The wider economic benefits of transport infrastructure

Peter Abelson
Applied Economics P/L
pabelson@appliedeconomics.com.au

Abstract
The standard evaluation of transport infrastructure focuses on transport user benefits. Wider economic benefits (WEBs) are claims for additional economic benefits. This paper reviews four main forms of WEB: agglomeration economies, the value of additional output in imperfectly competitive markets, labour supply effects, and induced commercial and residential developments. The paper finds that the wider economic benefits in each case are generally small or non-existent. The main potential exception is induced development. Where a claim for a substantial WEB is made, it needs to be supported by a reasoned narrative.

1 Introduction
In the 2018 Australian budget, the Government planned to invest over $75 billion in transport infrastructure over the next 10 years. State expenditures will likely double that amount. It is especially important therefore that these investment decisions be made efficiently and transparently. In this paper we discuss “wider economic benefits” (WEBs). Whereas the standard cost-benefit analysis (CBA) of transport infrastructure focuses on transport user benefits, WEBs are claimed additional economic benefits.

In a standard CBA, following Mackie et al. (2005), the net social benefit (NSB) of investment in transport infrastructure is given by:

\[
\text{Overall Economic Impact} = \text{Change in transport user benefits (Consumer Surplus)} + \text{Change in system operating costs and revenues (Producer Surplus and Government impacts)} + \text{Change in costs of externalities (Environmental costs, accidents, etc.)} - \text{Investment costs (including mitigation measures)}
\]
User benefits are estimated across the network and include existing and new trips. These benefits are principally savings in travel time and vehicle operating costs and often include reliability benefits. Externalities are principally safety and environmental impacts such as air quality and noise. Some externalities may be negative.

Over time, four main WEBs have been identified.

1. Agglomeration economies: the impacts of increased employment density on productivity
2. The value of additional output in imperfectly competitive markets
3. Impacts of lower trip costs on labour supply and productivity
4. Benefits from induced commercial and residential development.

The first three WEBs were identified by the UK Department for Transport (UK DfT, 2005) and have been the focus of most discussion since then. WEB 4, induced developments, were discussed as possible benefits in transport literature in the 1980’s and 1990’s. After dropping out of the literature, they have recently become live issues. A feature of WEBs is that they are usually associated with some form of market failure.

Estimated WEBs are sometimes large. In the London Cross Rail study, they added over 50% to the estimated standard set of benefits (UK DfT, 2005). In some NZ and Australian projects, they have added 30 to 50% to transport benefits (Douglas and O’Keefe, 2016). In some cases, they may be decisive in producing a positive net present value. They are often also contentious.

In this paper, each WEB is discussed in turn. The final section provides conclusions.¹

## 2 Agglomeration economies in transport

### 2.1 Introducing agglomeration economies

Agglomeration economies may be dynamic or static. Dynamic economies occur when productivity rises with greater urban employment. Static economies occur if productivity rises with “effective employment density”. This latter concept has been the major claim made for transport projects.

Many studies have found dynamic economies. In their major survey of agglomeration economies, Rosenthal and Strange (2004) found that doubling city employment increases productivity by between 3 and 8 per cent. Using meta-analysis, Melo et al. (2009) found an elasticity at the lower end of 0.03. The reasoning is based on scale effects: firms derive productive advantages from greater access to suppliers, labour and information. However, as discussed below, there are various explanations of these differences in productivity. Also, most studies of agglomeration economies are based on comparisons of metropolitan areas. Rosenthal and Strange (ibid) found little

---

¹ Abelson (2019) provides a fuller discussion of these issues.
research on localisation economies: where output increases with employment in the relevant industry in a city.

Graham (2005, 2006) and UK DfT (2005) introduced “effective density”. Effective density is a weighted sum of employment in a designated area and neighbouring areas. The latter has a lower weight as a function either of distance or of generalised trip costs (GTC) between the areas. Using cross-section analysis, Graham (2005, 2006, 2007) found that productivity rises with effective density based on a distance decay factor. Proponents of static agglomeration then argue that reductions in GTC increase interactions between areas and that this increases productivity, without any changes in employment levels or densities. In practice, reflecting relative ease of modelling, nearly all applications of agglomeration economies in transport studies are based on effective density.

The following section notes some general issues in agglomeration economies. Section 2.3 discusses the critical concept of effective density. Section 2.4 briefly discusses two major studies of effective density. Section 2.5 concludes.

2.2 General Issues in agglomeration economies

The general approach to estimating agglomeration economies is to estimate firm revenue within an industry as a function of inputs (labour, capital and other purchased inputs) and area employment:

$$\ln R_{ijn} = \beta_0 + \beta_1 \ln L_i + \beta_2 \ln K_i + \beta_3 OPI_i + \beta_4 \ln E_{jn} \quad (1)$$

where $R_{ijn}$ = revenue per firm $i$ in industry $j$ in area $n$, $L_i$ and $K_i$ are labour and capital inputs employed by firm $i$, $OPI_i$ is other purchased inputs, and $E_{jn}$ is employment in industry $j$ in area $n$. $\beta_4$ represents the estimated % increase in revenue for a 1% increase in employment in the industry. Sometimes wages per worker is the dependent variable, for example Hensher et al. (2012). Employment may be total employment or employment density (employment / size of area). It may also be effective density which includes employment in neighbouring areas discounted for distance or trip costs. The latter approach is common in transport studies, see Graham (2007), Mare and Graham (2009), Combes et al. (2010), Hensher et al. (2012) and Australian Bureau of Statistics (2017).

Several issues should be noted. First, the size of the geographical units is arbitrary. A large area may have high total employment but low employment density whereas a small area may have low total employment but high employment density. Theory does not tell us which is more important: employment over a large area or density in a small area.

Second, valuing output in dollar terms is problematic where prices of outputs vary. Prices are generally higher in large cities where they compensate workers for the higher costs of commuting and congestion (Glaeser, 2010). Also, wages fall with distance to the CBD, known as the “urban wage gradient” (Miles and Cheshire, 1986).
Unless revenues are adjusted for price differentials, estimated productivity differentials are biased. For example, in Sydney, petrol prices are typically 15% lower in outer suburbs than in inner suburbs. But this surely does not mean that petrol station workers are less productive further from the CBD!

Third, productivity has many causes. Cause and effect must be sorted, including natural advantages. Many cities grew up around ports, government centres and high amenity areas. These centres attract population and employment. As Glaeser (2010, pp.13-14) observed: “Productivity certainly attracts population...population density is not exogenous”. More productive workers may sort into denser areas. It cannot be assumed that labour is equally skilled in all centres or that all jobs even within an industry sector (such as banking or legal services) are similar across the urban area.

Finally, the relationship between transport and employment needs to be “explicitly modelled” (UK DfT, 2018d, p.10). Employment in one area may displace employment elsewhere. Indeed, transport infrastructure may decentralise employment.

### 2.3 Effective density

Formally, effective employment in area $j$ equals employment in area $j$ plus employment in adjacent areas ($k$) as a weighted function of distance (or generalised trip costs) between area $j$ and the other areas.

$$ED_j = E_j + \sum E_k D_{jk}^a$$  \hspace{1cm} (1)

where $E_j$ = employment in area $j$
$E_k$ = employment in neighbouring areas $k$
$D_{jk}$ = distance between area $j$ and areas $k$
$a$ = a decay parameter that reflects lower importance of employment further away.

Box 1 overleaf shows UK DFT (2006) advice on estimating the benefits from effective density.

Critically, effective density is almost always modelled as a function of distance between areas, which is more readily observable than trip costs.

Also, there is no theoretical basis for definition of an area or distance decay weighting. The default parameter value for $a$ in Equation (1) is -1.0, but this may be varied. Graham (2006) assumed a value of 1.0. Graham et al. (2009) found that 1.0 was appropriate for some sectors, but that higher decay factors of 1.6 was more appropriate for manufacturing and 1.8 for consumer and business service sectors.

---

Box 1 Estimating agglomeration benefits based on effective densities

UK DfT (2006) proposed that the agglomeration benefit for any area \((j)\) should be based on changes in estimated effective densities and calculated as

\[
WEB1_{j} = GDP_{j} \times \Delta ED \times EP
\]

where

- \(GDP_{j}\) = local economic output in area \(j\)
- \(\Delta ED\) = percentage change in effective employment density of the area,
- \(EP\) = the elasticity of total productivity with respect to effective employment density.

Summing agglomeration benefits over all areas and all industries,

\[
WEB1 = \sum_{ij} \left[ GDP_{ij} \times E_{ij} \times \left( EIP_{ij} \times \Delta ED_{j} / ED_{j} \right) \right]
\]

where

- \(GDP_{ij}\) = GDP per worker in industry \(i\) and area \(j\)
- \(E_{ij}\) = employment in industry \(i\) and area \(j\)
- \(EIP_{ij}\) = elasticity of productivity with respect to effective density of industry \(i\) in area \(j\)
- \(ED_{ij}\) = effective density of employment of industry \(i\) in area \(j\), and
- \(\Delta ED\) = change in effective density due to transport project.

Citing Graham (2006), DfT recommended values for \(EIP_{ij}\) from 0.04 to 0.11 depending on the industry.

On the other hand, estimated changes in effective density for transport evaluation purposes are based on changes in GTC between areas. Where GTC reflects business-to-business travel, this should reflect the main business modes and times of travel, principally in off-peak times. Critically, effective density rises without any changes in actual employment.

Is productivity likely to rise with effective density without any changes in employment locations? This question sets up three more questions. Do lower GTC significantly increase business travel between neighbouring areas? How should these new trips be valued? And, would these extra trips create agglomeration economies?

Most short-distance business trips are made by walking or taxi, not by train or bus (see study in New Zealand)\(^4\). New transport infrastructure usually reduces door-to-door GTC for such trips only marginally. Goodwin (1996) and Abelson and Hensher (2001) found that generated trips are small in relation to existing trips. Thus, lower GTC

---

\(^3\) The Australian Transport and Infrastructure Council (2016, p.6) noted that agglomeration benefits will occur only if there is a significant change in business travel between employment centres “because agglomeration benefits derive from business-to-business interaction”.

\(^4\) A large survey of bus and train users in Auckland, Christchurch and Wellington found that company business trips were only 1 per cent of total trips.

generally has little impact on short-distance business trips. Further, generated trips are usually of marginal business importance. Where lower GTC generate new business trips, these are valued in the standard evaluation process by the Rule of a Half as shown in the Annex. Thirdly, a small number of marginal business trips is unlikely to generate significant external agglomeration economies.

We conclude that agglomeration economies associated with generated business trips are likely to be exceptional and small.

### 2.4 Two studies of effective density

Despite the extensive discussion on effective density, there are few studies of the productivity effects. We comment briefly here on two major studies: by Graham (2007)\(^5\) and the Australian Bureau of Statistics (2017)\(^6\).

**Graham (2007)** estimated the relationship between firm revenue and effective density based on firms in designated industries in 8000 wards in the UK. Graham regressed firm revenue as a function of labour, capital inputs and effective density based on distance between areas with an assumed decay parameter \((\alpha) = -1\).

Graham (ibid.) reported high agglomeration elasticities, ranging from 0.07 for manufacturing to 0.197 for services, with an average urban elasticity of 0.129. However, after allowing for the heterogeneity of products within each industry, Graham et al. (2009) estimated lower, agglomeration elasticities averaging 0.04, ranging from 0.02 for manufacturing and consumer services to 0.08 for business services. Moreover, some of the estimated impact may reflect higher prices in denser employment areas.

Graham was aware of most of these issues. Graham (2006) noted:

- The concept of an area is arbitrary. Little research has been done into the effects of employment densities within cities.
- There is no firm basis for the distance decay parameter, the value of \((\alpha)\).
- In many industries, firms are heterogeneous.

**Australian Bureau of Statistics (2017) and KPMG (2017)**

The ABS (2017) estimated firm revenue as a function of labour, capital employed, intermediate inputs and effective density by industry in the eight Australian capital cities. All values other than effective density were estimated in dollar terms. In effect, all profits above a normal return on capital were assumed due to differences in effective density, and not to other factors such as intellectual capital. Other issues

---

\(^5\) Graham (2005, 2006 etc) produced several papers. They basically report the one major research study.

\(^6\) Abelson (2019) discusses Graham and Mare (2009), Combes et al. (201.0) and Hensher et al., (2012).
include: the arbitrary size of the areas, the arbitrary distance decay curve, the assumption of homogeneous firms within an industry, the high number of zeros in the data base, and scaling the coefficients in the production function to equal 1.0.

ABS (2017, Table 7) estimated many insignificant and indeed negative agglomeration effects. Of 152 estimates of agglomeration in 8 capital cities over 19 industry sectors, only 42% were positive and significant at the 90% confidence results. Notwithstanding, KPMG (2017, Table 3) recommended positive measures of agglomeration ranging up to 0.17 for 80% of industry sectors in all Australian cities, 20% with zero impacts and no negative economies. These recommendations are extraordinarily hard to reconcile with the analysis.

2.5 Conclusions

**Dynamic agglomeration economies:** There is evidence that productivity rises with total employment in an area. However, once factors such as the variety of services supplied within an industry, price variations between areas and differences in capital inputs are allowed for, the average elasticity of output to employment density appears to be in the order of 0.02 (doubling employment would increase output per worker by 2 per cent).

However, transport infrastructure may disperse, not raise, employment density. Thus, any suggestion of dynamic agglomeration economies must be supported by a narrative on employment density and modelling of land use and employment changes. The Australian Transport and Infrastructure Council (2016, p.6) concluded that WEBs may be negative and that “it is bad practice to apply a broad percentage up-lift to the results of the traditional appraisal”.

**Static agglomeration economies:** There is little evidence that lower GTCs, without changes in employment density, increase productivity. Transport infrastructure usually generates few new local business-to-business trips. Most such new trips are of marginal business importance and unlikely to generate significant agglomeration economies. As Douglas and O’Keefe (2016, p.12) observed static agglomeration “is invisible and largely unprovable”.

3 Value of additional output in imperfectly competitive markets

In the standard evaluation approach, markets are assumed to be competitive and workers paid the value of their marginal product. Output arising from business travel time savings is valued at the relevant wage rate plus direct overheads. However, where markets are imperfectly competitive, output prices may be set above marginal cost. This implies that the standard approach undervalues gains in output associated with business travel time savings. This point is generally accepted.
The uprate factor is the product of the price–marginal cost mark-up and the elasticity of demand. The UK (DfT, 2005) uprate factor of 0.1 reflected a price mark-up of 0.2 and an elasticity of 0.5 (ignoring the negative sign). KPMG (2017) recommended uprate factors for different Australian cities ranging from 5% to 25%.\(^7\)

However, this ignores two significant issues. First, much business travel time is spent productively (Hensher 2001; Wardman and Lyons (2015); Wardman et al., 2015). Second, some savings in business travel time may be converted into leisure time.

Accounting for these two effects, the average value of net output gained from business travel time savings could be below the average wage of business travellers rather than above as is commonly assumed (e.g. Transport for NSW, 2018).

4 Impacts on labour supply and productivity

UK DfT (2005) identified three potential labour supply effects due to lower GTC.

- Working longer hours in current occupations
- Increased participation in the workforce
- Moving to a more productive, higher paid, jobs.

As we will see, each of these benefits is picked up in the standard evaluation approach. The question is whether these valuations are reasonable.

**Working longer hours**

When a worker saves commuting time, the standard assumption is that she has a constant working week and will enjoy a preferred form of leisure to travel. Transport for NSW (2018) recommends that this preference is valued at 40% of the full time average weekly earnings for Australia.

Alternatively, someone may choose to work longer hours. Given a choice between extra leisure and work, workers are assumed to be indifferent at the margin between leisure and work. Taking on extra work increases after-tax wage income but foregoes leisure time. It follows that the value of the travel time saved is independent of whether the worker chooses improved leisure or extra work. However, increased earnings produce additional tax revenue, which is a social benefit, which is not counted in standard transport evaluations.

**Increased participation in the workforce**

In standard economic appraisals, the value of taking on work is derived from the “rule of a half”. Suppose GTC falls from $20 to $10 per return trip, or from $40,000 to $20,000 per annum. Following the rule of a half principle, the average benefit would be $5.00 per day or $10,000 per annum (see annex). Again, a tax benefit accrues to Government due to the additional work, which is a WEB.

\(^7\) In principle, a similar mark-up would apply to any other savings that are components of marginal costs.
There is also the issue of the extent of increased participation. In a modern city, most workers have several workplace options. DfT (2018c, p.9) suggests a low labour supply elasticity of 0.1. Thus, suppose average daily wage after tax is $250, GTC falls by $10 per day, and labour supply elasticity is 0.1. The net wage after transport costs would rise by 0.4 per cent and employment by 0.04 per cent.

**Moves to more productive jobs**

The valuation principles for moves to higher paid jobs are the same as those for entry to work. Without unusual barriers on access to jobs, the private benefit cannot exceed savings in GTC and the average private benefit is approximated by the “rule of a half”. The public benefit is again the extra tax revenue.

There appears to be little research on such employment moves. Absent a land use and transport interaction model to forecast employment relocation, the evaluation should provide an explicit explanatory narrative – not some arbitrary black box assumption.

5 **Induced commercial or residential development**

Induced developments may occur due to new transport infrastructure (Adler, 1987). In Australia, development of the major Adani coal mine in Northern Queensland appears to depend on construction of a major rail link. In Sydney, major benefits of residential development due to the proposed Sydney West Metro are claimed.8 In such cases, some economic surplus may be attributed to the infrastructure. This usually depends on some form of market failure, notably economies of scale.

**Induced commercial development**

Ferrari et al. (2019) cite several substantial commercial developments due to major transport infrastructure in Europe, China and India. Where transport infrastructure affects firm location and size, the issue is whether savings in transport costs represent all the benefits.

Absent changes in production costs, any new or relocated development reflects transport cost savings. For new trips, the rule of a half, described in the Annex, can be applied. As UK DfT (2018b) notes, in competitive markets, “user benefits will capture the entire welfare effects of a transport investment” including the benefit of induced development.

However, scale economies may occur, especially via access to larger markets. Venables et al. (2014) cite a retail development arising from economy of scale in a more populated area and show that the economic surplus is greater than under the rule of a half measure. Alternatively, multi-plant firms may operate from fewer sites.

---

8 The Australian Broadcasting Commission (14 August 2017) reported that the estimated BCR for the proposed metro with standard WEBs was 1.9 and that this rose to about 2.5 with high rise rezoning benefits included.
As shown in the Annex, increasing returns to scale may lower production costs as well as transport costs.

When production costs fall with economies of scale, the rule of a half valuation for new user benefits underestimates the producer surplus achieved. The true benefit is the sum of transport user benefits and savings in production costs. However, because economies of scale are case specific, the transport benefits and production savings need to be separately modelled in each case.

**Induced residential development**

Turning to residential development, there is an induced benefit only if the development depends on the transport investment. Also, care must be taken not to double count new user benefits and residential benefits that are capitalisation of these benefits.

Where there are scale economies in housing development or public provision of services, new housing may be purchased by households who do not use the new transport infrastructure and their benefits are not picked up as new user benefits. Scale economies create net social benefits when dwelling prices exceed all costs of development, including land, land development, any related public infrastructure costs, construction costs, marketing and finance costs. In this case, the net social benefit from new housing \( (NSB_{NH}) \) would be:

\[
NSB_{NH} = \sum (MP_{NH} - PRIVC - PUBC)
\]

(5)

where \( MP_{NH} \) denotes market prices of new housing and \( PRIVC \) and \( PUBC \) denote total private and public costs respectively of housing development. However, these benefits should be carefully described and estimated in a place making CBA.

Note that there is no reference here to displaced development. The assumption is that displaced development would be marginal development where the benefits, as reflected in market prices, approximately equalled the sum of private and public costs, so that there is no net social gain or loss associated with this displaced development.

### 6 Conclusions

The standard economic appraisal of transport infrastructure includes user benefits but may require marginal adjustments for additional economic benefits (WEBs) in a few cases. Claims of large WEBs are generally unjustified. When WEBs are claimed, an economic narrative and explanation is essential rather than applying “assumption laden black-box formulae as has increasingly been the norm” (Douglas and O'keefe, 2016, p.18).
More detailed conclusions are:

- Small agglomeration benefits may occur with increases in actual employment density. However, it needs to be demonstrated that the transport infrastructure will increase employment density.

- Changes in effective density due to lower trip costs are unlikely to have significant productivity effects without changes in actual employment densities.

- The value of output associated with travel time savings increases with imperfect competition, but this is likely offset by allowances for some productivity during work trips.

- Transport improvements may marginally increase labour supply or moves to more productive jobs. These benefits are captured by the rule of a half assessment in a standard evaluation method. However, there may be small additional benefits from increased tax revenue.

- Transport infrastructure can induce commercial and/or residential development and produce development surpluses where economies of scale are achieved. However, the link between transport and development needs to be demonstrated.

An excellent Norwegian report (Hagen, Chairperson, 2012) reached similar conclusions. This report was further supported by a review of international practice in 24 counties which found low use of most WEBs (Wangsness et al., 2016).

Fundamentally, any claims for WEBs should be carefully demonstrated in the context of any proposed new transport infrastructure. It is inappropriate to simply assume that a WEB exists.

Acknowledgements

I thank Professor Peter Mackie, Dr. Neil Douglas and Dr. Mark Harvey for detailed comments on a previous draft and Professors David Hensher and James Laird for some general advice and recommended references. I am fully responsible for the paper as presented.

References


Annex  Valuing user benefits and increased output

This annex outlines the basic method for estimating user benefits. The private generalised trip cost (GTC) is the sum of travel time and fares and other out-of-pocket costs, including taxes. The real social cost (RSC) excludes taxes or charges, such as road tolls, that do not reflect use of resources.

Figure A.1 shows the private GTC and the RSC for a given trip and mode before and after a transport improvement. There are \( Q_1 \) existing trips and \( Q_2 \) trips after the improvement. Post-improvement trips include trips diverted from other modes or routes and trips generated by the fall in GTC.

**Figure A.1  Benefits of existing, diverted and generated trips**

The benefits to existing trips are the savings in real social cost given by shaded area: \( Q_1 \times (RSC_1 - RSC_2) \).

Trip makers who divert to a new destination, route or mode are assumed to be willing to pay a price between GTC\(_1\) and GTC\(_2\). If the demand curve is linear, diverted trip makers would be willing to pay an average price of 0.5 \((GTC_1 + GTC_2)\). Thus, these benefits are often estimated as \(0.5(Q_2 - Q_1) \times (GTC_1 - GTC_2)\). This is known as the "Rule of a Half". Where GTC\(_2\) > RSC\(_2\), there is an additional benefit = \((Q_2 - Q_1) \times (GTC_2 - RSC_2)\).

The user benefits of new trips are calculated *in the same way* as benefits of diverted trips. The logic is as before. Some new trips would be generated on the improved infrastructure when the cost falls just below GTC\(_1\) but other trips would be generated only when the cost falls close to GTC\(_2\).
This evaluation model also captures the benefits of increased output when firms produce with constant returns to scale. Suppose that a firm sells 1000 widgets at a price of $100 and has the following cost structure (inclusive of transport costs) per widget:

<table>
<thead>
<tr>
<th>Labour</th>
<th>$ 50</th>
</tr>
</thead>
<tbody>
<tr>
<td>Capital plant and equipment</td>
<td>$ 10</td>
</tr>
<tr>
<td>Materials</td>
<td>$ 20</td>
</tr>
<tr>
<td>Transport costs</td>
<td>$ 20</td>
</tr>
<tr>
<td>Total cost per widget</td>
<td>$100</td>
</tr>
</tbody>
</table>

Now, if transport costs fall to $10 per widget, the firm makes a profit of $10 per widget which is the amount allowed for in the evaluation of the existing transport of goods.

In addition, firms that previously could produce and transport widgets at between $100 and $110 per widget can now do so at between $90 and $100 per widget and make an average profit of $5.0 per widget sold (assuming no price changes). Thus, some firms may expand output and others may relocate into this market. In each such case, given constant production costs other than transport, the rule of half the savings in transport costs is a realistic measure of the benefit of increased output.

Finally, suppose that there are economies of scale and that, as output increases, other costs fall from $80 to $60 per widget. There are then savings (benefits) of $20 per widget in addition to the direct transport benefits.