Rail demand forecasting – ten lessons of a lifetime

Lesson 1: Unexpected changes in the supply side can undermine demand forecasts

The case of Eurostar and CTRL is widely regarded as a good example of poor forecasting. But projections of very high rail market share of London- Brussels (60%+)/Paris (80%) markets for Eurostar based on logit models made in the period 1992-4 were accurate and borne out in practice.

The area of inaccuracy lay in (what seemed a less risky) presumption of a small market share of the much wider and multiple travel markets between the wider GB geography and the near European continent. But here the observable habits of long distance rail travel across EU member state borders simply failed to develop. There are three main reasons why this happened:

1. The UK opted out of the Schengen Agreement in 1999 and this meant that border crossing got harder not easier by rail and timed interchanges for onward travel became problematic
2. The various state railway companies and Eurostar failed to agree an integrated ticketing strategy that made through booking feasible
3. Services for which entire train fleets had been bought – the Regional Eurostar fleet and the Night trains fleet – were never introduced into service.

This is the same type of problem that arose with economists' failure to foresee the credit crunch: the demand models in use were acting within a shock-free continuity framework.

So Eurostar only has passenger volumes of 10mppa (about half of what was forecast). Whether new services will make a significant difference (Eurostar to Amsterdam later this year) and Deutsche Bahn to Frankfurt (announced 5 years ago) – i.e. more supply side changes – remains to be seen. But meanwhile Brexit threatens and border control anxieties grow, so the context for any market outlook has changed dramatically since the 1990s. These external changes would have seemed far-fetched in the early 1990s and not worth even a sensitivity test, not that the requisite variables featured in the demand models.

Turning from the specifics of Eurostar to the question of infrastructure use, CTRL now carries 20m passengers/year – there are now 10mppa using South Eastern's high-speed services (alongside the 10mppa on Eurostar). This reflects a supply side (train service) change but in the opposite direction to those affecting Eurostar. When the merit of CTRL was being debated, no domestic service plans had been developed. Moreover, the South Eastern high-speed demand looks set to continue to grow if sufficient additional capacity is provided (peak passengers stand from Ashford currently). As far as CTRL is concerned, it's a Lesson 2 case..

Lesson 2: Some demand forecasts are right for the wrong reasons….

A good example of this is the Borders railway where demand has exceeded expectations at some stations – massively so at Galashiels and Tweedbank – but has fallen well short at others (stations such as Eskbank and Newtongrange, closer to Edinburgh); overall ridership is somewhat ahead of first year forecasts, and this despite poor service reliability. The reasons for these discrepancies are perhaps easier to fathom than the Eurostar case.

The over-achievement at Galashiels/Tweedbank is likely to be due to a high number of trialists (a category often excluded from formal forecasts – their inclusion sometimes seen as 'scraping the barrel') and a significantly higher level of induced demand, partly tourist-based (and subject to fashion factors that are hard to estimate), but also because, taken in the round, the rail service *from* the remote Borders' towns represents a step change in quality for travel to central Edinburgh – and step changes are areas of intrinsic high risk in forecasting.

The reason why the close-to-Edinburgh demand is *lower* than expected are likely to be that housebuilding in these areas is behind schedule and the time-lag in changing patterns of commuting from existing residential areas takes a long time to work through (we estimated over 20 years in the case of the Bedford – St Pancras – Moorgate electrification scheme in a 1980s Steer Davies Gleave study for Network South East).

Lesson 3: ….and other forecasts are right despite everything going wrong!

Seen as a very high risk forecast at the time, the Virgin Trains bid for the West Coast franchise made 20 years ago projected a threefold increase in demand over the 15-year life of the franchise. Despite serious and highly publicised disruption during the renewal/upgrade programme in the 1999-2004 period; despite the late delivery of the planned transformational infrastructure upgrade and despite the technology on which faster speeds were predicated (a new cab-based train control system) being abandoned and the plan to operate at 140 mile/h also lost…the tripling of demand happened only one year later than projected in 1997.

Of course, not everything did go wrong and the long franchise timescale (15 years) was in practice a time horizon that would have seemed luxurious to (say) BR's management who were subject to annual budget-setting and it provided sufficient time to overcome problems.

When a business depends on hitting its demand and revenue targets, it is more likely to achieve them. Forecasting is not a pure science (where outcomes must be independent of key actors' behaviour).

Lesson 4: Markets change over time…

The majority of long distance rail travel in GB is now non-discretionary; thirty years ago two thirds of intercity travel *was* discretionary (holidays and visiting friends and relatives etc). Family travel was once a significant category; under-16s travelled unaccompanied and now rarely do; travel for sports and other events by rail was much more common; armed service travel by rail was also a significant category 30 years ago and has largely disappeared.

Business travel by rail has boomed as mobile comms/computing has made train travel highly productive. This effect has (so far) outweighed the substitution effect of using telecoms instead of travelling for face to face meetings, and probably helped create a more spatially dispersed professional/managerial workforce.

Looking at total demand, over time, trends may look stable, but they can be masking very large changes. Disaggregated forecasting, at the level of market segments, desirable though it may seem, may produce less accurate results.

Lesson 5: ….and so do market shares, although demand models regularly presume very limited modal switching

Over a period of 11-12 years from 1995, the rail market share of all travel over 25 miles (which is dominated by car, despite the existence of coach and air alternatives in some markets) grew from 8% to 14%. This is a very significant change, and it was hard to foresee. Between 1950 and 1997, rail travel in the UK had remained hovering around a total of 20bn passenger miles *per annum.* Since then it has grown to nearly 40bn.

Typically, demand models imply only modest levels of mode switching in the comparisons of do something *vs* do not which are used to underpin economic appraisals. So even with large changes, HS2 is reckoned to only attract 4% of its custom from car – attracting the complaint that 'it won't make much difference'.

But over time, it will make a difference. This is because a significant modal capacity increment, like HS2, will support a higher rail mode share than would otherwise occur. The models don't show this effect. HS2 demand is most likely under-estimated.

The lessons of history may or may not apply. Andrew Evans, analysing the effects of the WCML electrification project of the 1960s estimated that of the substantially increased volume of rail travel observed, roundly 50% was diverted from other modes and 50% was 'induced' – travel that would not have otherwise occurred (at least along the West Coast corridor). Induced (or generated) demand is often treated with scepticism (despite the evidence from the M25 for which qualitative research prior to its opening revealed expectations of small flows/no congestion (where would the traffic all come from?). Its economic value is also discounted by a loose consensus view of those engaged in transport policy that extra travel should be discouraged. But induced demand is significant in the case of many rail projects and policies.

In France, the huge success of the first TGV line in 1981 is generally attributed to the near-halving of the Lyon – Paris journey time to a nice memorable 2 hours. But qualitative research in France at the time (unpublished) showed that a major factor in mode switching was the introduction of a new seat reservation system, that could be changed/validated up to 5 minutes before the journey (a technological feat still not achieved in GB) and this could be done from home using the unique French Minitel system that used a domestic phone line: an internet capability, if you will, well before the world wide web.

Mode split models are not where the big shifts in mode use occur in forecasts such as those used for HS2: they arise from assumptions about 'base' demand in future years compared with today, where it is impossible to ignore the highly differentiated trends in travel trends by mode. These underlying mode-specific trend growth assumptions should be given much more scrutiny.

Lesson 6: Service frequency really matters – as do through trains

This was established through some ground breaking research commissioned by BR in 1980. I was part of the consulting team at Steer Davies Gleave that did the work using conjoint multivariate analysis. This introducing Stated Preference (SP) techniques for the first time into policy research in the UK. Stated Preference is a misnomer – at least if research is conducted the way that was used in this first 1980 research study, where a 'Journey Planning Game' tool was used to in an in-depth interview to replicate a structured set of travel choices, tailored to the very specific circumstances of each respondent's journey, including in terms of access to information and perceived choice sets.

The research showed that there were high levels of elasticity with respect to travel frequency - especially for business travellers and high levels of resistance to the need to change trains *en route* - especially for leisure travellers. This quantitative research, incorporated into the PDFH 'bible', underpinned the service uplifts and the transformation of ridership on the 'Other Provincial Services' that became the 'Regional Railways' sector.

Lesson 7: Changing mode car to rail ('park and ride') *en route* has not been accurately forecast (but could be)

In depth research with travellers shows high levels of anxiety about using station car parks, especially if they are remote from either end of the journey. Demand models of course may apply an interchange penalty, but they don't distinguish (assuming they have a mixed mode capability, which most do not) between proximate and non-proximate rail-head sites.

Major park and ride facilities that rely on a wide catchment, especially if combined with limited train frequencies that add to anxiety about return journeys, have in general under-achieved against expectations (Alfreton & Mansfield and East Midlands Parkways for example). Stations such as Bristol Parkway, on the other hand, where there are back-up taxis and multiple bus lines (so less anxiety) *and* high frequency train services that are quicker than those involving a city centre station *and* a generally closer catchment have out-performed expectations. In fact, Bristol Parkway is so far unique in demand generation. Non-Parkway stations such as Doncaster also perform well serving wider catchments because they offer choice and security for return travellers.

Sometimes, Lesson 2 type countervailing factors come to the rescue. Tiverton (mid Devon) Parkway has succeeded with a much higher than projected induced rail demand from nearby areas (Tiverton, for instance) and lower than projected numbers from further afield (North Devon, for example, it being a remote station for many candidate users).

Lesson 8: Fares elasticity effects are now buried in complexity

By the early 1990s, BR could predict the effects of fares level changes reasonably accurately, with measures disaggregated by journey purpose and trip length. But this was before yield management systems came onto the scene and a number of TOC-specific fares offers were overlaid on top of the already complex tapered mileage-based system of charges that BR used, complete with its set of railcard-based discounts. And it was also in an era of more normally distributed income levels so that

a single price-point had more traction. With increased income disparities have come much wider top and bottom fares to segment the market. Oh and then there's the deep discounts that split journey ticketing offers.

With fares level regulation half-applied, the scope to alter the applicability of fares restrictions case-by-case adds another layer of detail that makes it very difficult in a demand forecast to assume anything other than the same fares system applying in the do something/do not pair of cases. It just means that an investment partly predicated on a fares change, once considered by BR to be a near normality, for example, with a new train fleet, is impossible to forecast reliably and has thus fallen out of use: an odd case of analytical short-comings precluding a common-sense policy. Although there has been one major exception in recent years, namely the 30% premium applied to SouthEastern's fares of the high-speed service.

Rail fares: too complex for travellers, ditto modellers and regulators, yet commercial operators know how to work the system and each year they shuffle fares' baskets to ensure they deliver higher yield/passenger-km.

Lesson 9: the analytic presumption of demographic and land use stability has led to serious under-estimation of demand on new lines

But fortunately, poor benefit cost ratios (BCRs) and the opinion of rigid model-driven economists don't always win the day. Two examples…

The East London Line extensions project – together with a service plan change instigated by the SRA – led to the creation of the London Overground. Widely recognised as one of the major rail sector successes of the last 15 years, ridership has exceeded all expectations. The *ex-ante* models showed low levels of demand from a catchment demographic that made few journeys other than very local trips and were therefore unsuited to rail. As a result, the BCR was a bare 1:1 (and that was before construction costs climbed yet again). The argument used with Ministers of the day was that this was a project that served an arc of multiple deprivation through inner East and South East London (and it did).

Upon – and in expectation of – the project's implementation, these very catchment areas of the East London Line became fashionable and gentrified and travel increased hugely. Few saw it coming, and it would have been very difficult to make an investment case presuming that kind of dramatic change.

The better known example is the Jubilee Line Extension, again with a poor BCR (1.1:1). Opposed by DfT, and by TfL, and by the City of London Corporation, this scheme was forced through at Prime Minister level with the backing of Michel Heseltine who had championed Docklands regeneration. But the BCR and the demand forecasts (from TfL) presumed that there would be no change in land use in Docklands as a result of the project. Demand forecasts were low (indeed the London model used had no provision in its base case leave alone its do-something case, for Docklands growth: it had come from the Central London Rail Study (the clue is in the name)).

This analytical nonsense was apparent at the time to supporters of the project.

But not only was a very important project implemented despite a very poor demand forecast, it soon became clear that demand would be much higher than the models had projected. Meanwhile, the scope to secure even greater land value capture had slipped by: the models suggesting it wouldn't happen – a real price of analytical short-comings.

Lesson 10: We've run out of incremental capacity, so now the modelling gets much harder

Through the 1990s and 2000s, there was a lot of scope to increase rail capacity by lengthening trains and squeezing a few more services on to the network. But that scope is running out fast – especially in the south east. We are in a new era.

Even small levels of demand growth in the London commuter peaks now lead to extended journey times (at key stations such as Clapham Junction, the turnover of passengers on peak trains cannot be handled within the allotted station dwell times). On longer distance services, weekend peaks lead to lengthy journeys with standing passengers. In the North, more rolling stock can be accommodated more readily, but even so peak time pressures are significant.

Little is known about the feed-back effects in these circumstances. People assume that the advertised train service (on which they have made house and job location decisions) will indeed run and will accommodate them. But often trains at intermediate stations in the commuter peak have no spare capacity and cannot be boarded: journey times can be much greater than expected.

Travel adjustments are possible for some – working partly at home and on a less than a 5 day/week commute basis has grown substantially. So relationships between employment forecasts and peak travel demands will need to be revisited.

It is hard to deduce what the benefit is (and demand reaction to) a reduction in levels of overcrowding. This is rapidly becoming a key benefit of transport investment, yet its assessment is ill-served by models that seek to measure the effect of changes in the proportion of in-train times which are crowded – with such travel given an extra generalised cost weighting to reflect perceived discomfort.

The benefit metric is unclear and the effects of crowding relief on demand (including the release of congestion-suppressed demand) is very difficult to estimate. So this is an area needing much more research and development.

Jim Steer
January 2017

These notes are personal views and not the views of Steer Davies Gleave or of Greengauge 21.