Review of elements of methodology for HS2 business case

This review considers three of the significant changes to the calculation of the economic benefits generated by HS2. Using the terminology of the Macpherson Review of quality assurance of Government analytical models, this is an external peer review (a formal or informal engagement of a third party to conduct critical evaluation, from outside the organisation in which the model is being developed).

1. Rail demand matrices and the derivation of journey purpose

HS2 Ltd has replaced its previous journey purpose to ticket-type mappings with corridor-specific proportions taken directly from the National Rail Travel Survey. In light of this change, is the derivation of the base year rail demand matrices intuitively and mathematically sound?

The Planet model forecasts changes in long-distance trips by mode, including HS2, in response to changes in generalised cost. The development of the forecasts uses base year rail demand matrices, and hence the need to estimate these matrices.

The LENNON database is the rail industry's central ticketing system. It provides information on passenger kilometres, journey data, and ticket sales. Data from 2010/11 were available. The database does *not* give journey purpose (commuting, business, leisure). A decade ago journey purpose could be reasonably estimated by the ticket type purchased. Since then ticket types on offer have radically changed, and it is no longer reasonable to map a ticket type directly to a journey purpose.

The NPS (National Passenger Survey) collects passenger opinions of train services twice a year from a representative sample of journeys across the network. It is designed to assess consumer satisfaction, but does collect journey purpose (and indeed satisfaction is very different for commuters and leisure travellers). The NPS data indicates that purpose splits (the proportions of rail travel that is commuting/business/leisure) over the network has been stable nationally over the last decade while splits between ticket types have not.

The NRTS (National Rail Travel Survey) was a survey of passenger trips on the national rail system in Great Britain on weekdays outside school holidays. It combined data collected in London and the South East in 2001 with further survey data collected in 2004 and 2005 covering Wales, Scotland and the rest of England. It is a larger survey than NPS and gives purpose splits with geographical variation (indeed NRTS was designed to understand travel patterns for planning purposes). The NRTS data shows a pronounced difference of purpose splits by geography: in particular, the proportion of trips to and from London with a business purpose is higher than the national average.

Thus the NRTS gives reliable purpose splits for travel to and from London, but from older data. LENNON gives recent rail demand over each part of the network, but does not give purpose splits. NPS indicates that, at least nationally, purpose splits have been stable over the last decade. Given this data it is a reasonable assumption that the purpose splits for travel to and from London have not significantly changed over the last decade, and can be estimated from the NRTS data, while the travel volume can be estimated from LENNON data. This is, in my view, an improvement on the earlier derivation of base year rail demand matrices for journey purpose from ticket types.

Improvements

The Department for Transport should move to more extensive use of administrative data routinely collected for other purposes. Such data is unlikely to provide all that is needed for forecasting purposes but should be the starting point, with additional data collection designed to leverage existing administrative data. The LENNON database and the National Passenger Survey provide up-to-date data. The Department should explore whether reliable geographically differentiated purpose splits can be derived from NPS data: given sample sizes and sampling methodology the NPS data alone may not be enough and may need to be augmented; but since NPS is a rolling survey updated twice yearly there will be a 10 fold increase in sample size over five years.

2. Disaggregation and the 'rule of a half'

Benefits calculation - The method by which the model aggregates the economic benefits of HS2 following changes in demand and travel costs has been amended so that the calculation of benefits is now done at the most disaggregate level possible using the 'rule of a half' before being aggregated across all model zones. In light of this change:

Was the previous composite cost approach mathematically correct?

Is the current approach mathematically correct?

Is the current approach in-line with WebTAG unit 3.5.3 regarding the calculation of transport user benefits? ; and

Is there an alternative approach that should be investigated?

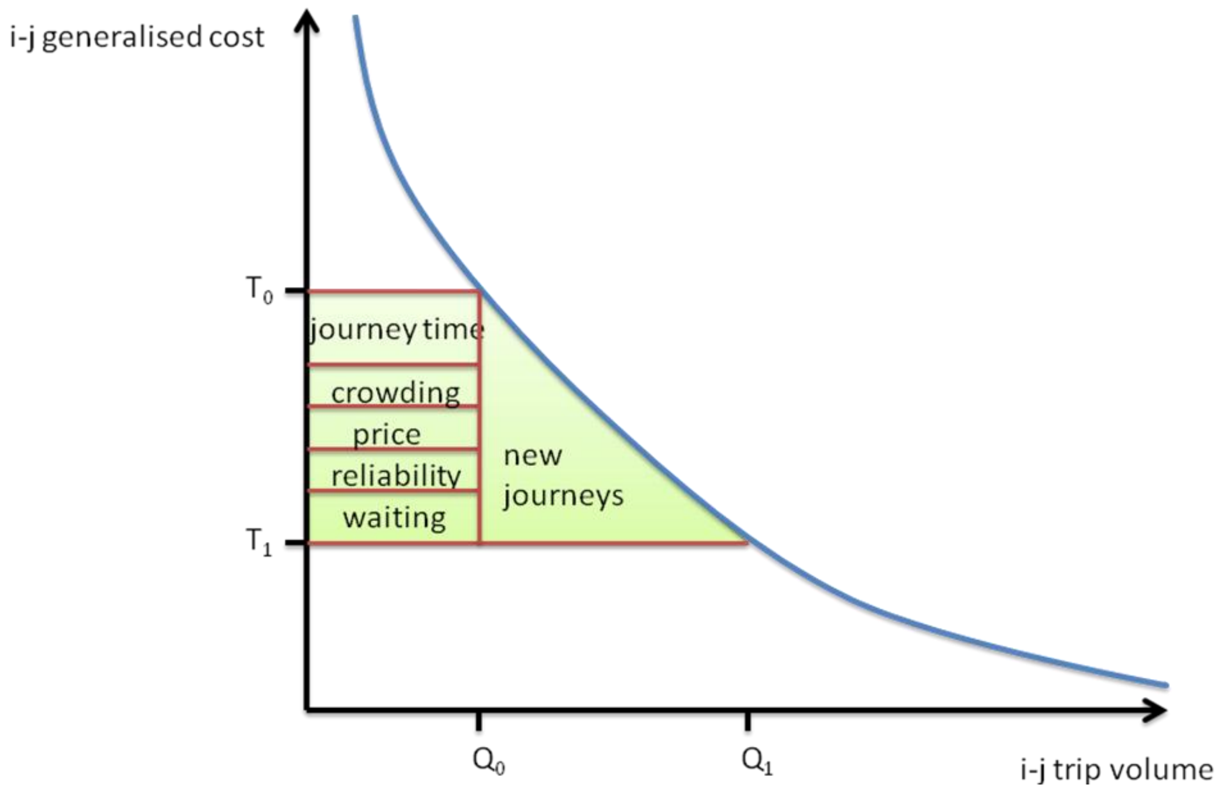
I greatly welcome the inclusion in the Economic Case for HS2 (October 2013) of distributions of benefit cost ratios, reflecting the inherent uncertainty around several of the model inputs. Models necessarily make simplifications and approximations, and when assessing whether a model is adequate the resulting errors need to be seen in the context of the inescapable uncertainty about any prediction of the future, and the use to which the model is put (for example whether appraising a single scheme, or comparing two schemes subject to broadly the same uncertainty about demand forecasts).

The previous approach using composite costs (a particular form of averaging generalised costs over a set of choices) and the current approach (calculating benefits at a more disaggregated level and using the 'rule of a half') both involve simplifications and approximations. A purist would conclude that neither is mathematically correct. But that is too high a hurdle.

All models are wrong, but some are useful (George Box, statistician).

To explore the nature of the simplifications and approximations, some background is needed.

Consider a transport improvement on a long, thin island where people live only at the two ends, labelled i and j, and where just one route, between i and j, is considered. The increase in consumer surplus for travellers on the trip i-j is the green shaded area in the diagram below; the point (Q_0,T_0) is before the transport improvement, and the point (Q_1,T_1), is after the transport improvement. Suppose the generalised cost is linear in its components, such as journey time, price, waiting, etc. Then the increase in consumer surplus for existing travellers comprises a sum of small rectangles, one for each component. There is also an increase in consumer surplus arising from new journeys, the near triangular region in the diagram.



For small changes, two mathematically well-founded approximations follow immediately. Suppose the change (between Q_0 and Q_1 and between T_0 and T_1) is of order ε where ε is small. (This is not the situation in the diagram, where to see what is going on the change is not small.)

First, the area of the near triangle will be negligible in comparison with the area of the rectangles (the near triangle's area is of order ε^2, while the individual rectangles are each of much larger order ε). Thus the increase in consumer surplus is well summarised by the improvements for existing travellers. In particular, the allocation of the benefit of the transport improvement across the components of generalised cost labelled crowding, journey time, etc. is straightforward.

Second, the area of the near triangle is well approximated by (T_0- T_1) (Q_1- Q_0)/2, the 'rule of a half'. Indeed the 'rule of a half' may be a reasonable approximation (as for example in the diagram), even when the first approximation does not hold. All that is needed is that the demand curve be nearly linear between (Q_0,T_0) and (Q_1,T_1). Otherwise, as explained in Section 8.3 of the Economic Case for HS2, it may overestimate or underestimate the benefit according to the convexity or concavity of the demand curve. As an example, if Q_0=0 in the diagram, corresponding to a new link, then the whole of the consumer surplus is in the not-so-near triangle, whose upper point may be arbitrarily high. The area can be estimated by numerical integration, but not reasonably by the 'rule of a half'.

What if the 'rule of a half' is a reasonable approximation or if numerical integration is used, but the first approximation does not hold? In this case the total benefit is well estimated, but it is not straightforward to allocate it across the components of generalised cost. Usually the area of the near triangle is allocated in proportion to the relative sizes of the rectangles. If the weights used for appraisal are the same as the weights used in the demand model, then the allocation is mainly a decision about presentation; but if the weights used for appraisal are *not* the same as the weights used in the demand model, then the (appraisal weighted) net benefit will depend on just how the area of the near triangle is allocated.

The increased social and economic opportunity offered by lower transport costs is represented by the near triangle, and for large transport improvements this near triangle may be a substantial component of the increase in consumer surplus. For at least some presentational purposes it may be misleading to label it as a reduction in crowding, journey time, etc. (even though all areas are measured in the same units). For example, it may allocate the benefit to an individual of making a new trip to the benefit labelled as a *reduction* in journey time even when the effect of the new trip will be to *increase* the total journey hours. (The theoretical justification is that the new trip was previously deterred by the journey time.) And, as we've noted, if the weights used for appraisal are *not* the same as the weights used in the demand model then this is more than a presentational issue.

The foregoing discussion concerned an isolated transport link served by a single route. For a transport network, we could repeat the above analysis for each link, and aggregate the various rectangles over the different links to get the aggregate benefit in crowding, journey time, etc. Is this mathematically well-founded, perhaps in the case of small changes?

This requires a more technical discussion. The demand curve illustrated could describe the relationship between trip volume and generalised cost on a single trip for a trajectory through the space of possible networks (imagine a continuous change in some feature of the transport network, for example the capacity and speed on a rail segment, both increasing gradually from zero to infinity). There will be a demand curve for each trip in the network, and the shape of these curves may depend on the trajectory chosen. Demand forecasting models that assign trips to routes are usually derived from an underlying utility model: this is important, since it implies the total benefit of the transport improvement, assessed by adding the increases in consumer surplus across all trips, will not depend on the precise trajectory chosen between the two networks describing the situation before and after the transport improvement. We should expect to be able to aggregate the small rectangles over different links if the change is small, so that the near triangles are negligible.

However with route choice, a difficulty arises from the nature of the underlying utility models for demand forecasting.

Consider a composite link, which comprises several routes between i and j. Travellers have factors affecting their utility that are not observable, and will influence choice between routes. Two routes between i and j on our earlier long thin island may have precisely the same generalised cost, and even the same components of generalised cost, but if they proceed along two different coastlines and one is by train and one by bus, the choice of coastline and/or mode may have value to travellers. More generally, if the diagram describes a composite link between i and j, then we need a further small green rectangle. This additional rectangle is the 'diversity' benefit. It is of a similar order ε to the other rectangles, and thus is significant even for small changes. If the weights used for appraisal differ from the weights used in the demand model, then the allocation of 'diversity' benefit is problematic, and this caused a significant problem with the earlier composite cost approach. (If the weights used for appraisal are the same as the weights used in the demand model then the composite cost approach has the advantage that it gives net new journeys on the composite link, taking into account shifts from one route to another.)

The current approach (calculating benefits at a more disaggregated level and using partial derivatives of the composite cost with respect to the components of generalised cost to allocate benefits across the components of generalised cost) avoids the problem diversity causes for composite links when weights differ between appraisal and the demand model (it reduces the order of error to ε^2), and used it in conjunction with the 'rule of a half' gives robust estimates for small changes.

If the change is large, underlying utility models will rarely allow disaggregation of the components of generalised cost. All of this is well understood in the literature: Jones (1977) writes "It may be noted that for presentational purposes the over-all benefit measure can be disaggregated by types of benefit, although these types of benefit should not be taken as a final measure of incidence" and Bates (2003) writes "The general principle is that while disaggregations of the total benefit may be indicative, only the total is theoretically unambiguous".

By using composite costs the earlier approach seems to have avoided the need for numerical integration, perhaps because on composite links the changes were smaller and seemed better approximated by `the rule of a half'. However the allocation of total benefit to the components of generalised cost was deficient, because of the problem caused by the `diversity' benefit on composite links, and when appraisal weights differ from demand model weights this affects the (appraisal weighted) net benefit.

The current approach (calculating benefits at a more disaggregated level and using the 'rule of a half') improves on the allocation of total benefit to components of generalised cost, but numerical integration should be used rather than the 'rule of a half' on disaggregated links where changes are large. WebTAG unit 3.5.3, in particular paragraph 2.1.10, refers to the TUBA guidance for the key issue here, which recommends numerical integration rather that the 'rule of a half' when there are large changes. There is little guidance in WebTAG itself on calculating benefits at the disaggregated level: it refers to the TUBA manual which recommends calculating benefits at the disaggregated level, as discussed in the Planet Framework Model (PFM v4.3), paragraphs 12.2.10-12.2.11.

In summary, neither model is mathematically correct, but both are useful. (For example, composite links are important for assessing impact on geographical regions.) I would suggest that calculating benefits at the most disaggregated level be combined with numerical integration rather than the 'rule of a half' where changes are large. The results are presumably those broadly outlined in Section 8.3.5 of the Economic Case for HS2.

Improvements

The difficulties faced are partly a consequence of the desire to disaggregate the total benefit into the components of generalised cost. Demand forecasting models need to use definitions of generalised cost that are statistically well-founded: in particular, the demand model weights associated with the components of generalised cost must predict reasonably well the decisions of individual travellers (whether to travel and which route to take). Policy makers may reasonably have other objectives, and appraisal weights that differ from demand model weights may be an attempt to recognise this. And there may be externalities associated with some components of generalised cost that are not accounted for in the utility formulation. Also, the total benefit is just a single number (or, better, a distribution for this number) and it is reasonable to try to understand how it arises.

I would recommend an explicit discussion of the effects not captured in the demand models, followed by an exploration of how these can accommodated consistently and transparently. A principled approach to externalities would include additional terms in the utility formulation to capture externalities (the utility formulation is described in de Jong et al, 2005). Since the demand forecasting model finds a user equilibrium it is unlikely that externalities can cause such perverse behaviour as Braess Paradox (Section 4.2.1 of Kelly and Yudovina, 2014), but including them explicitly will avoid the difficulties of appraisal weights differing from demand forecasting weights.

To understand how the total benefit arises, I would suggest that modern tools for the visualisation of data could add significant additional insight to the bare list of indicative components of generalised cost. Figure 6.3 and 6.4 from the Strategic Case for HS2 give examples for journey time, and a variety of graphical techniques could be used for the components of generalised cost, in order to communicate different facets of just how the total benefit arises.

3. Inflation

Inflation - The appraisal of HS2 is undertaken using the GDP deflator as an inflation measure to convert between nominal and real values. The fares assumption, however, is in line with current Government policy and is defined on the basis of an RPI inflation measure. The inconsistency in these measures of inflation means that real terms fares growth therefore requires adjustment so that it is based on the GDP deflator measure of inflation rather than RPI. Has the conversion between RPI and the GDP deflator been applied correctly?

The modelling framework uses the ONS GDP deflator to present costs, revenues, fares and benefits in real terms. In 2011 ONS changed the GDP deflator from an arithmetic to a geometric mean: it is now closer to the CPI than the RPI, although not quite the same as either.

Government policy is that regulated fares increase at RPI+1% per annum (with RPI measured from July to July). The model assumes all fares, regulated and unregulated, increase at this rate until 2036 from when they stay constant in real terms. RPI is forecast to grow at a faster rate than the GDP deflator index: over the forecast period fares growth of RPI+1% is equivalent to the GDP deflator+1.9%. This will increase the estimate of real fares reported by the model.

Demand forecasts use elasticities understood to be calibrated to fares growth deflated using RPI, and so the change does not affect these forecasts. Within the demand model there are some inconsistencies concerned with value of time growth relative to price growth, but this effect is not likely to be substantial.

This review is not a model audit (in the terminology of the Macpherson Review of quality assurance of Government analytical models), but I have looked through the spreadsheet calculations. The effect of using the GDP deflator rather than RPI to convert between nominal and real values is to increase the NPV of future revenue by 27%. Given the model assumptions on fares increases and on demand elasticity, this is, in my view, a reasonable estimate.

References

John Bates (2003) Economic Evaluation and Transport Modelling: Theory and Practice. Resource paper for 'Moving through nets: The physical and social dimensions of travel', 10th International Conference on Travel Behaviour Research, Lucerne.

Ian S. Jones (1977) Urban Transport Appraisal. Macmillan, London.

Gerard de Jong, Andrew Daly, Marits Pieters and Toon van der Hoorn (2007) The logsum as an evaluation measure: Review of the literature and new results. Transportation Research Part A: Policy and Practice, Volume 41, Pages 874–889.

Frank Kelly and Elena Yudovina (2014) Stochastic Networks. Cambridge University Press.

Frank Kelly, Cambridge, 18 November 2013.