Potential overlap between agglomeration benefits and other elements of appraisal

Prof Dan Graham, Imperial College London, London, SW7 2AZ, UK.
Email: d.j.graham@imperial.ac.uk

November 17, 2023

1 Introduction

This document provides perspectives on the current practice used to calculate wider economic impacts (WEIs) of agglomeration in transport appraisal, focusing on particular on potential sources of overlap with other elements of appraisal.

The objectives of the report are to:

a) Explain what is currently being done in terms of econometric estimation of agglomeration elasticities, and in the use of the econometric evidence for practical calculation of WEIs.
b) Comment on the theoretical and methodological case for use of Generalised Travel Costs (GTC) based measures of agglomeration in econometric estimation and appraisal calculations.
c) Evaluate the scope for overlap between direct user impacts (DUIs) and WEIs as currently calculated using a distance based measure of agglomeration.
d) Comment on the potential for overlap between the three categories of WEI that are assessed in UK appraisal.

The report is structured as follows. First, I give my interpretation of the thinking underpinning current appraisal practice for calculation of DUIs and WEIs. I then define the principle of additionality and explain the case used to justify additionality of DUIs and WEIs. In section 3, I describe the effects that transport investments are thought to have on productivity, noting that productivity effects should in theory be present in both DUIs and WEIs. Details concerning the calculation currently used to quantify agglomeration impacts are presented in section 4. Finally, in section 5, I assess the potential for overlap between WEIs of agglomeration other elements of appraisal.

Key points made in the report are as follows.

- Transport investments generate productivity effects through business time savings (captured by DUIs) and through WEIs.
- The fact that productivity benefits feature in changes in consumers surplus, and simultaneously in market failures, presents potential for double-counting in appraisal.

- The current approach to appraisal attempts to reduce scope for overlap between agglomeration impacts and DUIs by applying Euclidean distance based agglomeration elasticities to real changes in density measured by GTC.
- The logic underpinning this approach is that since some of the productivity effects of transport improvements are generated via time savings (e.g. DUIs), avoidance of a time based measure in the agglomeration elasticity would give a better approximation to a pure scale effect (e.g. WEI). Essentially, distance is being used as a proxy for GTC.
- It is feasible to estimate elasticities of productivity with respect to ATEM using GTC (or travel time), rather than distance, as an impedance measure.
- Without a change in the fundamental approach used to calculate WEIs of agglomeration, however, GTC based elasticities cannot be validly used in appraisal if overlap is to be avoided. This is because econometric estimation of GTC elasticities cannot separate productivity gains induced via the scale / externality effects of agglomeration from DUIs induced via time savings.
- Use of this elasticity in appraisal will certainly lead to double-counting of DUIs and WEIs.
- Use of a distance based ATEM measure remove variance in transport costs when estimating the agglomeration elasticity. This should reduce potential for direct capture of travel time / DUI productivity effects in the elasticity.
- But there are externality and non externality (e.g. DUI) effects on productivity. These effects are observationally equivalent in the sense that they both scale with ATEM while simultaneously raise productivity. This makes it difficult to successfully partition the externality effect of ATEM from non-externality influences. Furthermore, econometric adjustments used to address endogeneity may still fail to identify a pure externality effect.
- In particular, if transport costs vary systematically with ATEM (for example, are negatively correlated), the econometric model will pick such variation up in the agglomeration elasticity (e.g. due to omitted variable bias), and consequently, the elasticity of productivity with respect to ATEM will be biased as a measure of agglomeration externality.
- Furthermore, this will lead to overlap in appraisal because transport costs / time savings are valued in DUIs while also potentially captured in the agglomeration elasticity via OVB.
- There is therefore potential overlap between DUIs and WEIs in existing calculations.
- Overlap could likely be significantly reduced through:
  - a. further refinement of the econometric evidence base, and;
  - b. redesign of the fundamental approach used to calculate WEIs of agglomeration.
- It will, however, likely not be possible to eliminate overlap entirely, and it should be acknowledged that any attempt to evaluate WEIs of agglomeration provides only an approximation to real impacts. This is of course true of all other CBA calculations.
- From theory and empirical evidence, it seems likely that inclusion of agglomeration impacts in appraisal will provide a better approximation than exclusion.
- Regarding overlap between WEIS, the main issues to consider are
  - do imperfect competition and labour market effects scale with ATEM
  - have econometric adjustments in the model used to estimation agglomeration nullified their effects on the agglomeration elasticity

- It must surely be the case that competition varies systemically with ATEM, since geographic impedance is a determinant of competition. Adjustment for heterogeneity in output prices is at best uncertain, and consequently, the agglomeration elasticity may capture effects due to imperfect competition
- Labour market characteristics including skills and productivity do scale with ATEM. However, adequate adjustment is explicitly made for labour quality and for effects that arise due to the sorting of workers / firms. Therefore, there is likely more limited scope for overlap

2 Direct and wider impacts in transport appraisal

2.1 The scope of Cost Benefit Analysis

Cost benefit analysis (CBA) evaluates the ex-ante economic case for intervention by using concepts from economic theory to calculate benefits and costs in monetary values. The current version of the Green Book defines CBA as an assessment of

"...the impact of different options on social welfare. All relevant costs and benefits are valued in monetary terms, unless it is not proportionate or possible to do so" (HM Treasury 2022).

This definition emphasises that CBA should, as far as possible, be comprehensive in its coverage of costs and benefits. Inevitably, in practice, many scheme impacts prove unmeasurable, unquantifiable or unmonetisable; and therefore only a subset may be available for appraisal. Nevertheless, the principle remains the greater the range of authentic impacts included in CBA, the better.

2.2 Current Transport Analysis Guidance (TAG) calculations

In the UK, the practice of CBA is set out by the DfT in their Transport Analysis Guidance (TAG). TAG distinguishes two main categories of impacts that should be calculated for transport appraisal and that are relevant to this review.

1. **Direct user impacts (DUIs)** - impacts that are generated for both new and existing users of the transport system. These can arise in several ways, such as via changes in the generalised cost of travel (e.g. in time and fare / operating costs) or in the quality of transport services. DUIs typically constitute the largest component of benefits within conventional transport CBA. They are calculated using the concept of consumers' surplus as change in areas under the demand curve.
2. **Wider economic impacts (WEI) calculations** - WEIs refer to impacts on the economy that arise via market failures such as externalities, scale economies, and imperfect competition. They are calculated separately from the main DUIs using various sources of evidence.

2.3 The principle of additionality

Coverage of impacts in CBA must adhere to the principle of additionality. Simply put, this requires that distinct impacts, or categories of impacts, to be included within appraisal should be non-overlapping. DUIs and WEIs are considered to be non-overlapping for reasons explained below.

This basic principle of additionality underpins the DfT TAG approach to appraisal (e.g. see DfT 2019), and is widely regarded as crucial in CBA theory and practice to avoid the double-counting of impacts. Therefore, while CBA should be as comprehensive at possible, inclusive coverage should not come at the cost of double-counting.

To give a stylised example, let's say a transport investment causes a fall in GTC and a corresponding increase in consumers surplus. That change in consumers surplus may be transmitted to non-transport markets. For instance, it could appear in rents and subsequently in wages or other prices. The principle of additionality requires us not to credit the transport investment with the same benefit at three different stages of the economic transmission system. In this case, change in consumers surplus alone is sufficient to capture economic impacts.

2.4 Additionality of DUIs and WEIs

For many years UK appraisal practice sought only to assess DUIs. This focus was justified by the theoretical proposition that in the absence of market failures, and under conditions of perfect competition and constant returns to scale, DUIs alone capture all economic impacts of a transport improvement (for details see Mackie et al. 2012). The basic argument is that impacts in non-transport markets occur mainly via lower costs of production, which under competition lead to a fall in outturn prices. With elastic demand this causes an increase in employment and output, all of which is already captured under the change in consumers surplus in the transport market.

Thus in idealised conditions in which no market failures occur, DUIs capture all impacts and there is no need to undertake further calculations.

While this line of thinking dominated CBA practice for many years, it was at the same time well understood that the conditions for DUIs to capture all impacts (e.g. an absence of market failures, imperfections and scale economics) would generally not hold in practice. This led to interest in whether and how 'wider' impacts might be assessed. Note in this context 'wider' refers simply to being additional to the impacts captures under DUIs.

The argument for the existences of WEIs can thus be summarised as follows. When market failures are present, prices no longer reflect true marginal willingness to pay, and consequently, the area under a demand curve cannot be used to accurately assess consumers surplus. This means that DUIs will no longer capture the total economic effects of a scheme, and impacts on the economy that arise via market failures and are therefore wider, or additional.

Most of the attention on WEIs has been on impacts that manifest in changes in GDP or productivity. We will now consider how the productivity effects of transport improvements manifest within DUIs and WEIs.

3 The productivity effects of transport improvements

Transport investments can generate productivity effects via both DUIs and WEIs. For this reason, care is required to ensure that double-counting does not arise.

- **Productivity DUIs:** a reduction in GTC and / or an improvement in journey reliability will yield productivity gains for firms that use transport via business time savings. Firms value time savings, and thus such impacts represent welfare gains that are identified in appraisal via the DUI calculations for travel time savings.
- **Productivity WEIs:** are those that are not valued as TTS because they arise from market failures. In current UK CBA practice three categories of impacts are regarded as constituting WEIs.
  - **i Imperfect competition** - transport improvements can cause a decrease in the costs of interaction, thus potentially allowing firms to profitably expand output. Output expansion yields a welfare gain in monopolistic markets when willingness to pay for the increased output exceeds the cost of producing it.
  - **ii Tax revenues arising from labour market impacts** - the decisions that firms and workers make about where to locate is influenced by the accessibility offered through transport systems. If accessibility improves, and causes firms / workers to move to more productive locations / jobs, or have greater participation in labour markets, this will result in a tangible financial gain (i.e. higher wages or productivity). Most of this gain is captured in the consumer surplus based calculations of DUIs, but not the resulting change in tax revenue to the government (i.e. income tax, national insurance, and corporation tax).
  - **iii Agglomeration economies** - transport improvements can increase the potential scale of economic interactions available in the economy, with implications for the relative level of agglomeration experienced by firms. Essentially, improved transportation increases accessibility to economic mass (ATEM) and this in turn yields scale economies of agglomeration.

  Each of theWEIs is assumed to offer a benefit (rather than cost), and is regarded as additional (or wider) than the impacts measured as DUIs because it is a source of market failure.

Of the three defined sources of WEIs, most attention has focused on productivity effects that arise via agglomeration economies, because these are thought be the largest source of WEBs and because they can be quantified to a reasonable degree of accuracy via established econometric methods.

4 Calculation of WEIs of agglomeration

4.1 Calculating transport induced agglomeration-productivity impacts

The theoretical case for inclusion of agglomeration effects as WEIs within appraisal was developed by Venables (2007). The key insight of Venables' paper is that agglomeration economies can be induced without increasing the physical size of cities but by improving connectivity. This is because transport provision determines the effective level of agglomeration experienced by firms through its role in determining access to economic mass (ATEM). Where transport constraints exist, these serve to reduce the extent of agglomeration, while transport improvements increase the effective scale of economic activity accessible to firms essentially enlarging markets.

Venables (2007) developed a simple approach to calculate the agglomeration impacts of transport investments that required only two quantities to be measured

i The change in agglomeration caused by a transport improvement.

ii The productivity response to a change in agglomeration.

4.2 Current UK agglomeration calculations

The UK calculations implement the approaches suggested by Venables (2007) using two main sources of information

i GTC based changes in economic density ($d \log \rho_i^{\bar{g}}$) - transport models forecast changes in economic density based on GTC. For example, for n zones indexed by i, $i = (1, ...n)$, or j, $j = (1, ...n)$, the measure of economic density is of the form

$$\rho_i^{\bar{g}} = \frac{1}{n} \sum_{j=1}^{n} m_j \cdot f(\bar{g}_{ij}),$$

where m_j is a measure of economic mass at zone j and $f(\cdot)$ is a decreasing function of the mean modal GC of travelling from i to j

ii Distance based ATEM elasticities (η_{ω,ρ^d}) - elasticities that represent the strength of agglomeration by sector are estimated using a distance-based measure of economic density. For estimation, the measure of economic density is of the form

$$\rho_i^D = \frac{1}{n} \sum_{j=1}^{n} \frac{m_j}{d_{ij}^{\alpha}},$$

where d_{ij} is the Euclidean distance between zones i and j and α is a parameter that determines the decay of agglomeration impacts over space.

Using the two quantities described above, the transport induced WEIs of agglomeration (e.g. impacts on productivity) could be computed using a calculation such as

$$\begin{aligned} d \log \omega &= \sum_{i=1}^{n} \eta_{\omega,\rho^d} \times d \log \rho_i^{\bar{g}} \\ &= \sum_{i=1}^{n} \eta_{\omega,\rho^d} \times \left[\underbrace{\sum_{j=1}^{n} \eta_{\rho_i^{\bar{g}},\bar{g}_{ij}} d \log \bar{g}_{ij}}_{\text{static effect}} + \underbrace{\sum_{j=1}^{n} \eta_{\rho_i^{\bar{g}},m_j} d \log m_j}_{\text{dynamic effect}} \right] \end{aligned}$$

The second row above emphasises that current UK calculations recognise two possible effects on agglomeration (ρ_i) from a transport improvement

i **Static impacts** - arising from changes in GTC, holding masses constant. This component captures the effective increase in ATEM that occurs when GTC falls, e.g. $-\Delta \bar{g}_{ij} \rightarrow +\Delta \rho$. The change in GTC is forecast from a transport model

ii **Dynamic impacts** - arising from changes in the distribution of economic mass. This component seeks to predict the post intervention equilibrium of economic mass within the ATEM measure, e.g. $+\Delta m_j \rightarrow +\Delta \rho$, so called dynamic agglomeration. The change in distribution of masses is forecast via a spatial model (e.g. LUTI or SCGE).

4.3 Estimating the agglomeration elasticity

Agglomeration elasticities have to be estimated using an econometric model. Estimation based on disaggregate micro-level panel data (e.g. for workers and firms) has become the norm. A generic form of linear panel model used for this purpose is

$$\log \omega_{ict} = \delta \log \rho_{ct} + f_{ct} + u_{it} + z_{it} + e_{ict},$$

where ω_{ict} is the productivity of worker or firm i in areas c at time t, ρ_{ct} is the ATEM of area c at time t, f_{ct} is unobserved area effects potentially correlated with ρ_{ct}, u_{it} is unobserved firm effects potentially correlated with ρ_{ct}, z_{it} represents other measurable effects on productivity, and e_{ict} is a random error term.

Estimation of course requires a representation of spatial variance in productivity (i.e. ω) and its determinants. There are two approaches that are commonly used to achieve this.

1. **Total Factor Productivity (TFP)** – models for TFP estimate a statistical relationship between the economic output of firms and factor inputs (i.e. labour, capital, materials, etc.). This is done via a production function model which specified the nature of the relationship between the inputs that are used in production and the outputs that are produced. Specific determinants of productivity are included in the production function as shifters of productivity. The generic model form is

   $$y = g(z)f(X)$$

   where y is output, X is a vector of input factors, and z is a vector of firm and local market characteristics that are assumed to determine production levels.

   Estimation of this model requires geocoded data on production (outputs and inputs) and on the characteristics assumed as determinants of productivity.

2. **Labour Productivity (LP)** – models for LP typically make the assumption that factor inputs are paid the value of their marginal product, and they then use the worker wage rate as an implicit measure of labour productivity. Estimating equations for wages can be derived from production functions by assuming optimising behaviour on the part of firms (i.e. profit maximisation or cost minimisation). Alternatively, at the level of individual workers, we can specify a Mincerian type wage equation where the wage (productivity) of some worker in a given location is explained by a set of worker-specific variables (education, age, skills etc.) and a set of 'environmental' (e.g. local market) characteristics. The generic wage equation model form is

   $$w = g(z)f(u)$$

   where w is the wage rate, u is a vector of covariates relevant for wage determination, and z is a set of enviromental characteristics assumed to affect labour productivity.

Estimation of this model requires geocoded data on wages and on the characteristics assumed as determinants of labour productivity.

For either wage of production models, there are three key issues that determine the specific nature of the productivity models to be estimated.

1. **Sectoral breakdown** - existing empirical evidence indicates that the spatial determinants of productivity tends to vary by economic sector. For this reason estimation of TFP or LP models should, as far as possible, be conducted separately for some reasonable industrial disaggregation.
2. **Data type** - spatial productivity could be estimated using either cross-sectional or panel data. In a cross-section a sample of observational units are each observed at a single point in time. With panel data, cross-sectional units are observed at different points in time. The vast majority of recently published empirical work in the field makes use of panel data, and the general rule of thumb in the literature is that panel data are superior for reliable estimation. This is due to a number of advantages that panel data offer for estimation of TFP and LP models.
   - They allow for application of sophisticated methodologies capable of producing robust measures of productivity.
   - They allow dynamics and adjustment in behaviour (i.e. lagged effects) to be studied.
   - The precision of estimation can be increased by using both between unit and within unit variation.
   - Reverse causality and other problems of "endogeneity" can be addressed more readily using panel estimators.
3. **Data aggregation** - TFP and LP models can be estimated using data at various level of aggregation. Early studies of spatial productivity tended to use aggregate production function or wage models for spatial zones. More recent studies tend to be based on microdata for firms and workers, typically in panel form. In the recent academic literature, microdata are generally regarded as superior for the measurement of productivity because they offer a number of useful advantages in estimating the production function and wage equations. In particular, the use of micro panel data allows for application of more sophisticated methodologies to deal with potential inconsistency of the OLS estimator (i.e. via panel instrumental variables, fixed effects, random effects, dynamic GMM models and semi-parametric approaches). In addition, there is now a general consensus in the literature that analyses of productivity should be undertaken at the micro level because the behavioural assumptions inherent in the modelling framework (i.e. profit maximisation or cost minimisation) apply at this level.

Even with micro-panel data, the fact that we can construct only a partial representation of productivity, presents a number of challenges in estimating agglomeration effects, which ultimately dictate the form of the model used. These challenges are as follows.

- i **Endogeneity from unobserved firm / worker productivity** - our empirical representation of productivity at the level of firms and workers is likely incomplete. Unobserved attributes such as factor input quality, technology, and unmeasured capital can cause bias in estimation if they are correlated with ATEM.
- ii **Reverse causality** The relationship between ATEM and productivity may be simultaneously determined. Higher productivity locations may attract a greater level of private

investment over time leading to larger economic mass, and this increase in mass can feedback by raising productivity.

iii **Sorting / selection bias** - more productive firms / workers tend to survive / sort disproportionately into dense urban areas. Correlation between unobserved functional or occupational differences, and the level of agglomeration, can bias the estimated relationship between agglomeration and productivity.

iv **Confounding** - there are a variety of unmeasured local market characteristics that may be correlated with both productivity and agglomeration. These include specific characteristics of local input and output markets, such as price heterogeneity and the existence of public goods, including transport infrastructure. Thus the ATEM variable will potentially capture 'confounding' effects in addition to the marginal effect of agglomeration on productivity.

Several approaches have been used to attempt to address these potential sources of endogeneity, mainly based around instrumental variable (IV) and control functions (CF) approaches. The logic underpinning these approaches is as follows.

1. **Panel IV** - specify unit level fixed effects to address time-invariant confounding, and apply an 'instrument' correlated with ρ but uncorrelated with ω to nullify the effect of unobserved covariates.
2. **Panel CF** - include a function that proxies for unobserved productivity as an additional model component. Use an AR(1) specification to draw on variation over time rather than across firms.

Identification via IV requires that the instruments be valid, that is, genuinely exogenous to productivity while still being relevant to (e.g. correlated with) agglomeration. Identification via CF models requires 'structural' assumptions concerning firm behaviour to hold (for details see Graham and Gibbons 2019)

4.4 Current UK elasticities

The agglomeration parameter values used for appraisal in the UK were estimated by Graham et al. (2009). They use ONS firm level micro panel data, from the Annual Respondents Database (ARD), to estimate total factor productivity (TFP) within a Cobb-Douglas production function model. To represent agglomeration they use a distance based ATEM. They adopt a panel control function approach for estimation to addresses potential sources of endogeneity arising from unobserved productivity, including via heterogeneity in input quality.

Aomeration elasticities are estimated separately for four broad sectors of the economy: manufacturing, construction, consumer services and business services.

The results from this study yield an overall agglomeration elasticity of 0.04 across all sectors of the economy. For manufacturing and consumer services they estimate an elasticity of 0.02, for construction 0.03, and for business services 0.08. The distance decay parameter is found to be approximately 1.0 for manufacturing, but around 1.8 for consumer and business service sectors and 1.6 for construction. This implies that the effects of agglomeration diminish more rapidly with distance from source for service industries than for manufacturing. The relative impact of agglomeration on productivity is, however, larger for services than it is for manufacturing.

The key empirical results of their research are summarised in the table 1.

Table 1: Summary of UK agglomeration elasticity parameters estimated by Graham et al. (2009)

	*sic**	*agglomeration elasticity*	α
Manufacturing	15-40	0.024	1.122
		(0.002)	(0.127)
Construction	45	0.034	1.562
		(0.003)	(0.159)
Consumer services	50-64	0.024	1.818
		(0.003)	(0.190)
Business services	65-75	0.083	1.746
		(0.007)	(0.144)
Economy (weight aver.)	15-75	0.044	1.659

* Standard Industrial Classification

5 Potential overlap between WEIs of agglomeration and other elements of appraisal

5.1 Potential for overlap and current approach

The risk of overlap we consider here is that the elasticity of productivity with respect to ATEM captures some of the productivity (GDP) effects already includes in other elements of appraisal.

Since WEIs should arise solely from market failures, it is necessary that the agglomeration elasticity reflects, as far as possible, a pure externality scale effect, net of the effects of other transport cost attributes (e.g. time and money costs) that potentially scale with ATEM.

This is a difficult aim to achieve econometrically. The approximation that has been used to date employs a Euclidean distance based measure of ATEM in the econometric models to avoid the influence of time based productivity effects. The thinking was that since some of the productivity effects of transport improvements are generated via time savings, avoidance of a time based measure for estimation of the agglomeration elasticity would give a better approximation to an externality scale effect. Essentially, the distance based ATEM is being used as a proxy for the GTC based ATEM.

Graham (2006) showed that use of a GTC based measure of ATEM produced higher agglomeration elasticities than a distance based ATEM. This is because it captures both time and distance dimensions of effective density. However, this result should not be taken as evidence that the distance based approach achieves estimation of a pure externality effect. It simply indicates an obvious result: that inclusion of variance in transport costs in the ATEM covariate means that the direct influence of travel time / DUI productivity effects will be fully captured in the agglomeration elasticity.

In this section we explore three key issues related to overlap of DUIs and WEIs that arise at the intersection of the econometric modelling and appraisal calculations / transport modelling.

1. Can we estimate and use GTC based elasticities for transport appraisal?
2. Do distance based agglomeration elasticities capture effects that arise only from market failure?
3. Does overlap exist between the three categories of WEI that are defined for UK appraisal practice?

5.2 Can we estimate and use GTC based elasticities for appraisal?

The seeming mismatch in use of impedance measures, that is distance in the ATEM used in econometric modelling and GTC in the ATEM used in the transport modelling / appraisal calculations, has proven a source of confusion over the years. Transport consultants in particular have often criticised current practice on the grounds that use of quantities measured in different units introduces discontinuity in calculation.

The questions to address here are: can we *estimate* and *validly use* GTC based elasticities for appraisal?

5.2.1 Estimation of GTC based elasticities

I will deal first with the issue of estimation. It is eminently possible to estimate econometric models with a GTC based measure of ATEM (for example, as in Graham 2006), but there are three key empirical challenges that we face in doing so.

1. **Data availability** - GTC and travel time data are often unavailable as a spatial panel, often necessitating the use of approaches to infer GTC, which introduces uncertainty in estimation.
2. **Endogeneity** - GTC based ATEM tends to be more endogenous than a distance based measure because it incorporate the cost of network congestion, which is in part determined by the level of productivity and thus by agglomeration. Reverse causality is therefore a more serious concern than with a distance based ATEM.
3. **Complexity** - to construct a single GTC based ATEM measure, rules have to be introduced on how to incorporate variance across modes and times of day, and on how to address zero entries in OD matrices when it is not possible to travel from i to j by a given mode.

None of the above challenges are insurmountable, and with better data we can apply innovative approaches to address endogeneity problems.

Conclusion: it is feasible to estimate elasticities of productivity with respect to ATEM using GTC (or travel time) as an impedance measure

5.2.2 Use of GTC based elasticities in appraisal

Having estimated GTC based elasticities of productivity with respect to ATEM, could they be used validly in appraisal to calculate agglomeration impacts?

As explained above, the average effect of ATEM on productivity is estimated using panel data. Use of a GTC based measure of ATEM captures variances (across places and time) in

- Economic mass - m_j, often represented by total employment.
- Travel time and costs - g_{ij}, which will vary due to travel conditions and according to the quality and volume of infrastructure.

Focusing on the second element of variance in the ATEM, it is clear that when transport interventions are made their effects will show up in GTC. As explained above, transport improvements generate productivity impacts through two routes: business travel time savings (DUIs) and agglomeration effects (WEIs). The agglomeration elasticity estimated using the GTC based measure of ATEM, will thus capture a composite of these two effects and it will not be possible to distinguish them. Use of this elasticity in appraisal will certainly lead to double-counting of DUIs and WEIs.

Conclusion: without a change in the fundamental approach used to calculate WEIs of agglomeration, GTC based elasticities cannot be validly used in appraisal if overlap is to be avoided. This is because econometric estimation of GTC elasticities cannot separate productivity gains induced via the scale / externality effects of agglomeration from DUIs induced via time savings.

5.3 Do distance based agglomeration elasticities capture effects that arise only from market failure?

In this subsection we turn to the form of agglomeration calculation currently used in the UK, and ask whether the approach of using distance based agglomeration elasticities is really able to isolate the pure external scale effect of agglomeration. As mentioned above, strictly speaking, it is necessary that the agglomeration elasticities achieve this target of identification to avoid double counting.

The generic econometric model presented above (e.g. section 4.3)is used to estimate an agglomeration elasticity that captures how productivity responds to ATEM *ceteris paribus*. The key point relevant to the overlap issue we consider here is what *ceteris paribus* actually means in this context. Specifically, to what extent is estimation with a distance based ATEM able to net out the influence of non-externality effects leaving an elasticity that represents only the pure scale effect of agglomeration?

There are essentially two routes to netting out the influence of non-externality confounding effects.

- Adjustment for observed and unobserved confounders, via a combination of direct covariate adjustment and use of panel individual effects / within estimation.
- Estimation based on IV or CF to nullify the effect of other potential sources of endogeneity such as time-varying confounding, measurement error, or reverse causality.

These econometric approaches are used routinely in the empirical literature on agglomeration, but they are not foolproof. Furthermore, due to the typically high correlation that exists between externality and non-externality factors that could cause ATEM to induce productivity gains, in practice it can be challenging to know whether and to what extent ceteris paribus adjustments have really been achieved.

Essentially, we face a problem of *observational equivalence*. Externality and non-externality attributes are changing systematically with ATEM in the same direction (e.g. increasing), and they have the same general effect on productivity (e.g. creating gains). So it become very hard to distinguish them empirically.

Of particular concern for use of elasticities in appraisal is confounding bias from transport costs / time savings. If transport costs vary systematically with ATEM, and simultaneously influence productivity, the econometric model will likely pick up transport cost effects in the agglomeration elasticity as a result of omitted variable bias. Consequently, the model will fail to identify a pure externality effect, and the elasticity of productivity with respect to ATEM will thus be biased as a measure of the agglomeration externality. Furthermore, this will lead to overlap in appraisal because transport costs / time savings are valued in DUIs, but also potentially captured in the agglomeration elasticity via OVB.

The extent to which non-externality effects have been netted out is hard to determine. Regarding the estimates presently used for transport appraisal, I am confident that we cannot claim that these represent only a pure externality effect net of other confounding factors. They show only how productivity scales with ATEM, having made some econometric adjustments to address sources of endoegeneity. There is therefore, in my opinion, potential overlap between DUIs and WEIs in existing calculations.

Two final points are worth making here.

First, in making cost benefit calculations we are of course working with approximations and should not set the bar at an unattainably high level of accuracy.

Second, it is worth emphasising that the methodological issues raised in this section are incredibly difficult to address. The critique that agglomeration elasticities potentially capture more than pure externality effects is one could be made of the entire empirical agglomeration literature. However, the implications are more immediate when the elasticity is being applied to make appraisal calculations.

Conclusions:

- **Use of a distance based ATEM measure remove variance in travel costs when estimating the agglomeration elasticity. This should reduce potential for direct capture of travel time / DUI productivity effects in the elasticity.**
- **But there are externality and non externality effects on productivity which scale with ATEM. Observational equivalent makes it difficult to partition the externality effect of ATEM from other confounding influences. Econometric adjustments to deal with this use are not foolproof.**
- **Therefore, in my opinion, we cannot claim unequivocally that distance based elasticities of productivity with respect to ATEM identify only a pure externality scale effect.**

- **In particular, if transport costs vary systematically with ATEM, the econometric model will pick these up in the agglomeration elasticity (e.g. due to omitted variable bias), and consequently, the elasticity of productivity with respect to ATEM will be biased as a measure of agglomeration externality.**
- **Furthermore, this will lead to overlap in appraisal because transport costs / time savings are valued in DUIs while also potentially captured in the agglomeration elasticity via OVB.**
- **There is therefore potential overlap between DUIs and WEIs in existing calculations.**
- **Overlap could likely be significantly reduced through:**
  - **a. further refinement of the econometric evidence base, and;**
  - **b. redesign of the fundamental approach used to calculate WEIs of agglomeration.**
- **It will, however, likely not be possible to eliminate overlap entirely, and it should be acknowledged that any attempt to evaluate WEIs of agglomeration provides only an approximation to real world impacts. This is of course true of all other CBA calculations.**
- **From theory and empirical evidence, it seems likely that inclusions of agglomeration impacts in appraisal will provide a better approximation than exclusion.**

5.4 Does overlap exist between the three categories of WEI?

The final issue to be considered in this section concerns whether there is potential overlap between the three categories of WEI that are currently appraised: imperfect competition, labour market impacts, and agglomeration.

Again, the main issue to address here is whether imperfect competition and labour market effects scale with ATEM, and if so, whether adequate adjustments has been made in the econometric model for agglomeration to nullify their effects on the agglomeration elasticity.

First, we consider imperfect competition. It must surely be the case that competition varies systemically with ATEM. Indeed, the case for including imperfect competition effects in appraisal is essentially based on this premise: e.g. reduced interaction costs in the spatial economy induce firms to expand output.

There is no foolproof way of adjusting for heterogeneity in output prices in the econometric model for agglomeration, and consequently, the agglomeration elasticity may capture effects due to imperfect competition. In my opinion, therefore, there is scope for overlap between agglomeration and imperfect competition effects.

Regarding the labour market WEI component, it is the case that labour market effects scale with ATEM. However, the econometric model for agglomeration does include an explicit adjustment for labour quality and other productivity effects that arise from the sorting of workers of different skill levels into different locations. In my opinion, therefore, there is more limited scope for overlap between agglomeration and labour market effects when the agglomeration model is designed to distinguish these effects separately.

Conclusion: there is potential overlap between agglomeration and imperfect competition effects, but less scope for double counting between agglomeration and labour market effects.

References

DfT (2019). *TAG UNIT A2.1: Wider Economic Impacts Appraisal.* London: DfT.

Graham, D. J. (2006). *Wider economic benefits of transport improvements: link between agglomeration and productivity, Stage 2 Report.* London: DfT.

Graham, D. J. and S. Gibbons (2019). Quantifying wider economic impacts of agglomeration for transport appraisal: Existing evidence and future directions. *Economics of Transportation 19*, 100121.

Graham, D. J., S. Gibbons, and R. Martin (2009). *The spatial decay of agglomeration economies.* London: DfT.

HM Treasury (2022). *The Green Book.* London: HM Treasury.

Mackie, P., D. J. Graham, and D. Laird (2012). *Direct and wider economic benefits in transport appraisal*, Chapter in A de Palma, R Lindsey, E Quinet, and R Vickerman (eds) Handbook in Transport Economics, pp. 501–526. London: Edward Elgar.

Venables, A. J. (2007). Evaluating urban transport improvements: cost-benefit analysis in the presence of agglomeration and income taxation. *Journal of Transport Economics and Policy 41*(2), 173–188.