Expanding cities and connecting cities: appraising the effects of transport improvements.

Anthony J. Venables,

University of Oxford and CEPR

Abstract
Transport improvements facilitate the movement of people or goods within or between cities. The economic benefits of such improvements accrue through three main mechanisms. The first is the direct cost saving to existing and new traffic as captured by standard cost benefit analysis (CBA). The second occurs if there are urbanisation economies. Direct benefits are likely to increase economic activity in affected cities, and this may be amplified by increasing returns and urbanisation economies. The third mechanism applies principally to inter-city improvements which enable cities to specialise in particular sectors or tasks. If there are city-task level economies of scale (localisation economies) this is a further source of gain. This paper develops a model in which to analyse these mechanisms. An expression is derived for the multipliers that should be applied to a conventional CBA for intra- and inter-city improvements. Gains from specialisation and consequent scale economies may be large although, since there may be multiple equilibria, a transport improvement may be a necessary but not sufficient condition for the gains to be achieved.

Keywords: transport, cost benefit analysis, wider impact, urbanisation economies, localisation economies.

JEL classification: R1, R4, F12

Author’s Address:
A.J. Venables
Department of Economics
Manor Road
Oxford OX1 3UQ UK

tony.venables@economics.ox.ac.uk
1. Introduction

How do improvements in communications – the traffic of people, goods or ideas within and between cities – affect the distribution of activity between cities? What are the mechanisms through which these changes alter the level and spatial distribution of real income and welfare, and how should they be incorporated in the ex ante appraisal of transport projects? The objective of this paper is to address these questions in a simple analytical structure which includes both intra- and inter-city transport improvements and which encompasses a variety of different economic environments. These include production under constant returns to scale and under increasing returns which may be city wide (urbanisation economies) or sectoral (localisation economies). We derive expressions for multipliers that should be applied to standard cost-benefit analyses of both intra- and inter- city transport improvements and show how they depend on the extent and form of scale economies.

The mechanisms that we study fall under three broad headings. The first are those captured by a standard cost-benefit analysis (CBA): the value of the direct cost saving on existing traffic plus (for non-marginal changes) a ‘triangle’ of a benefit from traffic created. The second arises as the direct benefits of a transport improvement may cause a city to grow. If there are economies of scale at the city level (urbanisation economies) this growth is amplified, bringing additional benefits. This mechanism has been explored in previous work (Venables 2007, Kanemoto 2013) and is part of established appraisal techniques (Department for Transport 2005, Mackie et al. 2011). However, both the ideas and the application have been principally in the context of intra-city transport improvements. This paper extends the method to include inter-city transport projects.

The third mechanism arises as better inter-city communications may enable cities to specialise in production of particular sectors or ‘tasks’. Specialisation and the gains from trade are not, of themselves, a source of benefit beyond conventional CBA. However, if specialisation combines with market imperfections then additional gains may follow.1 To capture this we look at economies of scale at the city-task level (localisation economies) and at how better communications may facilitate urban specialisation, enabling these localisation economies to be achieved.2 The empirical importance of localisation economies is well

---

1 Absent market imperfections, the demand curve for transport captures willingness to pay. The ‘new trade theory’ includes market failures and additional gains from trade deriving from more intense competition, more product varieties, and increasing returns: this paper focuses on increasing returns. Differences in firm productivity and consequence firm selection (as in Melitz 2003) are not a source of market imperfection.
established (see the surveys by Rosenthal and Strange 2004, Combes and Gobillon 2015). Henderson (2003) finds evidence of such economies in manufacturing, particularly in high-tech sector, and Graham (2007) finds that they are larger in service sectors than in manufacturing.\(^3\) Clusters of service activities suggest that they are particularly important in business services such as finance, law, advertising, and media. Better inter-city transport facilitates business links and hence trade in these tasks, enabling firms in one city to develop their own specialisms and import the specialisms of other cities and so, if localisation economies are present, raising productivity in both.

Analysis is based on a family of simple models in which outcomes are determined by two key relationships for each city; labour supply and labour demand. The labour supply equation links wages to city size and to intra-city commuting costs, and comes from a long-term perspective in which there is mobility of people within and between cities. Underpinning this relationship are urban costs, i.e. the additional commuting costs and higher land prices found in large cities. To attract and hold population the wage offered by a city must therefore be higher the larger the city and the more costly its commuting. The formalisation of this is based on the standard urban model (the Alonso-Mills-Muth model). Intra-city transport improvements enter the model by changing urban costs and shifting this labour supply relationship.

Labour demand is derived from the production side of the economy and links productivity and wages to the level of employment and to inter-city trade costs. We initially assume a simple reduced form of this relationship and, in sections 3 and 4, develop this into a structured model of trade in which final output is produced using a range of intermediate goods or business services (collectively referred to as ‘tasks’).\(^4\) These tasks can be traded within and between cities, but trading them between cities is costly. This captures the idea that proximity to suppliers/ customers is important for business services and intermediate goods. Inter-city transport improvements lower these costs, thereby improving business links and enabling firms to access a wider range of lower cost and/or more specialist task suppliers.\(^5\) This in turn may trigger urban specialisation in particular tasks. This, as we shall see, is a source of ‘wider benefits’ of transport improvements if (and only if) there are market imperfections such as increasing returns or imperfect competition in production.

\(^3\) See de Melo et al. (2009) for a meta-survey.

\(^4\) The terminology ‘tasks’ was coined by Grossman and Rossi-Hansburg (2008). See Anas and Xiong (2003) for an urban application. For empirical work on task specialisation see Michaels et al. (2013).

\(^5\) Bernard, Moxnes and Saito (2015) study extension of the shinkansen network in Japan to establish the role of passenger transport improvement in increasing the number and spatial range of supplier-customer links.
The arguments are developed in three stages. Section 2 of the paper sets out the model and looks at the effects of transport or communications improvements under the assumption of constant returns to scale. Intra-city improvements shift the labour supply equation while inter-city improvements shift the labour demand relationship equation. We show that, in both cases, the welfare gains are as derived under a conventional CBA.

Section 3 deals with urbanisation economies, i.e. increasing returns to overall economic activity in the city, arising either because of monopolistic competition or positive externalities between firms economies. In this case a multiplier should be applied to a conventional CBA in order to capture the full welfare effects of a transport improvement. This is consistent with previous work on intra-city transport improvements which we extend to cover inter- as well as intra-city contexts. The multipliers are larger the greater are economies of scale and the smaller the city. Since both the direct effects and the multiplier are larger for a smaller city, it follows that inter-city improvements are a force for convergence.

Section 4 turns to the case where productivity is city- task specific (localisation economies). Transport improvements can then trigger the specialisation of cities in particular tasks, bringing additional benefits if localisation economies are present. This is an environment in which there are multiple equilibria arising because comparative advantage is endogenously determined by the level of activity in each task. Transport improvements can move cities to a point at which specialisation is triggered, this bringing (discontinuously) large welfare gains. It is also possible that transport improvements, together with urban development policies (such as development of city areas around transport hubs) can act as a coordination mechanism, triggering specialisation and associated real income gains. The analysis suggests that, at least in central cases, benefits accrue to both cities, but disproportionately more to the smaller city.

2. The model

For simplicity we work with just two cities (city specific variables subscripted $i = 1, 2$) and the rest of the economy. The rest of the economy is endowed with labour which is used to produce an outside good (or service) with constant returns to scale. We initially focus on a single city, describing labour demand and labour supply, and using comparative statics to derive the social value of changes in commuting and transport costs.

Production and labour demand: Each city produces a final consumption good which is freely tradable, faces an infinitely elastic demand curve from the rest of the economy, and
which we take to be the numeraire. For the moment, we characterise labour demand in each city by inverse demand curve,

\[ w_i = W(L_i, t). \]  

(1)

The wage in city \( i \) is \( w_i \), and the function \( W(L_i, t) \) is the value average product of labour, depending on employment in the city, \( L_i \), and the cost of inter-city trade or communication, denoted, \( t \). Production is assumed to have constant or increasing returns to city employment, so the wage is non-decreasing in city employment \( W_i(L_i, t) = \partial W(L_i, t) / \partial L_i \geq 0 \). Increasing returns are external to individual firms which break-even paying labour its average value product. Inter-city trade costs, \( t \), raise firms’ costs and hence reduce the wage that firms offer, so \( W_i(L_i, t) < 0 \). We will refer to this relationship as the ‘wage equation’ and give it a micro-economic foundation and specific functional form in section 3.

**Population and labour supply:** There is perfect labour mobility between locations (within and between cities and non-urban areas), and the real wage in non-urban areas is constant \( w_0 \). Following the classic Alonso-Muth-Mills model urban workers are employed in their city’s central business district (CBD), and each occupies a unit of residential space.\(^6\) The unit rental of space at distance \( z \) from the CBD is denoted \( r_i(z) \) and commuting costs are \( c_i \) per unit distance. Workers choose residential location within and between cities, so equilibrium rent functions in each city, \( r_i(z) \), must be such that real wages in the city (nominal wage minus commuting costs and rent) are the same everywhere, \( w_i - c_i z - r_i(z) = w_0 \) for all \( z \).

In a linear city in which there are \( K \) spokes from the CBD along which people live and travel takes place, population is \( L_i = K z^*_i \), where \( z^*_i \) is the edge of the city (length of each spoke).

At the city edge land rent is zero, so \( w_0 = w_i - c_i z^*_i = w_i - c_i L_i / K \) giving the labour-supply equation

\[ w_i = w_0 + c_i L_i / K. \]  

(2)

This equation says that the larger is the city the higher the wage it has to pay in order to cover commuting costs and rents incurred by workers, and thereby attract and hold its labour force.

---

\(^6\) See Duranton and Puga (2015) for a full exposition and discussion of this model and its extensions.
Urban equilibrium: The equilibrium of each city is the pair of values \( \{ w_i, L_i \} \) at which labour demand equals supply. Using eqn. (1) in (2), city employment \( L_i \) is implicitly defined by

\[
W(L_i, t) - c_i L_i / K = w_0. \tag{3}
\]

The economic surplus created in each city is the total value of rents, or equivalently is city income net of commuting costs and the opportunity cost of labour in non-urban use,

\[
R_i = L_i (w_i - w_0) - \int_0^{z_i} c_i z Kdz = L_i (w_i - w_0) - c_i L_i^2 / 2K = c_i L_i^2 / 2K \tag{4}
\]

(derived using eqn. 2 and \( L_i = K z_i^* \)). This measures the city’s welfare and is the criterion used to measure the value of transport improvements in this and the next section.

It is important to note that use of this as a national welfare criterion rests on the assumption that production in the rest of the economy operates under constant returns to scale and at price equal to marginal cost, assumptions which mean that changes in the scale of activity in the rest of the economy have no effect on real income. If there were decreasing returns in the rest of the economy then urban expansion would bring additional welfare gain as wages are bid up. If there were increasing returns then welfare effects would be reduced due to displacement effects, as expansion of the cities under study would reduce scale elsewhere (see Kanemoto 2013). We return to this issue in section 4 in the context of localisation economies, i.e. scale economies at the city-sector level. It is then no longer just total city population that determines productivity, but also its allocation between sectors. We modify the welfare criterion appropriately and show that transport can then bring benefits from scale economies even if total population in each city is held constant.

Policy change: We look at the effects of small changes in intra- city commuting costs, \( dc_i \), and inter-city transport costs, \( dt \), and start with a single city. An expression for the consequent welfare change is readily found. Differentiating the equilibrium condition eqn. (3), city population changes by amount \( dL_i \) given by

\[
[W_L(L_i, t) - c_i / K]dL_i = (L_i / K)dc_i - W_i(L_i, t)dt. \tag{5}
\]

Differentiating rent, eqn. (4), using (5) to eliminate the change in \( L_i \) and rearranging gives the effect of transport improvement on rent and hence welfare (see appendix for derivation),

\[
dR_i = (1 + \mu_i) L_i W_i(L_i, t)dt - (1 + 2\mu_i) L_i^2 / 2K dc_i \tag{6}
\]

where \( \mu_i \equiv KW_L / (c_i - KW_L) \).

\[
(7)
\]
The term $\mu_i$ is the multiplier on the welfare effects of policy which captures the effects of increasing returns to scale, and is central to what follows. Its sign is that of $W_t(L_i, t)$, since the denominator of the expression must be positive for the equilibrium to be stable.\(^7\) If there are constant returns then $W_t(L_i, t) = 0$, so $\mu_t = 0$, and expression (6) just gives the user-benefits of transport improvements. This can be seen by noting that total distance commuted is $C_i = \int_0^z Kdz = K(z^*)^2 / 2 = L_i^2 / 2K$, so the second term on the right hand side of (6) is the reduction in commuting costs times distance commuted. The first term is the impact of external trade costs on the wage times the level of employment; in the next section we show that this is equal to change in trade costs times the value of inter-city trade.

If there are increasing returns to scale, $W_t(L_i, t) > 0$, then $\mu_i > 0$ implying that the multiplier increases the welfare gains. Intuition comes from Figure 1 which has distance from the CBD on the horizontal axis (along a single spoke of the city) and wages on the vertical. The line $w_0 + c_i L_i / K$ is the labour supply equation (eqn. 2) and $W$ is the wage equation (labour demand, eqn. 1); this is upward sloping because of increasing returns to scale, and assumed to be concave. Equilibrium is at the intersection of these lines and the initial equilibrium is point $b$.\(^8\) Rent is the triangle $wbw_0 = (c_i / 2)(L_i / K)^2$ (times $K$ in the full city); this is the value of output $w_i L_i$ minus real commuting costs and the opportunity cost of not employing labour outside the city ($w_0 L_i$).

A reduction in inter-city communication costs, $-\Delta t$, shifts $W(L_i, t)$ upwards, as indicated by the vertical arrow. Absent increasing returns, the new equilibrium would be point $f$ and the gains from the change are area $efbw$. However, with increasing returns the shift in the labour demand function gives new equilibrium at point $g$, with larger city population and real income gain $hgbw$. The multiplier is the ratio of these two areas, $1 + \mu_i = hgbw / ef bw$.

Lower intra-city communication costs, $c_i$, flatten the city-size relationship. This increases city size, with an additional effect if increasing returns mean that size raises productivity. Notice that changes in $t$ and $c_i$ enter eqn. (6) differently, with the multiplier on $c_i$ multiplied by 2. The reason for this is that reducing $t$ raises average productivity, but a change in $c_i$ is a change in the marginal commuting cost, with associated average change of half the size; essentially the former brings a rectangle (shifting the wage equation) while the latter yields a triangle (rotating the city-size line) while. The productivity gain from a given amount of induced city growth are equivalent in both cases, but it is the smaller gain under constant

---

\(^7\) This says that the slope of the labour supply curve exceeds that of labour demand, see Figure 1.

\(^8\) The lower intersection is unstable.
returns that gives the coefficient of two on the multiplier for commuting cost changes. This points to the fact that any application of these multipliers depends on details of the transport project being undertaken and requires detailed modelling of the exact changes being considered.\(^9\)

Figure 1: Gains from reducing inter-city communication costs.

3. **Inter-city linkages and urbanisation economies**

In order to look at inter-city linkages in greater detail we now develop a full model of production and inter-city trade, this giving labour demand and the wage equation (eqn. 1) an explicit functional form. As before, we suppose that the two cities, \(i = 1, 2\), can each produce a single final good that is perfectly tradable on a wider national or international market; this good is the numeraire.

Production of the final good uses intermediate goods (or tasks) that are produced in each city and are traded (at a cost) between cities. These tasks are differentiated, as in Dixit-Stiglitz, and our approach builds on Ethier (1982) in applying this framework to intermediate goods

\(^9\) An intra-city change that shifted labour supply downwards (such as removing a transport bottleneck close to the CBD) would be associated with multiplier entering (6) with coefficient of unity.
and Abdel-Rahman and Fujita (1990) who apply it in the urban context. Tasks can be thought of as specialised parts or components and as the business services that are needed for production. We initially suppose that there is a wide range of these tasks, all of which are symmetric yet differentiated from each other. The number of tasks produced in city $i$ is $n_i$, each with price $p_i$. Tasks combine to produce final goods through a CES production function with elasticity of substitution $\sigma > 1$. The ‘assembly’ of these tasks into final output uses no other inputs. Inter-city task trade is costly, and this cost is expressed as ad valorem factor $t$ ($t \geq 1$, with $t = 1$ perfectly costless trade). With this technology the unit cost of the final good in city $i$, denoted $G_i$, takes the form

$$G_i = \left[ n_1 \left( p_1 \right)^{-\sigma} + n_2 \left( p_2 \right)^{-\sigma} \right]^{\frac{1}{1-\sigma}}.$$  

(8)

Each task is produced by labour alone with productivity $a_i$, potentially varying across cities. The price of a task produced in city $i$ is then

$$p_i = \frac{\theta v_i}{a_i} , \quad p_2 = \frac{\theta v_2}{a_2}.$$  

(9)

In some of the cases that we look at price is marked up over marginal cost, in which case $\theta > 1$; otherwise $\theta = 1$. We focus on the case in which both cities produce the final good at unit cost equal to the world price, i.e. $G_1 = G_2 = 1$. Using (9) in (8), with $G_1 = G_2 = 1$, gives

$$1 = n_1 \left( \frac{\theta v_1}{a_1} \right)^{-\sigma} + n_2 \left( \frac{\theta v_2}{a_2} \right)^{-\sigma} , \quad 1 = n_1 \left( \frac{\theta v_1}{a_1} \right)^{-\sigma} + n_2 \left( \frac{\theta v_2}{a_2} \right)^{-\sigma}.$$  

(10)

These can be solved to give explicit expressions for wages in each city,

$$w_1 = \left[ p_1 \left( 1 + t^{1-\sigma} \right) \right]^{\frac{\sigma-1}{\sigma}} a_1 / \theta , \quad w_2 = \left[ p_2 \left( 1 + t^{1-\sigma} \right) \right]^{\frac{\sigma-1}{\sigma}} a_2 / \theta.$$  

(11)

These equations give the level of wages at which final good production breaks even, so are inverse labour demand curves analogous to eqn. (1). They contain $t$ directly and dependence on $L$ can arise through several different mechanisms.

The first is increasing returns in total city production, so productivity depends on employment, $a_i = a(L_i)$, with elasticity $\alpha \geq 0$. The second is increasing returns at the level of

---

10 See Duranton and Puga (2004) for discussion of this approach and Fujita et al. (1999) for the role played by tradability of these goods.

11 Extending this to include inputs of labour is conceptually straightforward but algebraically cumbersome.

the firm combined with monopolistic competition in which case \( n_i \) is endogenous and depends on \( L_i \) and, in the standard model, \( n_i \) is proportional to \( L_i \). Thus, following Dixit and Stiglitz (1977), each intermediate variety or task is produced by a single firm that has fixed input requirement \( F \). The elasticity of demand for each variety is \( \sigma \), so the mark-up of price over marginal cost is \( \theta = \sigma / (\sigma - 1) \). Selecting units of measurement for firms such that \( F = 1/(\sigma - 1) \), firms break even when they produce one unit of output.

This means that the number of tasks produced in each city, \( n_i \), must satisfy

\[
n_i p_i = w_i L_i, \quad n_2 p_2 = w_2 L_2.
\]

The left hand side is the total value of production in each city (at scale one per firm) and the right hand side is total costs. Since prices are proportional to wages it follows that the number of tasks is proportional to employment, \( n_i = L_i a_i (L_i) / \theta \), (using (9) in (12)), so that eqn. (11) becomes

\[
w_i = W(L_i, t) = L_i \left(1 + t^{1-\sigma} \right) \left( a_i (L_i) / \theta \right)^{1/(\sigma - 1)}, \quad i = 1, 2.
\]

This is the explicit form of the wage equation, eqn. (1), giving the inverse labour demand curve from the production side of the economy.

Expressions for the elasticity of wages with respect to inter-city trade costs, \( t W_i / W \), and with respect to city employment, \( L W_i / W \), follow from this. Differentiating (13) the elasticity of wages with respect to \( t \) is \( t W_i / W = -t^{1-\sigma} / (1 + t^{1-\sigma}) \), and we show in the appendix that this is equal to (minus) the share of total task production that is traded between cities, denoted \( s(t) \).

Welfare change, eqn. (6) is therefore

\[
dR_i = (1 + \mu_i) s(t) L_i w_i ( -dt/t ) + (1 + 2 \mu_i) C_i ( -dc_i ).
\]

The first term is therefore the proportionate reduction in trade costs, \(-dt/t\), multiplied by the value of trade, \( s(t) L_i w_i \), and \( (1 + \mu_i) \). With constant returns, \( \mu_i = 0 \), this is simply the user benefit; the transport improvement increases trade but no wider benefits are attached unless there are increasing returns to scale and a positive value of \( \mu_i \).

---

13 Marginal cost is \( w_i / a_i \) and operating profit per unit output is \( w_i / a_i (\sigma - 1) \). The fixed input requirement is produced by labour so has cost \( w_i F / a_i \). Firms therefore break even at unit scale if \( F = 1/(\sigma - 1) \). Kanemoto (2013) looks at a wider class of monopolistic competition models.

14 As is apparent, this equals zero for infinite \( t \), and \( 1/2 \) for \( t = 1 \).

15 Several studies have suggested additional gains from trade expansion effects of transport.
The elasticity of wages with respect to employment is, differentiating (13) with respect to $L_i$, \[ LW_i/W = (1 + \alpha \sigma)/(\sigma - 1). \] Technological increasing returns ($\alpha > 0$) and monopolistic competition with product differentiation ($\infty > \sigma > 1$) combine to create increasing returns, as expected from earlier work such as Abdel-Rahman and Fujita (1990). Using this elasticity the multiplier, eqn. (7), can be expressed as

$$
\mu_i = \frac{1 + \alpha \sigma}{(\sigma - 1) \Delta w_i - (1 + \alpha \sigma)} \quad (7')
$$

where $\Delta w_i = (w_i - w_0) / \mu_i = L_i c_i / Kw_i$ is the urban to non-urban nominal wage differential.

Several points follow. As noted above, the multiplier is positive only if there are increasing returns, i.e. \[ LW_i/W = (1 + \alpha \sigma)/(\sigma - 1) > 0, \] and is larger the greater are economies of scale, $\alpha$, the greater are the variety gains from introducing new tasks (lower $\sigma$), and the smaller the city’s nominal wage premium, $\Delta w_i$.

Eqns. (7) and (7’) provide a basis for indicative estimates of the magnitude of the multiplier. For example, if \[ LW_i/W = 0.05 \] and $\Delta w_i = 0.2$ then $\mu_i = 0.33$. \[ LW_i/W = 0.05 \] means that doubling city size raises wages by 5%, at the lower end of the range of 3-8% suggested in the survey by Rosenthal and Strange (2004).\(^\text{16}\) The more recent literature reviewed by Combes and Gobillon (2015) suggests a lower elasticity, perhaps around \[ LW_i/W = 0.025 \] in which case (with $\Delta w_i = 0.2$) the multiplier drops to $\mu_i = 0.14$. As we have seen, the wage elasticity can be generated either by direct technological effects or monopolistic competition; if

- $\sigma \to \infty$ then \[ LW_i/W = \alpha, \] while if $\alpha = 0$ then \[ LW_i/W = 1/(\sigma - 1). \]

In the latter case \[ LW_i/W = 0.05 \] requires $\sigma = 21$, and \[ LW_i/W = 0.025 \] requires $\sigma = 41$. These numbers are at the very high end of the range of values of $\sigma$ estimated from demand systems (e.g. Broda and Weinstein 2006), and more representative (and lower) estimates of $\sigma$ would imply higher values of the multiplier.\(^\text{17}\)

---

\(^{16}\) Ahlfeldt et al. (2012) find $\alpha \approx 0.13$ in a spatially disaggregated model of Berlin. Graham (2007) finds an elasticity of 0.07 for UK manufacturing and a much higher elasticity of 0.197 for services.

\(^{17}\) For example, if $\alpha = 0.0, \sigma = 6, \text{and } \Delta w_i = 0.3 \text{ then } \mu_i = 2.$
4. Task specialisation and localisation economies.

Economies of scale may arise at the city-wide level (urbanisation economies, as in the preceding section) and at the city-task level. These are localisation economies, limited in sectoral as well as spatial range, and thereby creating city-task patterns of comparative advantage. For present purposes, this means that an additional mechanism comes into play. Better communications facilitate city specialisation, which in turn changes the scale of production of each task in each city. As cities come to specialise in a subset of tasks so economies of scale are achieved, creating or deepening comparative advantage. This is a further source of gains from transport and communications improvement and the effect may be large; in the simplest two-city case specialisation doubles the scale at which a city performs those tasks in which it specialises.

Analysis of this requires minor modification of the production side of the economy. So far we have assumed that a particular task is produced in just one city, for use in that city and export to the other. We now assume that each task can potentially be produced in both cities, and the pattern of specialisation – which tasks are produced where – depends on city-task productivity and the level of inter-city trade costs, \( t \). The simplest case is that in which there are just two tasks, \( A \) and \( B \), both produced under perfect competition and sold at marginal cost. The unit production costs of each task in city \( i \) are \( w_i / a_i^A \), \( w_i / a_i^B \), these now carrying task superscripts as well as city subscripts. If tasks are sourced from the lowest cost city then the unit costs of producing final output in each city are

\[
G_1 = \left[ \left( \min \left[ w_1 / a_1^A, w_2 / a_2^A \right] \right)^{-\sigma} + \left( \min \left[ w_1 / a_1^B, w_2 / a_2^B \right] \right)^{-\sigma} \right]^{1/(1-\sigma)},
\]

\[
G_2 = \left[ \left( \min \left[ w_2 / a_1^A, w_2 / a_2^A \right] \right)^{-\sigma} + \left( \min \left[ w_1 / a_1^B, w_2 / a_2^B \right] \right)^{-\sigma} \right]^{1/(1-\sigma)}. \tag{14}
\]

Thus, task \( A \) produced in city 1 is a perfect substitute for task \( A \) produced in city 2 (similarly task \( B \)) and each task is sourced according to its lowest cost of supply inclusive of transport cost factor \( t \). Analogous to (8) and (10), the unit cost of final output equals the world price, so \( G_1 = G_2 = 1 \).

---

\(^{18}\) These might be better thought of as broad ‘functions’, such as finance, legal, advertising, engineering. Localisation economies are external to firms, each of which takes price and productivity as exogenous.
Evidently, if \( t \) is very high tasks are sourced locally and no task trade occurs. Both tasks are performed in both cities, so (14) takes the form
\[
G_i = \left( \frac{w_i}{a_i^A} \right)^{1-\sigma} + \left( \frac{w_i}{a_i^B} \right)^{1-\sigma} \right]^{1/(1-\sigma)} = 1
\]
for \( i = 1, 2 \), and the wage equations derived from this are
\[
w_1 = \left( a_1^A \right)^{\sigma-1} + \left( a_1^B \right)^{\sigma-1}, \quad w_2 = \left( a_2^A \right)^{\sigma-1} + \left( a_2^B \right)^{\sigma-1}.
\]
(15)
Wages in each city depend simply on the city’s productivity in both tasks.

At sufficiently low \( t \) trade takes place and specialisation occurs. Tasks are labelled such that city 1 has a comparative advantage in task \( A \) and city 2 in \( B \), in which case city 1 will export task \( A \) to city 2 if
\[
t w_i / a_i^A \leq w_2 / a_2^A,
\]
(16)
(and conversely for task \( B \), see eqn. (14)). When cities are specialised it follows from (14) that wage equations, analogous to (13), are:
\[
w_1 = \left( 1 + t^{1-\sigma} \right)^{\sigma/(\sigma-1)} a_1^A, \quad w_2 = \left( 1 + t^{1-\sigma} \right)^{\sigma/(\sigma-1)} a_2^B.
\]
(17)
Thus, wages depend on productivity in the task in which each city has comparative advantage and on inter-city communication costs incurred on trade in tasks.

**Ricardian comparative advantage:**

We look first at the case in which city-task productivities are exogenous (Ricardian comparative advantage), and then endogenise these with localisation economies. In the Ricardian case \( a_i^A, a_i^B \) are constants and, assuming symmetry and city 1 comparative advantage in \( A \), they satisfy \( a_1^A = a_2^B > a_2^A = a_1^B \). Symmetry means that wages are the same in both cities so trade and specialisation occur if and only if \( t \) is less than a critical value denoted \( t_b \), which from (16) is
\[
t < t_b \equiv a_1^A / a_2^A = a_2^B / a_1^B.
\]
Concentrating on this symmetric case, changes in \( t \) above \( t_b \) have no effect on trade or welfare; comparative advantage differences are not large enough for any trade to occur. As \( t \) crosses this value specialisation occurs and the value of trade jumps up to share
\[
t^{1-\sigma} / \left( 1 + t^{1-\sigma} \right) \]
of the value of output, then increasing continuously as \( t \) falls. Wages do not

---

19 With specialisation eqns. (14) are \( \left( w_i / a_i^A \right)^{1-\sigma} + \left( w_2 / a_2^B \right)^{1-\sigma} = 1 \), \( \left( w_1 / a_1^A \right)^{1-\sigma} + \left( w_2 / a_2^B \right)^{1-\sigma} = 1 \), and eqns. (17) come from solving this pair of equation for wages.
jump but move continuously, as can be seen by comparison of the wage equations without and with specialisation, i.e. (15) and (17) evaluated at \( t = t_b = a_1^A / a_2^A = a_1^B / a_1^A \). City size and total urban rents \( R \) are also continuous. Figure 2 illustrates rents as a function of inter-city transport costs, and is constructed with \( t_b = 1.1 \). Point \( t_0 \) gives an initial position and reducing \( t \) from this point (in the Ricardian case along the lower curve, \( \alpha = 0 \)), rent follows the path \( t_0 \rightarrow b \rightarrow d \). There is no effect until \( t = t_b = 1.1 \) (point \( b \)) at which it becomes profitable to specialise and trade. At this point trade jumps upwards, and real income follows the dashed – and continuous – line towards \( d \).

Furthermore, standard CBA gives an accurate measure of the real income change created by better communication. As is apparent from equations (15) and (17), the wage equation does not depend on scale (i.e. \( W_L \) and \( \mu \) are zero). There are gains from city specialisation and trade but, since there are no external effects or other market imperfections, these gains do not extend beyond those captured by standard CBA.

**Figure 2: Gains from task specialisation**

![Diagram showing gains from task specialisation]

**Localisation economies and endogenous comparative advantage:**

We now add localisation economies, so that productivity differences and comparative advantage become endogenous, shaped by the pattern of city specialization. These can be
incorporated by expressing labour productivity in each city-task as $\alpha_i = \bar{a}_i^{i} a(L_i^i), i = 1, 2, J = A, B$, where $\bar{a}_i$ is a constant, capturing any Ricardian productivity differences, and $a(L_i^i)$ gives returns to scale as a function of city-task employment. This is analogous to the modelling of increasing returns in section 3, but with scale effects driven by city-task employment, $L_i^i$, rather than by overall city employment.

Once again, at high $t$ it is efficient to produce both tasks in both cities and not trade tasks. At sufficiently low $t$ trade will occur and there will be complete specialisation so, with the assumed assignment of tasks to cities, $L_i^A = L_1, L_i^B = L_2 = 0, L_i^B = L_2$. The wage equations (17) are therefore

$$w_i = \left(1 + t^{1-\sigma} \right) \bar{a}_i^A a(L_i), \quad w_2 = \left(1 + t^{1-\sigma} \right) \bar{a}_2^B a(L_2)$$

The efficient path is illustrated on figure 2 as the solid line $t_0 \rightarrow s \rightarrow e \rightarrow f$, computed with $\alpha = 0.05$. Point $s$ is where transport costs are such that it is efficient for specialisation to occur, so employment in each sector goes from being $L_i/2$ to $L_i$. Trade commences, jumping up to a positive value, although wages, city employment and rents (and hence welfare) are continuous through this point. Conditions defining this critical value, $t_s$, are given in the appendix. Further reductions in $t$ beyond this point give continuously increasing city size and rents. These arise from the direct effect of lower $t$ and from localisation economies. The multiplier takes the form $\mu = 1 / [(\Delta w / \alpha) - 1]$, where $\alpha$ is the sector specific scale economy from (18).

While $t_0 \rightarrow s \rightarrow e \rightarrow f$ gives efficient outcomes as $t$ falls, these outcomes are not necessarily achieved by the choices of firms which fail to internalise the externality generated by localisation economies. As a consequence, there is a range of values of $t$ at which both specialised and non-specialised outcomes are equilibria. The intuitive reason is the following. If there is specialization then economies of scale are being achieved so there are wide differences in costs, with city 1 having comparative advantage (in $A$) and city 2 (in $B$); given these cost differences trade takes place, even if $t$ is quite large. Alternatively, if there is non-specialisation then costs are quite similar as both cities divide their labour force between both tasks; even relatively small values of $t$ are then sufficient to prevent trade, thereby confirming the non-specialised outcome.

---

20 It must be the case that $t_s > t_b$, as (on the efficient path) specialisation occurs when the gains from specialisation (now including realisation of scale economies as well as Ricardian advantages) come to equal the costs, $t$ per unit trade.
The range of values of \( t \) in which both these outcomes are an equilibrium are, on the figure, \( t \in [t_b, t_s] \). The lower edge of the range is the point at which Ricardian productivity differences are large enough to cause trade to occur (\( t_b = \frac{\bar{a}_i^A}{\bar{a}_i^S} \)), below which specialisation is the only equilibrium; the upper edge is the point at which specialisation is efficient (see appendix for derivation of \( t_b \) and \( t_s \)). Thus, an equilibrium path is \( \text{to} \rightarrow \text{b} \rightarrow \text{e} \rightarrow \text{f} \). In the interval \( t \in [t_b, t_s] \) it would be efficient for both cities to specialise but, given non-specialisation, it is cost minimising for final goods producers to source tasks locally, implying that cities remain non-specialised. Only when transport costs fall to \( t_b \) are Ricardian productivity differences large enough to create trade and specialisation; as this occurs the efficiency gains from specialisation and localisation economies are reaped by a discontinuously large increase in rents and in welfare, the jump from \( b \) to \( e \).

The theoretical possibility of multiple equilibria accords well with the often heard practical idea that transport improvement is necessary but not sufficient for the full gains of better communication to be achieved. Improvements that leave costs in the interval \( t \in [t_b, t_s] \) will not achieve gains unless there is some coordinated action to promote specialisation, i.e. to build the cluster of task producers in each city. Coordination may be achieved by public action (e.g. to promote development in light of new possibilities brought by the transport improvement) or private, such as investment by a ‘large agent’ to redevelop the area around a rail station or airport.

This can be summarised as follows:

1) If the economy is at the equilibrium with specialisation and intra-city task trade, then reductions in \( t \) bring gains evaluated with the multiplier, \( \mu_i = \frac{1}{[\Delta w_i / \alpha] - 1} \).

2) If the economy is at the equilibrium with diversified cities and no intra-city task trade then reductions in \( t \) may trigger specialisation and trade, bringing a jump in welfare.

3) A reduction in \( t \) is not a sufficient condition for this to occur as coordination failure means that firms fail to move to the new equilibrium.

One further important point can be made. In section 3 gains from economies of scale depend on growth of the city as a whole, this drawing in labour from elsewhere. However, with localisation economies gains come also from task specialisation as each city expands the

\[21\] The ‘break point’ and ‘sustain point’ in the terminology of Fujita et al. (1999).

\[22\] The coordination failure arises as each firm takes productivity as constant, failing to internalise the productivity benefits that would arise from coordinated activity.
sector in which it specialises. This generates gains additional to those from standard CBA, even if city population were to remain constant.

Figure 3 illustrates (parameters given in the appendix). The lower two curves in the figure give rent as a function of inter-city costs, $t$, when city size is fixed. Rent remains the welfare indicator if city size is fixed because of constrained land-area and land-use. The lowest curve gives rent with constant city size and constant returns to scale. Specialisation occurs as $t$ is reduced, but standard CBA techniques apply; because there are no imperfections specialisation occurs at the point where the productivity gain from specialisation is exactly offset by the transport costs incurred. Adding localisation economies (middle curve, $\alpha > 0$) means that the gains from specialisation are not fully internalised by private decision takers so that, exactly as in figure 2, there is a jump in welfare as specialization occurs. The top curve gives, for comparison, the case with localisation economies and a perfectly elastic supply of labour and land to the city. There are gains both at the point where specialisation occurs, and from overall employment growth as $t$ falls further.

**Figure 3: Gains from task specialisation with fixed city size**

$R_t: \alpha = 0$

$R_t: \alpha > 0$

Size fixed

Size endogenous

---

23 Identical results hold if city size is constrained not by fixed land but by fixed labour supply, although the welfare indicator is then land rent plus excess wage income arising as real wages are bid up above the outside wage, $w_0$. 

16
**Urban asymmetries: convergence or divergence?**

Finally, we relax the assumption that the two cities are symmetric in order to address the question, do inter-city transport improvements cause divergence or convergence of city size and hence land rents and real incomes?\(^{24}\) Asymmetry is captured by assuming that city 1 has lower commuting costs per unit distance so, other things equal, is larger than city 2 and has higher total rent, \(R_1 > R_2\). The equilibrium path of rents as \(t\) is reduced is illustrated in Figure 4, based on a numerical example (see appendix). There are three phases.

At high \(t\) both cities produce both tasks. Rents in city 1 are higher because of its assumed lower commuting costs.\(^{25}\) At sufficiently low \(t\) both cities are fully specialised, and both gain from reductions in \(t\), with rent equalisation in the limit. In the intermediate range (between \(t^*\) and \(t^{**}\)) the larger city, 1, contains all production of task A and part of production of task B, while city 2 just has the remaining production of task B. Essentially city 1’s size advantage creates break point at \(t^{**}\), at which it captures all of task A and is large enough to retain some of task B. Evidently, there is discontinuous divergence at this point, with city 2 contracting and experiencing a decline in rents and nominal wages. Further reductions in \(t\) in the interval \(t^*, t^{**}\) cause a process of convergence with city 2’s share of task B increasing in response to its lower nominal wage. This process is continuous, with full specialisation of both cities occurring at \(t^*\).

This pattern of divergence followed by convergence during a process of reductions in trade barriers is one seen in other models in which there is an interaction between increasing returns to scale and specialisation (e.g. Krugman and Venables 1995). Essentially, increasing returns allow small advantages to become magnified creating divergence, but at very low trade costs free trade spreads the benefits of scale evenly across locations. There are, of course, many potential sources of asymmetry, of tasks as well as of cities, so these results are only illustrative of possibilities. However, there is an important message. Better communication will bring structural change to each city. In asymmetric cases there may be gainers and losers at various stages as production relocates, but the benefits of higher productivity are ultimately felt in all locations.

---

\(^{24}\) Real wages are equalised by the assumption of perfect labour mobility, but land rents and nominal wages are higher in the larger city, see eqn. (4).

\(^{25}\) Rents are not independent of \(t\) because, for ease of computing, the elasticity of substitution between tasks of the same type produced in different cities is assume to be large, but finite, see appendix.
Large transport projects such as high-speed rail networks or expansions of airport capacity have non-marginal effects on economic geography and the structure of activity in cities and regions that are affected. These effects are inevitably hard to predict, although there is an important and fast growing empirical literature assessing impacts (e.g. Duranton and Turner 2012 and the survey by Redding and Turner 2015). For purposes of project appraisal predicted effects have to be valued, and it is important that this is done in a rigorous micro-founded way. Such a framework is needed both to discipline some of the more exaggerated claims that are made for transport improvements and to establish and quantify circumstances in which ‘wider benefits’ are likely to be generated.

This paper develops such a framework and shows how inter-city, as well as intra-city, projects can generate these wider benefits. For inter-city improvements in particular, the paper captures the idea that transport improves business links as firms are better able to draw on intermediate goods and services (such as legal services, finance, design, and advertising) supplied from other cities. This is trade creating, but is not a source of additional benefits (beyond those captured by standard CBA) unless trade interacts with economies of scale, in which case the gains from expanding production can be captured by a multiplier applied to the user-benefits of transport improvement. An important distinction arises between urbanisation and localisation economies. The latter mean that productivity levels are city-task specific and that transport improvements can promote city specialisation. This can give
large (and discontinuous) welfare gains from transport cost reductions, although coordination failure means that such reductions are a necessary, but not sufficient condition for securing the spatial reorganisation of production that delivers these gains.

Needless to say, the framework developed on the paper is built on numerous assumptions which need to be relaxed in application. In particular, the analysis is based around two monocentric-cities located in a perfectly competitive rest of the economy. A richer geography in which cities could be polycentric and in which transport improvements between some locations could displace activity from others would give a richer menu of policy choices and outcomes. This is beyond the scope of analytical work of the type presented in this paper and requires larger scale modelling. However, the ideas and techniques developed in this paper can be used to inform such future work.
Appendix:

Section 2: Differentiating (4) and using (5) to eliminate $dL_i$ gives

$$dR_i = \left( \frac{L_i^2}{2K} \right) dc_i + \left( \frac{c_i L_i}{K} \right) dL_i = \frac{1}{KW_L/c_i - 1} \left[ \frac{L_i^2}{2K} \left( \frac{KW_L}{c_i} + 1 \right) dc_i - L_i W_i dt \right].$$

Eqn. (6) follows using the definition of $\mu_i$, eqn. (7).

Section 3: Eqn. (9) with $G_1 = G_2 = 1$ implies $n_1 p_1^{-1} = n_2 p_2^{-1} = 1/(1 + t^{-1})$. The gross value of final output produced in city $i$ is $w_i L_i$ and city 1 expenditure on tasks supplied by cities 1 and 2 are respectively $n_1 p_1^{-1} w_i L_1$ and $n_2 p_2^{-1} w_i L_2$. The value of imports relative to the value of output is therefore $n_2 (p_2)^{-1} = t^{-1} (1 + t^{-1})$, and similarly in country 2.

Section 4: We assume symmetry throughout, so $a_1^A = a_2^B > a_1^A = a_2^B$. Each city has labour force $L$. If there is no specialisation, $L/2$ is employed in each task. With specialisation, $L$ is employed in task A (city 1) or task B (city 2). Equilibrium wages are the same in each city, $w_1 = w_2$.

Efficiency: Point $s$ is where wages (and hence employment and rents) are the same with and without specialisation.

Without specialisation (from 16):

$$1 = (w_1)^{-1} \left\{ \frac{a}{\bar{a}^A} a(L/2) \right\}^{\sigma-1} + \left( \frac{a}{\bar{a}^A} a(L/2) \right)^{\sigma-1} = (w_1)^{-1} \left( \frac{a}{\bar{a}^A} a(L/2) \right)^{\sigma-1} + 1 + \left( \frac{a}{\bar{a}^A} \right)^{\sigma-1}.$$

With specialisation (from 17): $1 = (w_1)^{-1} \left( \frac{a}{\bar{a}^A} a(L) \right)^{\sigma-1} [1 + t^{1-\sigma}]$.

The critical value $t_x$ is that at which both equations give the same value of $w_1$.

Break point: Non-specialisation is an equilibrium if $t > \bar{a}^A a(L/2)/\bar{a}^A a(L/2)$, i.e. it is not profitable for city 1 to export task A to city 2. At the symmetric equilibrium this reduces to $t > \bar{a}^A/\bar{a}^A$, and the ‘break point’ is $t_b = \bar{a}^A/\bar{a}^A$. $t_b > 1$ only there is some exogenous Ricardian comparative advantage, $\bar{a}^A/\bar{a}^A > 1$.

Sustain point: Specialisation is ‘sustainable’ as an equilibrium if $t < \bar{a}^A a(L)/\bar{a}^A a(0)$, i.e. it is profitable for city 1 to export task A to city 2, given that productivity is evaluated at the specialised employment levels. The ‘sustain point’ is $t_s = \bar{a}^A a(L)/\bar{a}^A a(0)$. With increasing returns $a(L)$ is increasing in $L$, so $t_s > t_b$, giving an interval in which both outcomes are equilibria.

Parameters for simulation:
Figure 2: $\sigma = 2; \alpha = \{0, 0.07\}; c_1 = 0.5; c_2 = 0.5; w_0 = 0.9; \bar{a}_1^A = \bar{a}_2^B = 1.44, \bar{a}_1^B = \bar{a}_2^A = 1.58$.

Figure 3.4: For ease of computation, task $J$ produced in cities 1, 2, are assumed to be highly, but not perfectly substitutable. Thus, expressions $\min\left[w_1/a_1^J, tw_2/a_2^J\right]$ in eqns. (14) are replaced by a CES cost function with elasticity of substitution equal to 20. Other parameters are $\sigma = 1.5; \alpha = 0.1; c_1 = 0.5; c_2 = 0.5; w_0 = 0.9; \bar{a}_1^A = \bar{a}_2^B = 1.44; \bar{a}_1^B = \bar{a}_2^A = 1.58$.

In figure 4, $c_2 = 0.55$.

References


