production located in a region.

[2] We use the term "city" to refer to all urban places.

We argue that there are two major changes in the economy that need to be recognised to make the most of this opportunity. First, on average, cities' optimal sizes continue to grow, as has happened for centuries. Second, the optimal location for cities has moved south over the last century, particularly towards London. These changes have been driven by technology, the rise of the service sector, and globalisation.

II. A SIMPLE MODEL OF THE SPATIAL ECONOMY

To consider the question of where to locate these houses in terms of raising living standards this paper sets out a basic model of spatial economics, using the graphical framework developed in Combes et al (2005).

Urban economics' basic intuition is simple: concentrating economic activity in cities leads to "agglomeration economies". Just like economies of scale at firm level, agglomeration economies mean that larger cities have higher productivity. Higher productivity, in turn, attracts more firms and workers, further increasing the city's size. In short, spatial concentration of economic activity generates self-reinforcing benefits. (Duranton and Puga 2004 surveys this literature).

Marshall (1890) described three possible sources of agglomeration economies: knowledge or human capital spillovers, labour market interactions and input-output linkages between firms. Knowledge or human capital spillovers arise because spatially concentrated firms or workers find it easier to learn from one another. Larger labour markets may, for example, allow a finer division of labour. Finally input-output linkages occur because lower transport costs between buyers and sellers means firms benefit from locating close to suppliers and customers. Economists have formalised these different sources and their implications (Fujita et al, 1999, Fujita and Thisse, 2002).

Figure 1

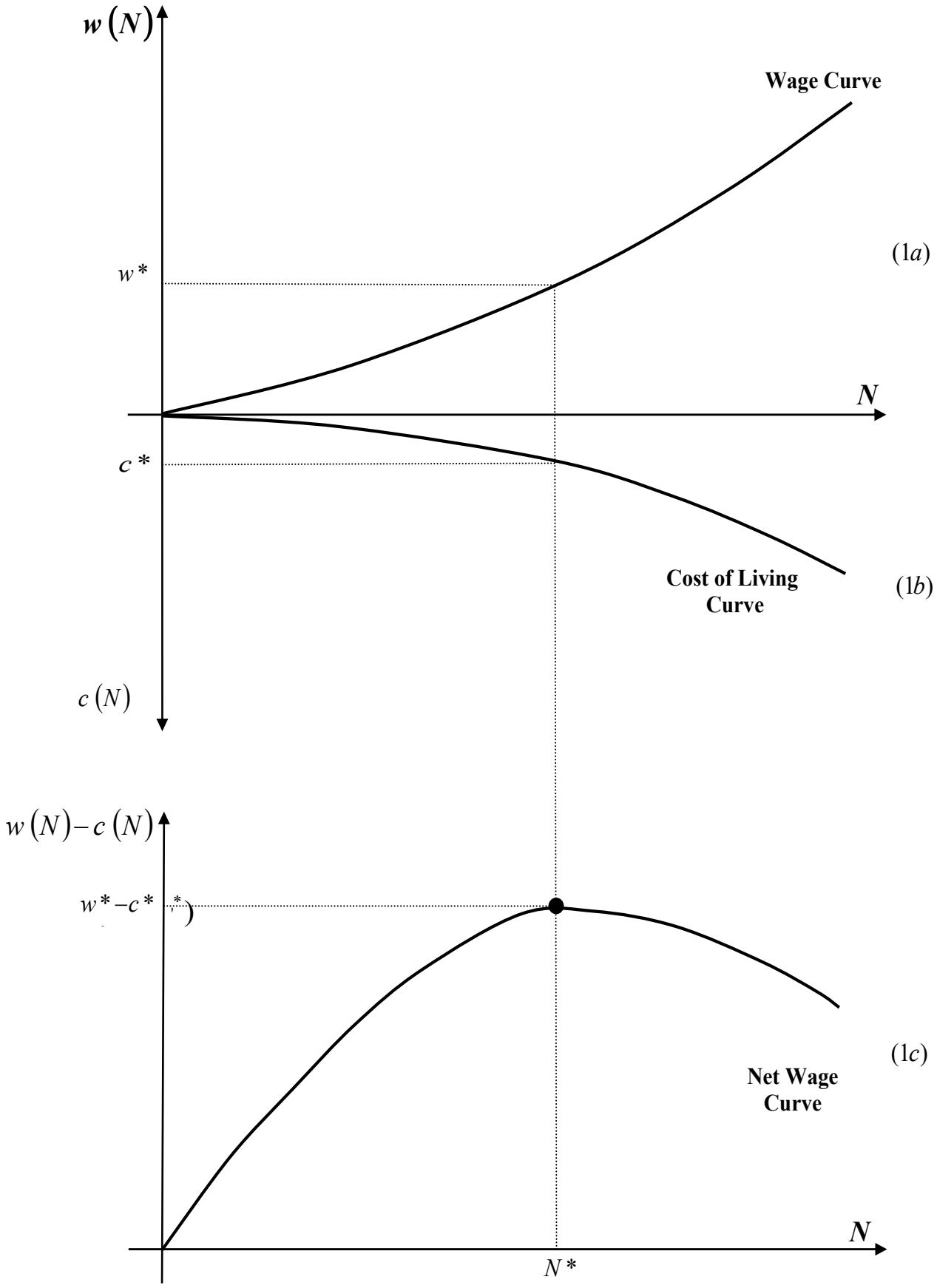

w(N)
Wage Curve
(1a)
w*
N
c*
Cost of Living Curve
(1b)
c(N)
w(N)−c(N)
w*−c*
(1c)
Net Wage Curve
N
N*

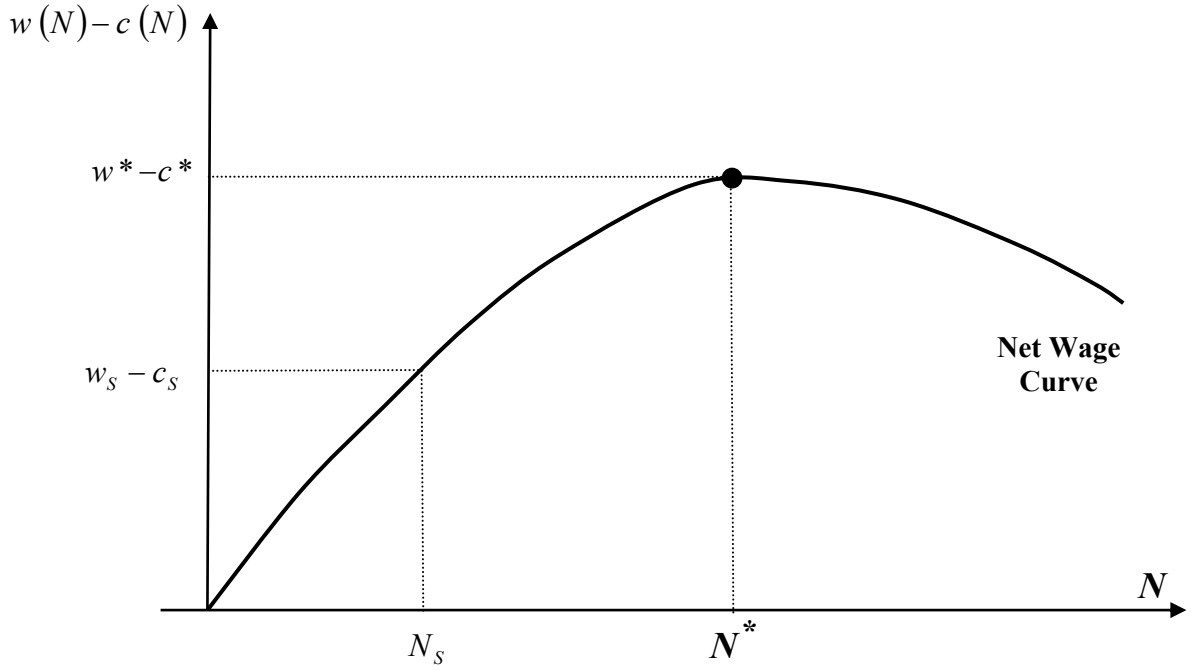
Figure 1

For these reasons, urban workers' marginal revenue product is higher when the city is larger, so that firms in larger cities can pay higher wages. This relationship between wages and city size is captured in figure 1a, taken from Combes et al (2005), which shows the wage (*w*) that firms can pay increasing as the city's population (*N*) increases. Offsetting this, the cost of living increases as the city grows: in particular, ceteris paribus, housing costs are always higher in large cities than small cities – the inevitable outcome of competition for scarce urban land. In addition, congestion means that the time and money spent travelling to work are generally higher in large cities. From the worker's point of view the cost of living is a negative item: workers want low costs of living. As such we draw it in figure 1b as a downward sloping line below the axis: as population (*N*) increases, living costs (c) increase, that is, the negative item becomes larger.

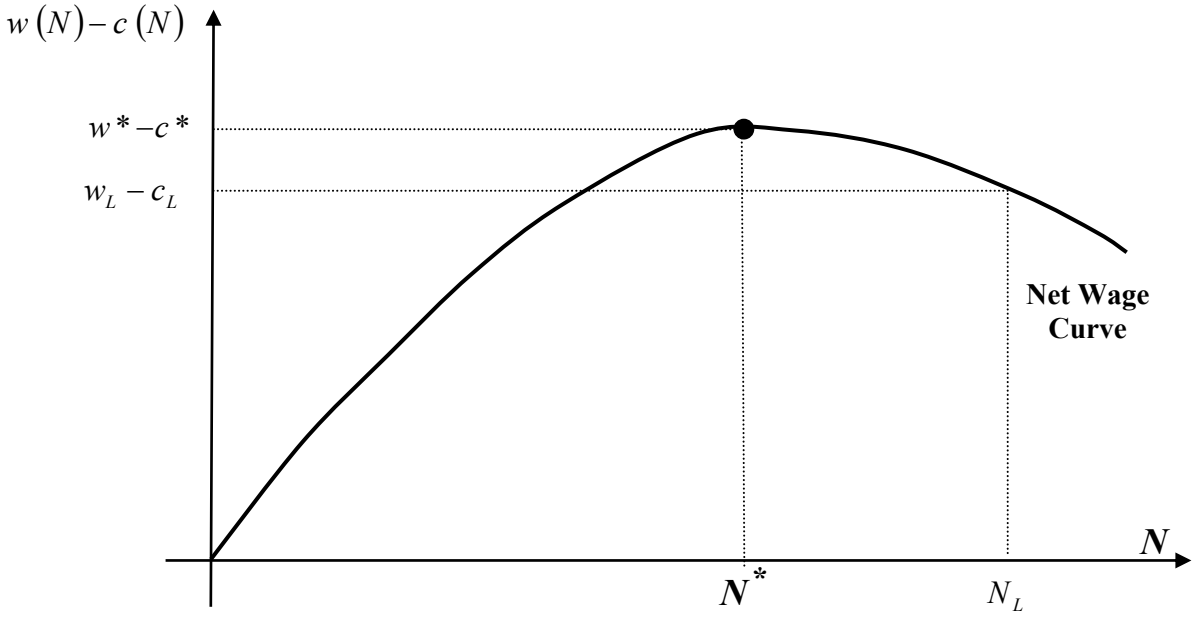
Workers' overall living standards are given by the wage minus the cost of living. In panel 1c we show how living standards vary with city size by drawing the 'net wage curve' (wages minus living costs) for different sized cities. It rises initially, as agglomeration economies increase more rapidly than living costs, peaking at population N^*, before falling. After N^* increasing population causes living cost increases that outweigh the wage increases generated by the additional agglomeration economies.

The optimal city size is clearly N^*, at which point net wages are maximised at w^*. So long as migration and city formation are costless, N^* is the equilibrium city size. Imagine that we had many equally-sized, sub-optimally small cities, each of population N_S and with wages w_S. In these conditions, developers would enlarge some cities by $N_S - N^*$. This leads people in cities that had not grown to move to the newly-built houses in the now-optimal sized cities, because their new net wage $w^* - c^*$ exceeds their old net wage $w_S - c_S$. That is, they improve their standard of living by moving. This is shown in figure 2a. Notice that those already living in the expanding city favour development: their net wages also increase from $w_S - c_S$ to $w^* - c^*$.

Figure 2



(2a)



(2b)

Figure 2

Figure 2b shows the opposite case, with too few, too large cities, of size N_L. In this case developers would build new cities of size N^*. Some people from existing cities then move to the new cities, since new cities offer wages of $w^* - c^*$, which exceeds $w_L - c_L$. Previously too large cities fall in size from N_L to N^*, but wages rise from $w_L - c_L$ to $w^* - c^*$. Both those who move and those who stay are better off because of the new cities creation.

So long, therefore, as developers can enlarge existing cities and build new ones, and so long as workers can migrate, observed city sizes will be optimal. This is a fundamental insight. In this simplified world, market forces (operating via wage and house price signals) will maximise living standards. Of course, we will refine this finding as we extend the model. But too often the popular debate neglects arguments in favour of the beneficial role of market signals in the rush to plan our cities' sizes and locations and constrain developers.

Of course, city sizes change over time. There are two proximate reasons why optimal city size may change: changes in the extent of agglomeration economies and changes in the cost of living in urban areas relative to non-urban areas. Increased agglomeration economies shift the wage curve upwards. It is intuitive and simple to show diagrammatically that this in turn shifts the net wage up and to the right and hence increases the optimal city size. Decreases in living costs (which given the reversed axis shift the cost of living curve "up", that is closer to the horizontal axis) have similar implications. The converse is also true: if cities become less productive, or city living becomes relatively more expensive, optimal city size falls.

We represent this in figure 3, which, for visual simplicity, excludes the falls in living costs. As can be seen, the optimal size of cities grows from N_A to N_B, with the optimal wage rising from w_A to w_B.

Figure 3

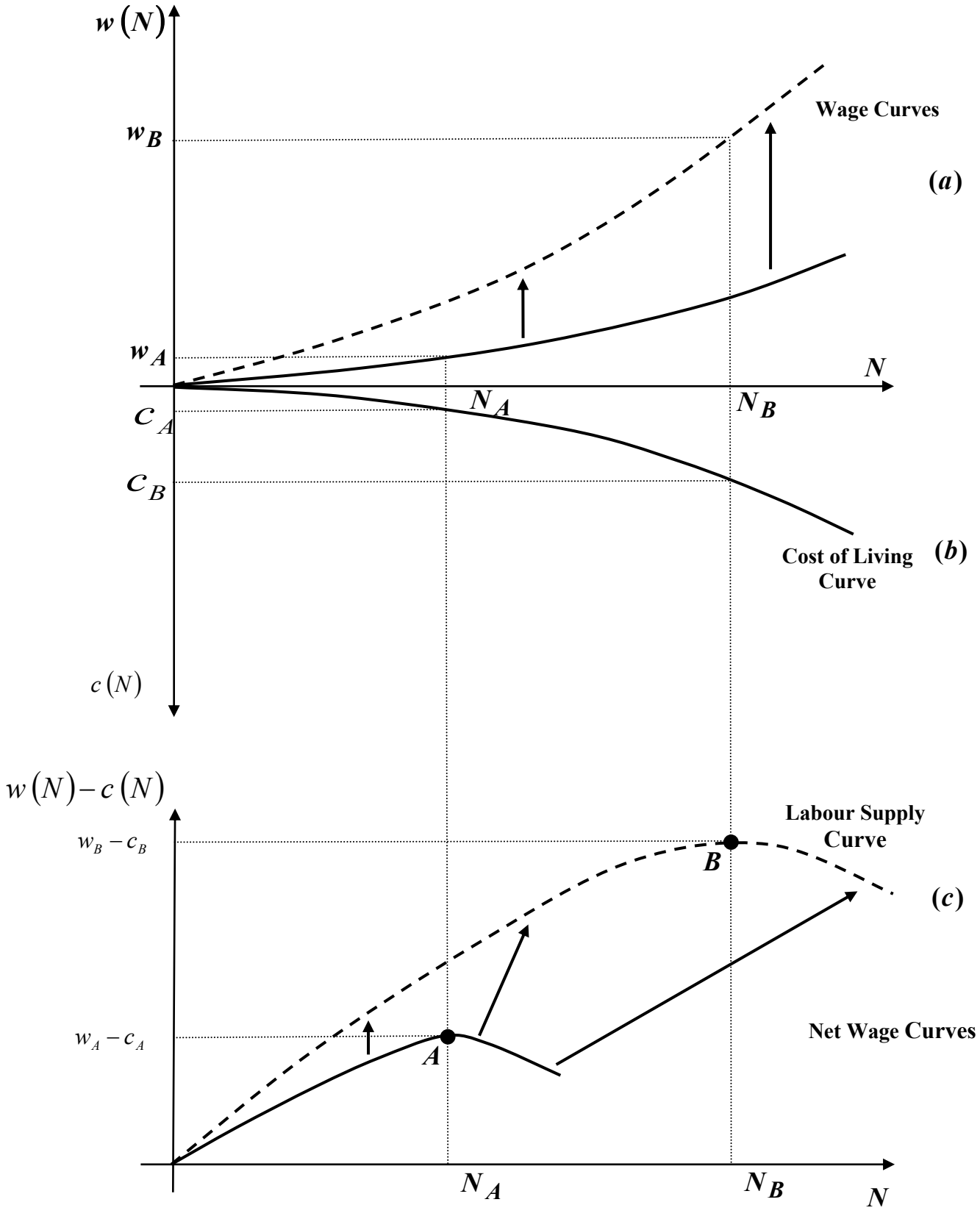

w(N)
w_B
w_A
C_A
C_B
N_A
N_B
N
Wage Curves
(a)
Cost of Living Curve
(b)
c(N)
w(N) − c(N)
w_B − c_B
w_A − c_A
A
B
Labour Supply Curve
(c)
Net Wage Curves
N_A
N_B
N

Figure 3

In outlining the model, we explained how the combination of entrepreneurial developers and free migration will ensure that all cities are of optimal size. This is true not only in the static sense, but also dynamically. The combination of developers and migration ensure that as the optimal city size changes, so actual city size will also change. Spatial differences in house prices provide good incentives to developers to build houses in the right places, while spatial differences in wages provide good incentives to workers to move to these new homes. Together, these mechanisms ensure that observed city sizes reflect changing optimal city sizes.

We note one proviso, however. Formal models of the city formation process have traditionally relied on the existence of "large agents" (large scale land developers or city governments) to ensure that house price and wage signals lead to optimal city sizes. Such large agents are needed because optimality requires those building new houses to be able to extract the additional benefit to existing city residents of increasing city size. See, for example, Henderson 1974. In the absence of such large scale agents the resulting city sizes need not be optimal. Note, crucially, however, that house price and wages still give the correct signals. It is the absence of any agents that respond to these signals that causes sub-optimality, not the nature of the signals themselves.

III. CHANGE OVER TIME 1: NINETEENTH CENTURY BRITAIN

This model will be useful for understanding reality if workers are mobile and developers free to develop new housing. The planning regime means that this is not currently the case in Britain. But nineteenth century Britain comes close to these conditions. In particular, entrepreneurs could build houses pretty much wherever they owned land, and migration was high (Baines and Woods, 2004). Studying nineteenth century experience is valuable for three reasons. First, to test whether the model explains a situation in which the economy matches the model's assumptions. Second, to get a benchmark against which to judge the effect of current restrictions that prevent the economy behaving as the model predicts. Third, because housing is durable, inherited settlement patterns affect current spatial patterns, that is, what happened in the nineteenth century matters for the twenty-first century because many of today's houses and cities were built in the nineteenth century.

The nineteenth century was critical for urbanisation in Britain: it saw the proportion of people living in cities rise from 28% to over three-quarters (Wrigley 2004 p. 88, Crafts 1985 p. 62). During this time the urban wage curve rose and relative urban living costs fell (that is, the cost curve shifted "up"). The wage curve shifted up for three reasons. First, steam power allowed firms to locate near each other, whereas water power dispersed them because any given water power site generated limited amounts of power. Second, steam ships, canals and later railways dramatically reduced transport costs. Without good transport the agglomeration advantages of producing in one place and selling more widely were outweighed by the transport costs of so doing. At the international level, the effect of reduced transport costs were reinforced by free trade policies and international peace, which substantially reduced international trade costs. Internationally production concentrated

geographically, with industrialising Britain the big winner (Crafts and Leunig, 2006).[3] The effects of the rise in the wage curve are set out in figure 3, above.

We also identify three factors that reduced urban living costs. First, transport costs fell, reducing the cost of bringing food and fuel into cities. Second, increasing imports of food meant that cheap food arrived directly into big coastal cities such as Liverpool. Third, the decline in domestic (i.e. household) production of goods such as clothing or furniture particularly benefited urban dwellers, with their better access to shops (Hudson 2004).

For all these reasons the optimal size of cities grew dramatically. Entrepreneurs responded by constructing what were, in effect, new cities. Crewe, for example, was a small village of just 46 houses and 289 residents in 1800. By 1900 entrepreneurs had built an additional 9,173 houses, and the population grew to 42,074. Although Crewe is an extreme case, many cities were transformed, particularly in North West Britain: Burnley grew from 3,305 to 97,043, Blackburn and Preston from under 12,000 to over 100,000. Large cities expanded as well: both Liverpool and Manchester grew by over 500,000, while London grew from 1m to 6.2m (Census of Population 1801 and 1901, Great Britain Historical GIS). As we noted, this rise in optimal city size was matched by massive rises in actual city sizes. The economy moving in line with the model is, in essence, British spatial economic history before the First World War.

IV.EXTENSIONS TO THE MODEL

Before turning to more recent events and current debates about the location of new housing, we need to enrich our model to include a number of important factors.

The traditional model assumes that agglomeration economies are industry-specific (Henderson 1974), so that, for example, film makers benefit from locating near other film makers but do not benefit from, say, the presence of insurance companies nearby. The model therefore predicts that each city will specialise in one industry. There is no additional benefit to locating industry X and Y together, but there is an additional cost because industry X workers increase living costs for industry Y workers and vice-versa. For this reason, these types of agglomeration economies are termed “localisation” economies. Empirical evidence suggests that many manufacturing activities benefit from such localisation economies (see, for example, Henderson 1988).

We can move beyond the prediction of single industry cities by assuming that worker productivity in one industry rises when local employment in other local industries rises. These agglomeration economies are often termed “urbanisation” economies and arise, for example, where industries share suppliers. An obvious case would be specialist business services, such as accountants and parcel delivery services. Workers having non-industry specific skills would also generate multi-

[3] The role of changing transport costs has been formalised in a recent branch of spatial economics called ‘new economic geography’ (Fujita, et al, 1999; Baldwin et al, 2003). Combes et al (2005) present this in the diagrammatic framework used here.

industry cities: industries wanting the same skills as each other would locate together. Finally, if workers live in multiple worker households such as dual earner couples, then the employers of one person in the couple will benefit from locating near the employer of the other person in the couple. Clearly, where urbanisation economies occur for a group of industries in addition to localisation economies specific to each industry, then optimal city sizes will tend to increase. Empirical evidence (see, for example, the survey by Rosenthal and Strange, 2004) suggests that urbanisation economies are larger than localisation economies and more likely to occur for service industries than for manufacturing.

If all people and industries have the same characteristics then the model predicts that all cities will be the same size. Instead, imagine two industries, one with strong and one with weak agglomeration economies. It is obvious that the strong agglomeration city will be larger because the incentives to locate together are greater. Referring back to figure 1c, both cities will have a bell-shaped net wage curve, but the curve will be further to the right for the strong agglomeration city. So long as workers can be employed in each industry, the ability to migrate ensures that net wages are equal in each place. Workers in the larger city have a higher gross wages and living costs, but in net terms are neither better nor worse off than those in smaller cities.

People's heterogeneous preferences can generate different city sizes. The living costs curve includes monetary and non-monetary aspects. For example, train commuters spend time as well as money on travel. The total cost of commuting is higher for those who dislike it more. Reverse logic holds for urban amenities: ballet companies and opera houses need large audiences and are generally found in cities so that ballet and opera lovers find the psychic cost of city life low. Specialised tastes generally make the city attractive, since providing for that taste is only worthwhile when there are many people.[4] Once we allow living cost curves to be heterogeneous across individuals, the model predicts different size cities even within the same industry. If some people dislike urban life, then a firm that employs them can still be profitable away from the city and its agglomeration economies: it pays lower wages (reflecting its workers lower productivity) but its workers are happy with lower wage since they dislike city living so much. Equally a firm in a weak agglomeration economies industry may locate in a city. It will not pay higher wages (since productivity is no higher), but there may exist workers willing to work for the firm nonetheless, because they wish to live in the city.

We have nearly all the building blocks for a detailed consideration of the spatial patterns of development in Britain, and the policy implications. Before doing so we revisit the prediction that market forces will generate optimal size cities. In reality there are various reasons why the actual city sizes will not necessarily match the optimum.

Workers deciding if higher urban wages are sufficient to compensate for higher urban living costs will compare the "take-home" wage premium with the cost premium. This means that cities will be suboptimally small. Were there no payroll taxes, workers could cover more urban living costs, and

[4] This may explain why homosexuals often live in cities. Assuming that 90% of people are heterosexual, and neglecting bisexuals for simplicity, it follows that 45% of the population are potential partners for heterosexuals (50% are of the other gender, 90% of whom are heterosexual) whereas only 5% are potential partners for homosexuals (50% x 10%), Homosexuals therefore need to meet nine times as many people as heterosexuals to be as likely to find a partner: thus cities are particularly attractive. This logic underlies the *Little Britain* "only gay in the village" sketch.

therefore would choose to live in bigger cities. Given that payroll taxes are typically 40-50% in Britain (including employer and employee payroll taxes), this bias is potentially large, and means that there is good reason to expect actual cities to be economically too small. Against that, any urban subsidies, such as for transport, reduce urban costs, so the overall effect is ambiguous.

In all cases an additional person moving into the city creates externalities that they do not consider when making the migration decision. These can be positive – one more merchant banker increases the productivity of other merchant bankers - but they may be negative, as congestion increases. The solution is not arbitrary restrictions on urban growth, but to internalise the externalities, for example, by using congestion charging. More generally, some form of "construction charge" can be levied on new housing to take account of wider externalities. Britain has such a charge: so-called "section 106" payments.[5] Similarly, if we are worried by the market's failure to include the amenity value of undeveloped land we should impose Greenfield development charges. If we are worried about the carbon footprint of development, then a carbon tax provides the right incentives for developers and purchasers to internalise the global warming externalities. Even if we abolish planning constraints, therefore, "overbuilding" will only be an issue if we fail to internalise these external costs.

We should note one final point: it is very easy to show that there exists an optimal city size, but very hard to calculate it. No amount of research gives a clear answer to this question. Instead we look to market signals to tell us if cities are too large or too small, and seek a policy framework that ensures that workers, entrepreneurs and policy makers react to these signals appropriately.

V. CHANGE OVER TIME 2: BRITAIN SINCE 1945

In section III we discussed the 19th century development of Britain's urban structure. We explained how economic changes rapidly increased the optimal size of cities and showed that the economy responded as theory predicted, with substantial urbanisation, including the creation of new cities. In this section we consider changes that have affected Britain since 1945.

There are many reasons to believe that Britain's largest cities' optimal size has grown since 1945. First, domestic and international transport costs have fallen as road transport did for the 20th century what steam did for the 19th. Second, human capital levels (i.e. skills) have risen. Third, the service sector has become increasingly important. Fourth, improving information and communication technology (ICT) has allowed the separation of company headquarters from production. Fifth, there has been a rise in dual-earner households, and particularly households with two people looking for skilled work.

Section III set out the effects of falling transport costs; the same logic applies here. Rising human capital levels and increased global integration have shifted the British economy from manufacturing towards services. From 1950, service sector employment rose from 48% to 70% of the workforce (Broadberry 2006). This has increased the optimal size of Britain's larger cities because service

[5] As we write the government is considering replacing section 106 payments with a planning gain supplement. (CLG 2007). Alternative methods of achieving the same aim include Cheshire and Sheppard (2005) and Leunig (2007).

industries generally benefit more from urbanisation economies (Graham 2007) and, as section IV argued, urbanisation economies are likely to generate large, multi-industry cities.

This move to greater optimal sizes for larger cities is reinforced by improvements in ICT which allow headquarter's activities to be separate from production activities. Headquarters provide service type activities – management, research and development, etc – and so benefit from urbanisation economies. Historically, head offices were generally located with their manufacturing plants, in single industry cities. ICT improvements allow firms to take advantage of urbanisation economies by moving their headquarters to service sector oriented cities. Thus ICT improvements have changed the optimal spatial economy from industrial to functional specialisation with headquarters located in larger cities and production plants elsewhere (possibly overseas). Diagramatically, high ICT costs meant that the net wage curve for industrially specialised cities was above that for functionally specialised cities. Recent ICT improvements have reversed this (Duranton and Puga 2005).

Finally, the big rise in two career households has increased the importance of urbanisation economies. Labour Force Survey figures show the proportion of graduate households with two earners increasing from 53% in 1977 to 77% in 2006. The returns to graduate human capital are often location sensitive. This matters when the two people are not in the same industry, since then labour in industry A is in joint supply with labour in industry B. In essence, both people in the couple want to work in the same city, and therefore industry A and B have incentives to locate together, leading to the rise of large multi-industry cities.

Given these changes we would expect to find cities that have benefited most growing since 1945. This has not been the case. The population of London has fallen by over a million since its 1951 peak. Nor, with the exception of the handful of post-war planned new cities such as Milton Keynes, have new cities sprung up, or existing smaller cities grown dramatically. Given the rising importance of high-tech spillovers, for example, we might expect Oxford and Cambridge to be the Burnley and Blackburn of the post-war era. Instead their post-1951 populations grew at only 0.5% per year, from 105,898 to 134,248 and 80,311 to 108,863 respectively. (Great Britain Historical GIS, ONS)

Since the arguments for the increasing importance of agglomeration economies since 1945 are compelling, either increased urban living costs have offset increased agglomeration economies or something has prevented the economy from adjusting in line with the predictions of our spatial economic model.

In fact, a major policy change has made the cost of living curve very steep with respect to rising city size. In our discussion of the 19th century, we noted that most land owners could develop their land as and how they saw fit. As such, the cost of housing did not increase dramatically as the city became larger: supply expanded to meet demand. After the 1947 Town and Country Planning Act would-be developers needed local authority permission to build new houses. In addition, national government heavily restricts development on around half of Britain's land, including much of South East England (Barker, 2006, pp. 33-4). The result is that growing cities cannot grow, or, to put it another way, the cost of a new house is infinite!

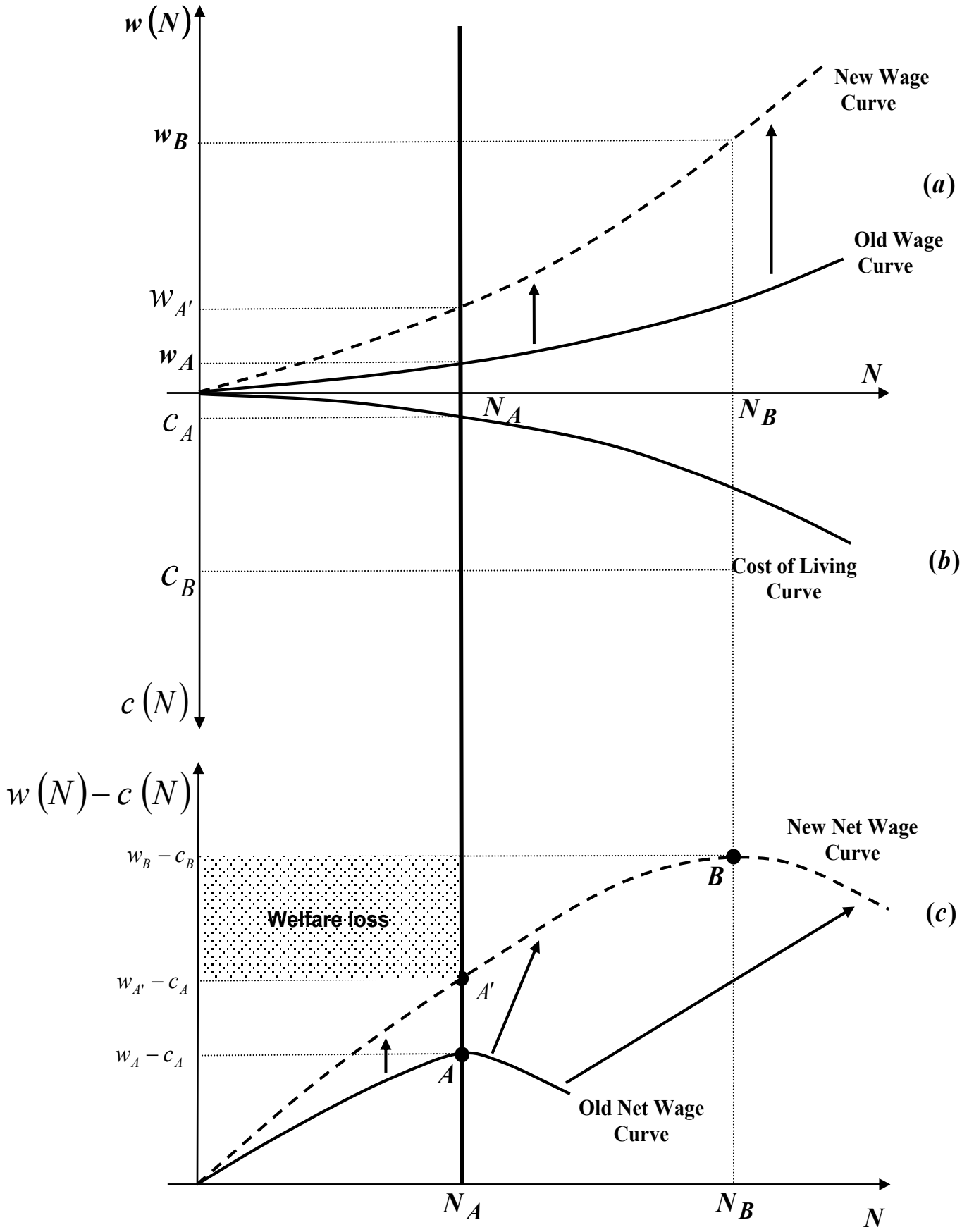

w(N)
w_B
New Wage Curve
(a)
Old Wage Curve
w_A'
w_A
N
N_A
N_B
c_A
c_B
Cost of Living Curve
(b)
c(N)
w(N) − c(N)
New Net Wage Curve
w_B − c_B
B
Welfare loss
(c)
w_A' − c_A
A'
w_A − c_A
A
Old Net Wage Curve
N_A
N_B
N

Figure 4

We can demonstrate the effects of such restrictions graphically. For simplicity, we assume a city's population is capped at the initial level. Clearly planning policy is not that restrictive, but such an extreme assumption is not an unreasonable approximation given the low levels of city growth since 1945. Figure 4 shows the effect of planning restrictions in the context of a positive productivity shock that we believe occurred for our larger cities after 1945. Were this period to have been like the nineteenth century, the economy would have moved from equilibrium A to equilibrium B, with cities growing from N_A to N_B, and wages and net wages increasing correspondingly. In reality, planning restrictions meant that cities remained at N_A. Technological change was still beneficial with wages rising from w_A to $w_{A'}$, but both gross wages and living standards (shown in panel c) are lower than they would have been without planning restrictions. The shaded area shows the welfare loss to those already living in that city, with additional losses for those who were unable to move to it.

The welfare state also reduces adjustment of the spatial economy. To see this, consider a former industrial city that has experienced a negative productivity shock. The diagram (figure 5) is the reverse of figure 3, that is, the wage curve shifts downwards. In the standard model, both the wage curve and the net wage curve move from the upper to the lower solid lines, and the city's population falls from N_A to N_B. Imagine now a level of benefits set between the original and new wage curve, shown as the dotted line in figure 5a. We show benefits rising as population increases because housing benefit is linked to housing costs which are higher in larger urban areas. Now the new wage level is not feasible, because it is below the level of benefits.

Figure 5

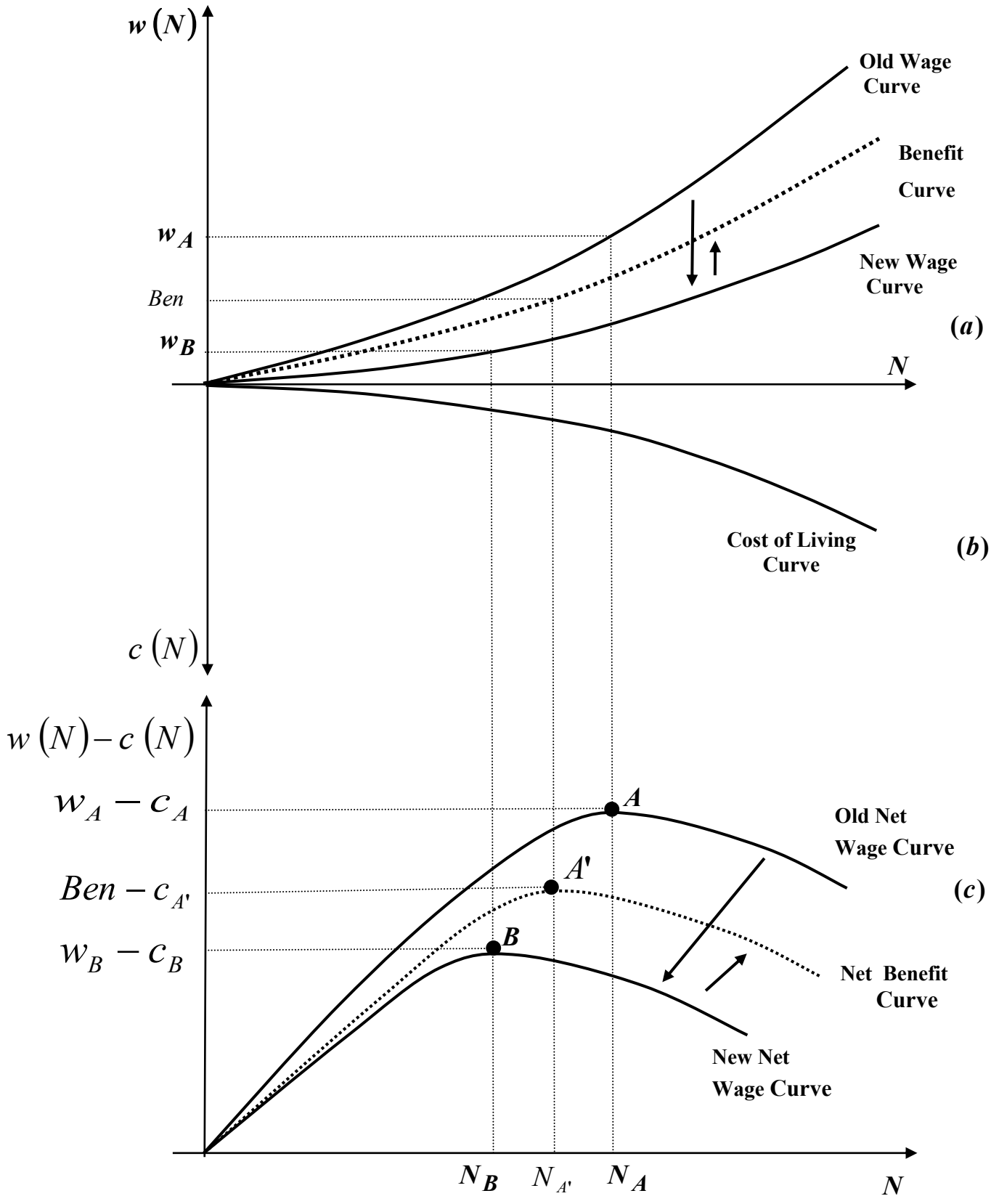

w(N)
Old Wage Curve
Benefit Curve
w_A
Ben
w_B
New Wage Curve
(a)
N
Cost of Living Curve
(b)
c(N)
w(N) − c(N)
w_A − c_A
Ben − c_A'
w_B − c_B
A
A'
B
Old Net Wage Curve
(c)
Net Benefit Curve
New Net Wage Curve
N_B
N_A'
N_A
N

Figure 5

This has two effects. First, it reduces migration. Panel c shows that city size falls only from N_A to N_C, so that population in the declining city is larger than optimal by $N_B - N_C$. Furthermore, unemployment is higher. Strictly this model predicts that everyone is unemployed, which is obviously unrealistic. Instead imagine skill-segmented labour markets in the city. For high skill, high wage employment, benefits are not an issue, and the high skill population migrate in line with the original model, with those who stay remaining in employment. For the low skilled, in contrast, benefits exceed the new equilibrium wage, and therefore they prefer unemployment to migration. This gives two predictions: skilled people are more likely to migrate than unskilled people, and many unskilled people who remain will not be working. The evidence supports these predictions: only 19% of those without GCSE or equivalent qualifications in Windsor are not employed, while 33% of the equivalent group in Blackburn are not in work (Labour Force Survey, 2007).

We have set out good reasons why Britain's successful cities are too small and some declining cities are too large. There are many independent pieces of evidence to support this – we highlight two that are particularly telling given the spatial economic framework we have adopted.

The construction of Canary Wharf added substantially to London's office space. If the previous level had been sufficient, then Canary Wharf would have caused London office rents to fall. They did not: City of London rents are £620/sqm, the highest in Britain after Mayfair (Valuation Office Agency 2007). Rather, Canary Wharf has allowed London's financial centre to grow dramatically. Note that as it has done so, wages have risen, not fallen. This is entirely consistent with London's financial centre exhibiting strong agglomeration economies, and is inconsistent with any other story. There is no reason to think Canary Wharf has exhausted London's agglomeration possibilities, and it is plausible that the proposed 82 storey Shard London Bridge, 48 storey Leadenhall Street, and 46 storey Heron Tower will add to London's agglomeration economies.

More generally, if we are right that planning restrictions are preventing urban growth, then urban land values will be higher in cities than in smaller places nearby. This is correct: prices in both London and Manchester, for example, are over double those of their regions respectively (Valuation Office Agency 2007).

VI. OPTIMAL LOCATIONS

The argument so far suggests that Britain's service sector centres are too small (because of planning constraints) while some declining manufacturing cities are too large (because of the benefits system). A cursory knowledge of Britain's economic geography suggests a broad spatial pattern as to these types of cities' locations. Spatial economics emphasises that these patterns are the result of the economic system reinforcing the initial advantages of different locations.

In the 19th century, as Britain industrialised, the best locations were those with access to two key resources – coal and water. The 19th century was powered by coal, which is heavy and expensive to move. This meant that 19th century industry had huge incentives to locate near the coal field, or at least near a port. Ports also provided access to the shipping routes that served demand throughout

the Empire and wider world. Consistent with this story, the Victorian era saw many people move to London, and to cities near to coalfields or ports. In terms of our model, these places had the highest net wage curves. Households responded to higher net wages by moving, and developers responded by building more houses. That is, wage and house price signals reinforced the spatial differences arising from location advantages.

Clearly, many of these cities are no longer economically successful. Three factors explain this: changes in energy sources, transport, and the location of demand.

Coal is no longer an important source of energy, replaced by oil, gas, and electricity. All of these are relatively cheap to transport over long distances, so that firms no longer have a large incentive to locate near to any particular energy source.

Second, the nature as well as price of transport has changed. Internally, road haulage has replaced rail, canal and coastal shipping. In 1953 more freight was moved by rail than by road, today road is seven times as important (Department for Transport 2006 table 4.1). When coastal shipping was important, ports were good places to have warehouses. In 1900 Britain's biggest grocer, Lipton's, had three warehouses: London, Liverpool and Glasgow. Today Tesco's biggest warehouse is at Milton Keynes, far from any coast, but right in the middle of Britain's motorway network. The vast majority of Britain's supermarket warehouses are located along Britain's geographical backbones: the M1-M6 corridor from London to Leeds-Manchester, and the M4 from London to Cardiff (Leunig and Swaffield 2007).

19th century international merchandise trade went by ship, today much goes by air. This change has meant a move from seaports towards airports, and, furthermore, a move towards London whose airports account for over three quarters of air freight tonnage to the British mainland (CAA 2006, table 13.2). Even within shipping, there has been a move from North to South. As late as 1972 Liverpool was responsible for 11% of British merchandise exports, by 1992 it had fallen to 3%. In contrast the combined shares of Dover and Felixstowe rose from 10% to 29% (Overman and Winters 2005).

This southerly shift in the pattern of external trade is part of a larger change in Britain's key markets. In 1900 the British economy was globally oriented, with particular emphasis on the Empire. Today it is still globally oriented, but with a greater European emphasis. Furthermore, while Liverpool was Britain's closest port to North America in 1900, today that honour is held by London, whose airports offer excellent transatlantic links.

Economists assess the degree of connectivity by looking at 'market potential', which measures the income in and near an area taking into account transport costs (Ormerod et al 2006). It is straightforward to see how this can be useful in measuring potential agglomeration economies: areas with large market potential have better prospective customer-supplier relationships, thick labour markets and ample learning opportunities.

Table 1 shows that since 1911, the already low market potential of British regions has fallen back relative to London. The exception is the West Midlands. It is not particularly that London is now

more populous or more wealthy, but rather the changing transport system that accounts for almost all of the changes. As Crafts notes 'as the world moved from rail and ship to road and ferry, the 'peripherality' of the North, Wales and Scotland was seriously accentuated' (Crafts 2005 p. 1164).

Table 1: Market potential

Per cent of London and South East values	1911	1985
Scotland	56.1	37.5
Wales	69.1	53.6
North	62.0	47.5
West Midlands	61.3	71.5

Source: Crafts (2005) table 7

Just as the optimal location for much of the population moved to the coal fields and major ports in the nineteenth century, so it has moved to the South and to London in the twentieth. This process began after the First World War, and has continued apace since 1945. That is not, of course, to say that there will be no areas of economic success outside of the South East (witness the nascent resurgence of Leeds and Manchester). Some theoretical models (see, for example, Fujita, Krugman and Venables 1999) even suggest that the trend towards the concentration of economic activity in the South East might be reversed if long distance transport costs fall sufficiently far while congestion increases in the South. That said, such circumstances would be best addressed, in the first instance, with measures to tackle congestion. Once these were implemented, it is indeed possible that the "optimal" congestion remaining in the South East might indeed lead economic activity to move away from the area. However, on balance, Britain's centre of economic gravity is currently shifting southwards, implying that more of Britain's population and more of Britain's larger cities should be located there (Scott 2007).

VII. IMPLICATIONS

This paper started from the premise that Britain will build 3 million new houses, and asks where they should be built to maximise living standards. We argue that spatial economics offers four insights (i) that the average size of cities should rise (ii) that within this, some cities need to expand while others need to contract (iii) that cities with high land values should expand (iv) that the optimal degree of expansion may be considerable.

Spatial economics tells us that the optimal size of cities will grow under certain conditions. We described various factors that caused the average optimal size of cities to grow since 1945. Since actual city growth has been limited since 1945, it follows that many of our successful cities are currently too small. There is no reason to think that the changes we identify as causing the rise in optimal city size – such as increased skill levels, service sector growth, or the rise of dual career households – have stopped, and as such optimal city size are likely to continue growing. The first implication, therefore, is that the optimal size of many of our successful cities has grown, and continues growing.

Second, we saw that the optimal location for big cities today may be radically different from the optimal location during the industrial revolution. In particular, there are theoretical and empirical

reasons to believe that the size of a small number of high skill, service orientated cities should increase, perhaps dramatically. The question then is how to identify these places. We need to find places whose size is currently constrained to be a long way below their optimal size, which in turn means that their net wages are below their optimal levels. The constraints imposed by the planning system means that land prices are an effective way to do this. There are two reasons why land prices may be high. First, there could be an unexpandable local amenity, for example, a seaview, or a village feel, that makes an area desirable. Or, more commonly, people may want to live in an area because it is an area with good jobs. We need to be able to distinguish between the two, since expanding the size of the former would be value destroying, but expanding the latter value creating. This can be done via local auction mechanisms, whereby local people can decide on the price at which they are prepared to accept local development. If an extra house will destroy the quality of the area, the price that they will demand will be so high as to make that extra house uneconomic to build. But if the area is popular because of employment, then rational local people will be prepared to accept money from developers in exchange for the right to build additional housing in the area (Leunig 2007). Indeed, if additional housing increases agglomeration economies, local people may be willing to accept additional housing without requiring direct payments.

In areas of high derived demand for land, the number of houses is lower than the number of potential good jobs, leading people to bid up the price of land in the area. We can, therefore, have a strong *a priori* belief that the planning system is constraining city size below its optimal level in such places. In addition to London and Manchester, two areas stand out as having particularly high land prices: the university cities of Oxford and Cambridge. In both cases, a hectare of land designated for housing is worth £8m, more than twice the UK average. Other university cities, such as Bath, Bristol, Cardiff, Durham, Edinburgh and York also have reasonably high land prices, suggesting possible potential for market-led expansion in these areas as well. In addition, there are strong market signals telling us that it would be economically advantageous to expand London, including suburban London and London commuter cities such as Guildford and Reigate.

Figure 6

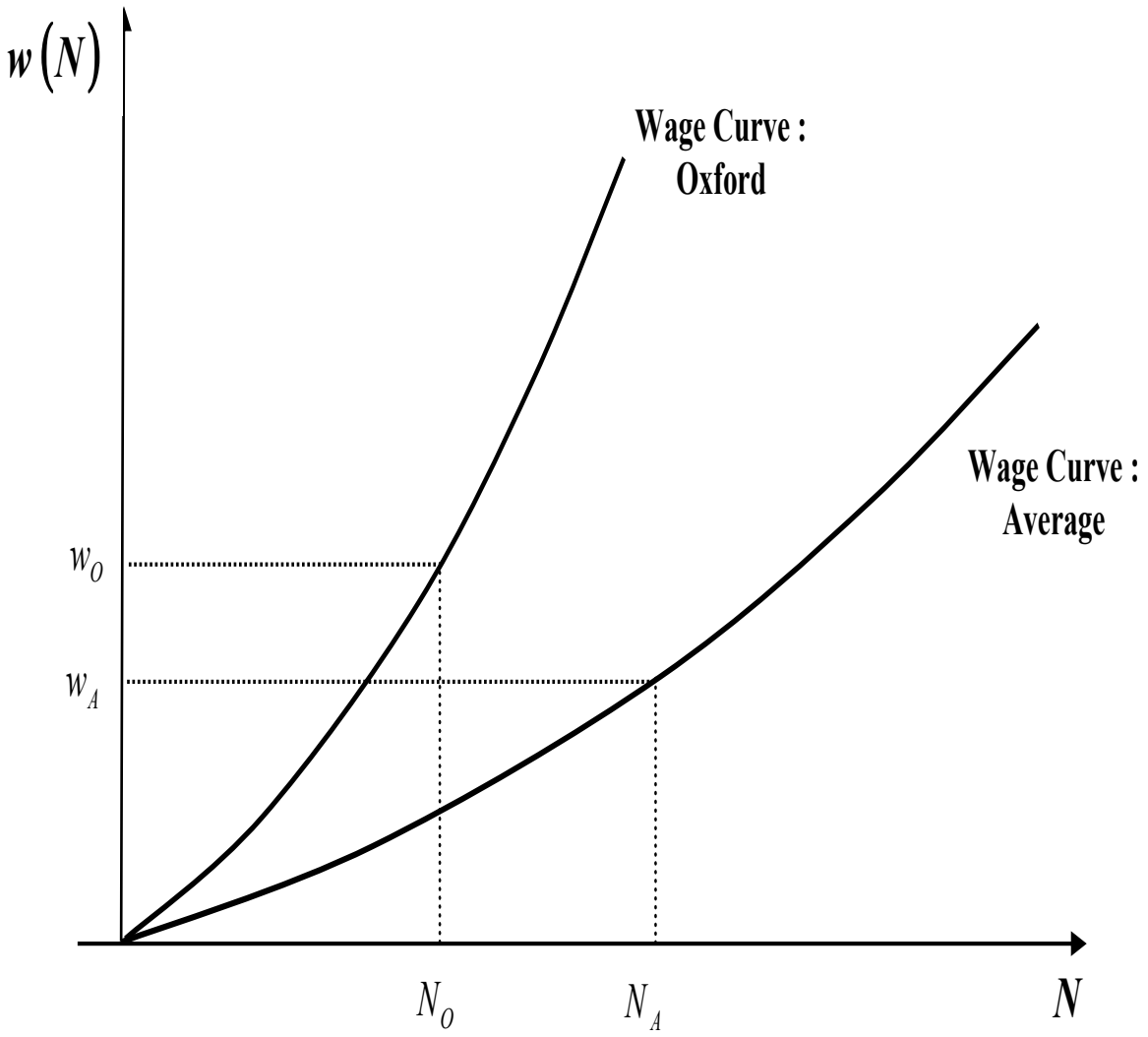


Figure 6

At this point, it is important to reiterate a key issue raised earlier - it is not possible to calculate the optimal size of cities such as Oxford and Cambridge. What we can say for certain is that so long as land prices in those cities remain markedly higher than elsewhere, then spatial economics implies that we will maximise living standards by expanding these cities. The self-reinforcing nature of agglomeration economies means that these cities' optimal sizes might be very large indeed. Oxford, for example, is smaller than the average British city, but has higher GVA. Diagramatically figure 6 shows that the Oxford wage curve must lie above that of the average city by a sufficient amount to ensure that actual Oxford wages (w_O) exceed actual average wage (w_A), even though Oxford's population (N_O) and so agglomeration economies are smaller than those for the average city (N_A). In short, there is good reason to think that the wage curve in Oxford is much higher than the wage curve for a typical British city, which would in turn imply that Oxford's optimal size is significantly larger than that of the typical British city.

Agglomeration economies mean that allowing developers to respond to market signals will not necessarily cut house prices, but *will* improve affordability and living standards. Imagine building many new houses in Cambridge. At first sight, more houses means, ceteris paribus, lower house prices. In this case, however, the ceteris paribus clause does not hold, since increasing population means more agglomeration economies and higher wages, which in turn means that workers are willing to pay higher house prices, which in turn raises land prices. Thus the effect of building more houses on house prices is ambiguous.

Notice three things. First, if land prices remain relatively high, that is a signal to continue to build more houses in Cambridge until the Cambridge land price premium falls: if land prices remain high we know that the marginal job in Cambridge continues to have a sufficiently high net wage relative to jobs elsewhere, to persuade the newest worker to pay the high cost of land in Cambridge. In essence, if building more houses in an area does not cut land prices, that shows that not enough new houses have been built!

Second, house prices will not rise once Cambridge grows beyond a certain point: at some point the additional population will increase the commuting and congestion costs by more than the increase in agglomeration economies. At this point the value of housing will stop rising, since the demand for housing is derived demand. The price people are willing to pay for a house in Cambridge is equal to the basic wage, minus the other costs of living in the area. When commuting and congestion costs increase by more than the rise in agglomeration-economy induced wage rises, the amount people are willing to pay for housing in Cambridge starts to decline, and at that point the value of land in Cambridge declines. Notice that this does not necessarily mean that too many houses have been built: the Cambridge wage minus Cambridge non-housing costs could still exceed the equivalent values elsewhere. In short, falling land prices (relative to those elsewhere in the country) in currently high land price areas like Cambridge is a necessary but not sufficient condition for Cambridge reaching its optimal size.

Third, housing does not become less affordable even when prices rise, because the rise was induced by rising wages.. Indeed, in the presence of mortgage regulations that prevent all of marginal income being used for housing, post housing incomes must rise. Third, affordability improves nationwide, for two reasons. First, as people move from other areas to Cambridge, housing supply eases in other cities even if no additional houses are built there. Second, as incomes increase in Cambridge, real wages increase elsewhere, as competition for the remaining workers forces the least efficient, lowest-paying firms elsewhere out of business (Rice and Venables 2003). Both these factors make housing more affordable across Britain. Finally, higher average wages mean higher tax revenues, since tax revenues are related to the level of economic activity, allowing the government to improve housing affordability either by cutting taxes or by non-market means, such as building more social housing.

We make three further comments. First, houses take time to build. Given that we do not know the optimal size of any city today, and cannot predict it for 2020, the best approach is not to say where the 3 million houses should go, but to set out either the principles that government will use to decide on location, or to construct incentives that will allow the invisible hand to locate new

housing in areas with the highest land values. The slow rate at which houses can be built means that oversupply in any particular city is most unlikely.[6]

Second, housebuilding levels will match optimal levels if the externalities discussed earlier are internalised. As we noted, this will often include congestion charging and development charges. Note, however, that the ratio of residential land to farm land prices – close to 1,000:1 in areas such as Oxford and Cambridge – is so extreme as to render it extremely unlikely that the benefits will outweigh the costs. The question is not whether to build more, but how much more to build.

Finally, commuting allows people today to live and work in different places, leading income levels between cities and their hinterlands to (partially) converge, particularly between London and the South East.

Table 2: Ratio of London to South East GDP per head,

1871	1911	1954/5	2001
159%	192%	141%	113%

Source: Crafts (2005)

A century ago the south east outside of London was a place to migrate from, today it is a place to commute from. As a result society can choose whether to build new houses in London (and other metropolises) themselves, or in their commutable hinterlands. Land values suggest both are appropriate for London, but less so for other major cities, where nearby cities have much lower land values, reflecting the fact that it is harder to commute to big cities in areas that do not have London's quantity of public transport.

Transport also has a role in connecting places that already exist. As agglomeration economies imply larger cities, it may be cheaper to connect two already existing cities with high speed transport links than to demolish one and double the size of the other. We should not, however, expect too much of transport. We might be able to integrate Bolton into the Manchester economy, but cannot hope to integrate it into the London economy. Transport can act as a substitute for migration over short distances, but it cannot replace it over medium or long distances.

VIII. CONCLUSIONS

This article shows that spatial economics can make a useful contribution to the debate on the optimal location for the 3 million houses that the government wants to see built. Spatial economic theory shows that in terms of living standards there are optimal sizes and locations for cities, which change over time as the economy changes. Although spatial economics cannot quantify the optimal size and location at any time, it can show how to use price signals to assess whether specific cities and cities in general are too small. Increasing numbers of high skilled workers and dual career households, the growth of the service sector, the growth of air travel and the reorientation of the

[6] Note that so long as we constrain ourselves to thinking about the construction of new housing, rather than the demolition and replacement of existing housing, there is no need to consider the cost of demolishing and rebuilding schools and so on. As noted earlier, the cost of constructing new infrastructure could be covered by section 106 money.

economy towards Europe all point towards a rise in the optimal size of cities, and of a move in the optimal location from the industrial revolution cities of Northern England, Scotland and Wales towards the South East. This conclusion is strongly supported by the available price data. As a result, significantly expanding London and its commuter satellites, as well as from expanding other high skill places in the South East, is likely to generate substantial rises in both wages and the standard of living.

IX. REFERENCES

Baines, D. and Woods, R. (2004), 'Population and Regional Development' in *The Cambridge Economic History of Modern Britain* Volume 2: Economic Maturity 1860-1939 Cambridge University Press.

Baldwin, R. Forslid, R., Martin, P., Ottaviano, G. and F. Robert-Nicoud (2003), *Economic Geography and Public Policy*, Princeton University Press.

Barker, K. (2004), *Delivering Stability: Securing our Future Housing Need*s, H.M. Treasury, London.

Barker, K . (2006), *Barker Review of Land Use Planning: Final Report*, H.M. Treasury, London.

Broadberry, S. (2006), *Market Services and the Productivity Race, 1850-2000: British Performance in International Perspective,* Cambridge University Press.

CAA (2006), UK Airport Statistics London.

Cheshire, P. and Sheppard, S. (2005) 'The introduction of price signals into land use planning decision-making: a proposal', *Urban studies*, 42 (4). 647 -663.

Combes, P. P., Duranton, G. and Overman, H. G. (2005) 'Agglomeration and the Adjustment of the Spatial Economy', *Papers in Regional Science*, 84, 311-349.

CPRE (2004), *A Basis on Which to Build*: A report by Europe Economics, CPRE, London.

CPRE (2007), *Planning for Housing Affordability: Why providing more land for housebuilding will not reduce house prices*, CPRE, London.

Centre for Cities (2007), *A Question of Balance*, Centre for Cities, London.

Crafts, N. (1985), *British Economic Growth during the Industrial Revolution,* Oxford University Press

Crafts N. (2005), 'Market potential in British regions, 1871-1931', *Regional Studies*, 39(9): 1159-1166

Department for Transport (2006), *Transport Statistics for Great Britain* London

Crafts N.F.R. and Leunig T. (2006), 'The historical significance of transport for economic growth and productivity' in The Eddington Transport Study Research Annex volume 1: *Understanding the relationship: How transport can contribute to economic success*

Duranton, G. and Puga, D. (2004), 'Microfoundations of urban agglomeration economies' in J. V. Henderson and J.-F. Thisse (eds.) *Handbook of Regional and Urban Economics*, Volume 4, pp. 2063-2117.

Duranton, G. and Puga, D. (2005), 'From sectoral to functional urban specialisation', *Journal of Urban Economics*, 57(2), 343-370.

Fujita, M. and Thisse, J. F. (2002), *Economics of Agglomeration: Cities, Industrial Location and Regional Growth*, Cambridge University Press, Cambridge, England.

Fujita, M., Krugman, P. and A. J. Venables (1999). *The Spatial Economy: Cities, Regions and International Trade*, MIT Press, Cambridge, MA

Graham, D. (2007), 'Identifying urbanization and localization externalities in manufacturing and service industries', *Paper in Regional Science,* forthcoming.

Great Britain Historical GIS Project, University of Portsmouth, http://vision.edina.ac.uk/index.jsp

Henderson, V. (1974), 'The Sizes and Types of Cities', *American Economic Review*, 64, 640-656.

Henderson, V. (1988), *Urban Development: Theory, Fact, and Illusion*, Oxford University Press.

Hudson, P. (2004), 'Industrial organisation and structure' in Floud, R. and Johnson, P., *The Cambridge Economic History of Modern Britain* Volume 1: Industrialisation 1700-1860, Cambridge University Press.

Leunig, T. (2007), *In my back yard: unlocking the planning system,* CentreForum, London

Leunig T. and Swaffield J. (2007), *Cities Limited: An Appraisal of Urban Policy,* Policy Exchange, London.

Marshall, Alfred (1890), *Principles of Economics,* London: Macmillan and Co., Ltd.

Ormerod P., Cook W, Rosewell, B. (2006), 'Why Distance doesn't die: Agglomeration and its benefits', GLA Economics Working Paper 17

Overman H. G., and Winters A. L., (2005), 'The geography of UK international trade', *Environment and Planning A*, 37(10): 1751-1768

Rosenthal, S. and Strange, W. (2004), 'Evidence on the nature and sources of agglomeration economies' in J. V. Henderson and J.-F. Thisse (eds.) *Handbook of Regional and Urban Economics*, Volume 4, pp. 2119-2171.

Rice, P. and Venables, A.J. (2003), "Equilibrium Regional Disparities: Theory and British Evidence", *Regional Studies*, 37 (6/7), 675-687.

Scott, P. (2007), *Triumph of the South: Regional Development in the Twentieth Century,* Ashgate: Aldershot.

United Kingdom: Population Census of 1801: Population Act, 41 Geo III, 1800.

United Kingdom: Population Census of 1901: Census of England and Wales 63 Vict c 4.

Valuation Office Agency (2007,) *Property Market Report* January edition, VOA, London.

Wrigley, E. A., (2004), 'British population during the long eighteenth century' in Floud, R and Johnson, P, *The Cambridge Economic History of Modern Britain* Volume 1: Industrialisation 1700-1860, Cambridge University Press.