Development of integrated demand and station choice models for local railway stations and services

Marcus Young
Transportation Research Group
University of Southampton
4 October 2017
The railway in Britain has experienced considerable growth in recent years.

Passenger journeys have doubled over the past 20 years.

58 new stations have opened in the last 10 years.

Data source: Office of Rail and Road

Data source: Railfuture
Growth looks set to continue, with new stations and lines under consideration and being campaigned for.

There could be 12 new railway stations built in Wales.

The locations under consideration include three in Cardiff and two each in Swansea and Wrexham.

28.07.17
DfT names five winners of fresh £16m stations fund.

New and potential line re-openings.

Source: www.railengineer.uk
Demand forecasts for new stations have sometimes not performed well

Borders Railway – forecast vs. actual

<table>
<thead>
<tr>
<th>Location</th>
<th>Trips (thousands)</th>
<th>Business Case</th>
<th>Year 1</th>
<th>2016/17</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tweedbank</td>
<td>43</td>
<td>474</td>
<td>338</td>
<td>47</td>
</tr>
<tr>
<td>Galashiels</td>
<td>47</td>
<td>202</td>
<td>342</td>
<td>12</td>
</tr>
<tr>
<td>Stow</td>
<td>12</td>
<td>48</td>
<td>66</td>
<td>48</td>
</tr>
<tr>
<td>Gorebridge</td>
<td>180</td>
<td>93</td>
<td>75</td>
<td>180</td>
</tr>
<tr>
<td>Newtonrange</td>
<td>106</td>
<td>97</td>
<td>137</td>
<td>106</td>
</tr>
<tr>
<td>Eskbank</td>
<td>261</td>
<td>133</td>
<td>228</td>
<td>261</td>
</tr>
<tr>
<td>Shawfair</td>
<td>124</td>
<td>17</td>
<td>21</td>
<td>124</td>
</tr>
</tbody>
</table>
Trip end models are typically used in the UK to forecast demand for new local stations.

### TABLE 3.1 SUMMARY OF DEMAND FORECASTING METHODOLOGY

<table>
<thead>
<tr>
<th>New Station/Line</th>
<th>Methodology Used</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alloa</td>
<td>No Information supplied</td>
</tr>
<tr>
<td>Aylesbury Vale Parkway</td>
<td>Trip rate and accessibility modelling (using HEs)</td>
</tr>
<tr>
<td>Chandlers Ford</td>
<td>Logit model, trip rate model and MOIRA</td>
</tr>
<tr>
<td>Coleshill Parkway</td>
<td>Trip rate model and logit mode choice</td>
</tr>
<tr>
<td>Corby</td>
<td>Trip rate, MOIRA and station access model</td>
</tr>
<tr>
<td>East Midlands Airport Parkway</td>
<td>GIS catchment analysis, elasticity based model 6 air</td>
</tr>
<tr>
<td></td>
<td>port mode share assumptions</td>
</tr>
<tr>
<td>Ebbw Valley Line</td>
<td>Logit model and uplift for trip generation</td>
</tr>
<tr>
<td>Edinburgh Crossrail</td>
<td>No Information supplied</td>
</tr>
<tr>
<td>Edinburgh Park</td>
<td>Trip rate and logit mode choice</td>
</tr>
<tr>
<td>Glasshoughton</td>
<td>Trip rate</td>
</tr>
<tr>
<td>Imperial Wharf</td>
<td>RAILPLAN strategic forecasting model</td>
</tr>
<tr>
<td>Larkhall-Milngavie</td>
<td>4 stage land use model</td>
</tr>
<tr>
<td>Laurencekirk</td>
<td>Trip rate</td>
</tr>
<tr>
<td>Liverpool South Parkway</td>
<td>Elasticity based model, airport accessibility model,</td>
</tr>
<tr>
<td></td>
<td>mode switch (logit) model</td>
</tr>
<tr>
<td>Mitcham Eastfields</td>
<td>Trip rate</td>
</tr>
<tr>
<td>Shepherds Bush</td>
<td>Trip rate</td>
</tr>
<tr>
<td>Vale of Glamorgan Line</td>
<td>Trip rate</td>
</tr>
<tr>
<td>Warwick Parkway</td>
<td>Parkway Access Model and Mode/Route Choice models</td>
</tr>
</tbody>
</table>

2010 report for DfT

- LLANTWIT MAJOR & RHOOSE
- SHEPHERDS BUSH
- MITCHAM EASTFIELDS
- GLASSHOUGHTON
- EDINBURGH PARK
- COLESHILL PARKWAY
- CHANDLERS FORD
- AYLESBURY VALE PARKWAY

% DIFFERENCE FROM FORECAST

-60  -30  0  30  60  90  120  150
This presentation will focus on the development of improved national trip-end models.

Addressing the station catchment problem

Calibrating a national trip end model using probabilistic catchments

Applying the model to generate forecasts
This research builds on previous work to calibrate a national trip-end model for local stations

\[ \ln \hat{V}_i = \alpha + \left( \ln \sum_{z} P_z w_z \right)^\beta + \ln F_i^\gamma + \ln T_i^\delta + \ln J_i^\xi + \ln P_k^\eta_i + T e_i^\kappa + E l_i^\nu + B_i^\tau \]

\( V \) = annual entries/exits for Category E & F stations (England and Wales)
\( Z \) = all zones where the closest station is station \( i \)
\( z \) = census output area

\( P = \) Resident population
\( w = \) decay function (distance/time based)
\( F = \) Weekday train frequency
\( T = \) Distance to nearest Cat A-D station
\( J = \) Jobs within 2 minute drive of station
\( P_k = \) No. of parking spaces
\( B = \) Travelcard boundary (y/n)
\( T e = \) Terminus station (y/n)
\( E l = \) Served by electric trains (y/n)
In reality station catchments are more complex entities, they overlap and stations compete.
Revised model uses postcode as zone and weights population by probability using a station choice model

\[ \ln \hat{V}_i = \alpha + \left( \ln \sum_{z}^{Z} Pr_z P_z w_z \right)^{\beta} + \ln F_i^{\gamma} + \ln J_{it}^{\eta} + \ln P k_i^{\eta} + T e_i^{\kappa} + E l_i^{\nu} + B_i^{\tau} \]

\( Z = \) all zones with station \( i \) in their choice set
\( z = \) unit level postcode (1.4 million)

Calibration dataset - 1,792 Category E and F stations in mainland GB
A station choice model was calibrated using 14,422 observations from on-train passenger surveys

\[ P_{r_{nik}} = \frac{\exp(N_k^\beta + \sqrt{D_{ik}}^\gamma + U_k^\delta + \ln F_k^\zeta + C_k^n + P_k^\mu + T_k^\kappa + B_k^\lambda + \ln A_k^\mu)}{\sum_{k=1}^{K} \exp(N_k^\beta + \sqrt{D_{ik}}^\gamma + U_k^\delta + \ln F_k^\zeta + C_k^n + P_k^\mu + T_k^\kappa + B_k^\lambda + \ln A_k^\mu)} \]

\( K = 10 \) nearest stations to each postcode

Pseudo R\(^2\) = 0.71
A postcode probability table was then generated for every postcode in mainland GB.

For each unit postcode:

1. Find 10 nearest stations (the choice set).
2. Generate predictor variables for each station.
3. Calculate probability of each station being chosen.
A railway station choice predictor app was created using R Shiny to aid interpretation.

Railway station choice predictor

Enter Postcode

B74 4PX

Select model

Combined (TE24)

Predict station choice

The nearest station to B744PX is BUL and the nearest major station is BSW (by distance).

<table>
<thead>
<tr>
<th>station code</th>
<th>station</th>
<th>probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>BKT</td>
<td>BLAKE STREET</td>
<td>0.04</td>
</tr>
<tr>
<td>FOK</td>
<td>FOUR OAKS</td>
<td>0.25</td>
</tr>
<tr>
<td>BUL</td>
<td>BUTLERS LANE</td>
<td>0.20</td>
</tr>
<tr>
<td>SUT</td>
<td>SUTTON COLDFIELD</td>
<td>0.14</td>
</tr>
<tr>
<td>WYL</td>
<td>WYLDE GREEN</td>
<td>0.03</td>
</tr>
<tr>
<td>CRD</td>
<td>CHESTER ROAD</td>
<td>0.02</td>
</tr>
<tr>
<td>ERD</td>
<td>ERDINGTON</td>
<td>0.01</td>
</tr>
<tr>
<td>SEN</td>
<td>SHENSTONE</td>
<td>0.00</td>
</tr>
<tr>
<td>PRY</td>
<td>PERRY BARR</td>
<td>0.00</td>
</tr>
<tr>
<td>HSD</td>
<td>HAMSTEAD (BIRMINGHAM)</td>
<td>0.00</td>
</tr>
</tbody>
</table>
A probabilistic catchment was then defined for every station in the calibration dataset

\[
\ln \hat{V}_i = \alpha + \left( \ln \sum_z P_r P_z w_z \right)^{\beta} + \ln F_i^\gamma + \ln J_{it}^\eta + \ln P_k^\eta_i + T e_i^\kappa + E l_i^\nu + B_i^T
\]

For each station

Identify all postcodes with this station in choice set

Weight population of each postcode by the station’s probability and \( w \)

Sum weighted population for all postcodes

Trip end model calibration input
The trip end model was calibrated using both deterministic and probabilistic catchments

\[ R^2 = 0.851, \text{ AIC} = 3763 \]

\[ R^2 = 0.843, \text{ AIC} = 3848 \]

\[ R^2 = 0.822 \text{ (different dataset)} \]
A methodology was then developed to generate demand forecasts for new stations using the new models.

1. Identify unit postcodes within 60 minutes of proposed station.
2. Find 10 nearest stations for each postcode.
3. Retain postcodes with proposed station in choice set.
4. Calculate probabilities using station choice model.
5. Run trip end model to generate forecast.
Case study: forecasts were generated for the Borders Line which opened in 2015

Map of new line (source: Wikipedia)
Various catchment representations for Tweedbank station

Probabilistic catchment for Tweedbank

Observed catchments from Transport Scotland user survey

Deterministic catchment for Tweedbank
Various catchment representations for Galashiels station

- Probabilistic catchment for Galashiels
- Observed catchments from Transport Scotland user survey
- Deterministic catchment for Galashiels
Under-forecast for Galashiels may be due to new parking, bus interchange and tourism

New interchange at Galashiels – 1400 bus departures per weekday

Galashiels’ gateway to the Borders

New 43 space Pay and Display car park opposite interchange

![Graph showing mode of transport](image)

Source: Borders Railway Year 1 Evaluation, June 2017
A methodology has also been developed to model abstraction from existing stations.

\[ \text{Trips}_{\text{old}} \times \left( \frac{\text{Pop}_{w_{\text{prob} \text{ w_{dist} \text{ new}}}}}{\text{Pop}_{w_{\text{prob} \text{ w_{dist} \text{ old}}}}} \right)^{1} = \text{Trips}_{\text{new}} \]
In conclusion, results suggest it is possible to develop a robust and transferable national forecasting model. This may be preferable to models developed on an ad hoc basis. Useful ‘sense-check’ model to assess reliability of local models. Opportunity to extend approach to flow models.