Published paper

Lythgoe, W.F., Wardman, M., Toner, J.P. (2004) Enhancing Rail Passenger Demand Models to examine Station Choice and Access to the Rail Network - AET European Transport Conference, 4-6 October 2004, Strasbourg.

ENHANCING RAIL PASSENGER DEMAND MODELS TO EXAMINE STATION CHOICE AND ACCESS TO THE RAIL NETWORK

W F Lythgoe
Institute for Transport Studies, University of Leeds, UK
M Wardman
Institute for Transport Studies, University of Leeds, UK
J P Toner
Institute for Transport Studies, University of Leeds, UK

1 INTRODUCTION AND OBJECTIVES

Much analysis of rail travel demand in Great Britain has been undertaken using time-series direct demand models, for example Jones and Nichols (1983), and Owen and Phillips (1987). In these models, changes in demand over time are explained as a function of independent variables that change incrementally over the same time period. However, such an incremental approach is of no use for forecasting the demand from new stations, or for other new rail services. Furthermore, this approach does not handle competition between different stations, nor the impact of access on either rail demand or rail elasticities.

There is, therefore, a need for cross-sectional models which can forecast demand for journeys from new stations, or in response to population changes, changes in station accessibility or radical service quality changes. Previous examples include Tyler and Hassard (1973), Holt and White (1981), Shilton (1982), Jones and White (1994), and Wardman (1996). These authors were unable, for obvious reasons, to take advantage of the new opportunities for developing such models which have been presented by the increased availability of machine-readable Geographical Information Systems (GIS) data on populations and road networks. Such data can be combined with data on rail passenger flows and revenues and on rail service quality, these latter data being those already used to develop time-series models.

Arising on the growth of computing power, a further opportunity is now presented for potentially more sophisticated cross-sectional models, which may not be amenable to linear regression, to be calibrated using non-linear regression.

Some initial attempts to build more sophisticated models have already been reported (Lythgoe and Wardman, 2002; 2004). The objective of this paper is to generalise the station choice model from the earlier work and to show how various limitations have been overcome. There is an emphasis on replacing the MNL station choice form by a particular cross-nested logit form, with different dissimilarity parameters between given station i and each of its competing stations. Introducing such a cross-nested logit form enables the proportion of new journeys from station i abstracted from its competitors to be dependant, inter alia, on the proximity of station i to each of those competitors.

In particular, what we propose here is an improvement in that the previous model could only be applied to a subset of origin stations, namely Parkway stations. Also, when fitting the data, a specification error that had been previously identified is remedied by introducing a population elasticity.

The origin station choice model described in this paper builds on the original Parkway station model (Lythgoe and Wardman, 2002; 2004) and can predict the demand for inter-urban rail journeys of over 40km between pairs of stations in Great Britain. It is based upon 10,324 observed demand levels from 329 existing stations to 334 destination stations. The aim is that it should be more straightforward to apply than existing techniques for forecasting demand from new or greatly revised stations and services, and that it should provide consistent results.

2 BACKGROUND: SUMMATION MODELS

Before the station choice model described here was developed, rail travel had been analysed using a series of simpler cross-sectional direct demand models. The most recent of these was the 'summation model' and this model informs the derivation and refinement of the origin station choice models which are discussed later.

In a summation model, the populations around the origin and destination stations are allocated to zones. The number of journeys between an origin zone a and a destination zone b, using rail between stations i and j, is given by:

$$Q_{aijb} = Kp_a F_{ai} p_b F_{jb} F_{ij} \qquad (1)$$

where:

Q_{aijb} is the number of journeys from origin zone a to destination zone b

p_a, p_b are the populations in zone a and zone b, respectively

F_{ai} is a function of the utility of access, U_{ai}, from zone a to the origin station i

F_{jb} is a function of the utility of egress, U_{jb}, from the destination station j to zone b

F_{ij} is a function of the utility of the rail journey, U_{ij}, from station i to station j

The number of journeys between origin station i and destination station j, is derived by summing the numbers of journeys between origin zones a and destination zones b, so that:

$$Q_{ij} = \sum_a \sum_b Q_{aijb} = K \left(\sum_a p_a F_{ai} \right) \left(\sum_b p_b F_{jb} \right) F_{ij} \qquad (2)$$

It has been found that population elasticities are needed in the summation model when it is estimated to the data, so that:

$$Q_{ij} = K \left(\sum_a p_a F_{ai} \right)^{\delta_i} \left(\sum_b p_b F_{jb} \right)^{\delta_j} F_{ij} \qquad (3)$$

A hybrid summation model can combine a summation model at the origin, and a dummy parameter for the destination station j, so that the number of journeys is given by:

$$Q_{ij} = K \left(\sum_a p_a F_{ai} \right)^{\delta_i} B_j F_{ij} \qquad (4)$$

Reference will be made to this model as the origin station choice models are developed below.

3 THE NEW MODEL: ORIGIN STATION CHOICE MODELS

The origin station choice models, which are the focus of this paper, are hybrid models which combine station choice at the origin, and a dummy parameter for the destination station.

Figure 1 shows the choices for a potential traveller who might choose to travel from origin zone a to destination station j. The annual number of journeys from zone a, via station i, to destination station j, is given by:

$$Q_{aij} = n p_a P_{aij} \qquad (5)$$

where:

Q_{aij}	is the number of journeys from zone a, via station i, to destination station j
n	is the (unknown) number of travel choices per year
p_a	is the population in zone a

P_{aij} is the probability of an individual resident in zone a choosing to travel via station i, to destination station j

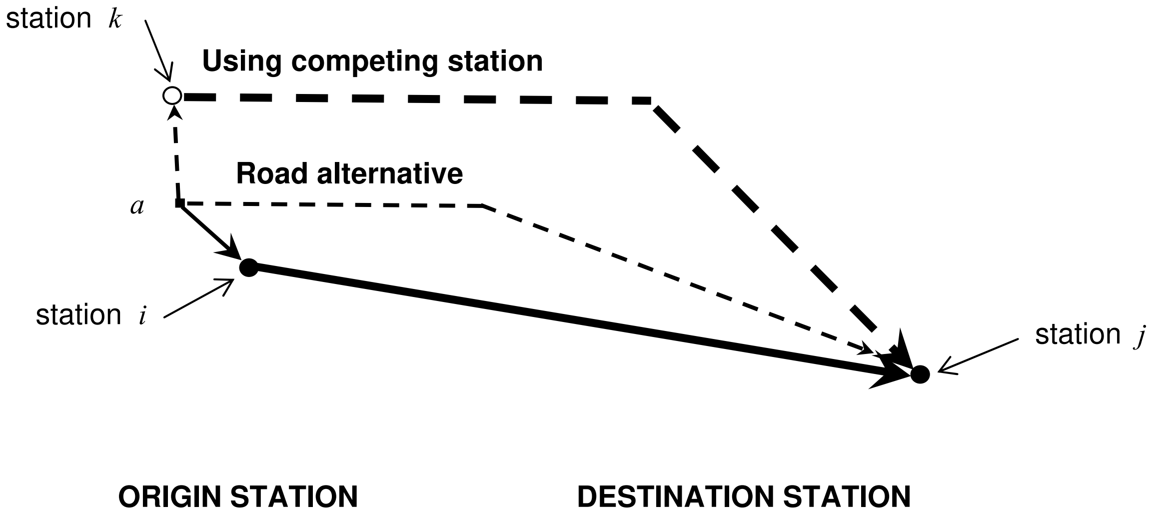

Figure 1 **General model - journeys from origin zone a to destination station j**

Equation (5) can be re-written as:

$$Q_{aij} = np_a P_{aij|a*j} P_{a*j} \quad (6)$$

where:

$P_{aij|a*j}$ is the probability of an individual resident in zone a choosing to travel via station i, to destination station j given that rail will be used (the station choice form)

P_{a*j} is the probability of an individual resident in zone a choosing to travel by rail via any origin station to destination station j

Note that the asterisk (*) used in the subscripts in Equation (6) is used as a 'wild card'; in this case indicating travel by rail via any origin station.

The probability P_{a*j} is determined empirically when the models are estimated. At this stage, for clarity, this probability can be assumed to take the following form:

$$P_{a*j} = \frac{1}{n} K F_i e^{U_{a*j}} F_{\overline{a*j}} B_j \quad (7)$$

where:

K	is a constant (to be estimated)
F_i	is a factor which is applied for origin station i
U_{a*j}	is the composite utility of travelling from zone a by rail via any station to destination station j
$F_{\overline{a*j}}$	is a function of the composite utility, $U_{\overline{a*j}}$, of <u>not</u> travelling from zone a by rail via any station to destination station j (i.e. travelling by an alternative mode from zone a to destination station j, or not travelling at all)
B_j	is the dummy parameter (to be estimated) for destination station j

Note that the bar ($-$) used over the subscript in Equation (7) is used to negate the usual meaning of the subscript.

Substitute Equation (7) into (6) to give:

$$Q_{aij} = KF_i p_a P_{aij|a*j} e^{U_{a*j}} F_{\overline{a*j}} B_j \tag{8}$$

The number of journeys between origin station i and destination station j, is derived by summing the numbers of journeys between origin zones a and destination station j, so that:

$$Q_{ij} = \sum_a Q_{aij} = KF_i \left(\sum_a p_a P_{aij|a*j} e^{U_{a*j}} F_{\overline{a*j}} \right) B_j \tag{9}$$

It was noted above (in Section 2) that population elasticities are required when models described in that section are estimated to the data. By an analogy to the hybrid summation model defined by Equation (4), an origin population elasticity can be introduced into the origin station choice model by setting:

$$F_i = {A_L}^{O_L} \left(\sum_a p_a F_{ai} \right)^{\delta_i - 1} \tag{10}$$

where:

A_L is the dummy parameter (to be estimated) for London as origin station i

O_L is dummy variable which takes the value 1, if London is the origin station, or 0 otherwise

F_{ai} is a function of the utility of access, U_{ai}, from zone a to the origin station i

δ_i is the origin population elasticity (to be estimated)

Substitute Equation (10) into Equation (9) to give:

$$Q_{ij} = KA_L{}^{O_L}\left(\sum_a p_a F_{ai}\right)^{\delta_i - 1}\left(\sum_a p_a P_{aij|a*j} e^{U_{a*j}} F_{\overline{a*j}}\right) B_j \qquad (11)$$

Two alternative forms for the station choice, $P_{aij|a*j}$, are discussed below (in Sections 4 and 5).

4 MULTINOMIAL LOGIT STATION CHOICE

The most obvious form for the station choice is Multinomial logit (MNL), as represented in Figure 4. This form is the same as that used in Lythgoe and Wardman (2002; 2004). The MNL station choice form is:

$$P_{aij|a*j} = \frac{\left(e^{U_{aij}}\right)^{\frac{1}{\mu}}}{\left(e^{U_{a*j}}\right)^{\frac{1}{\mu}}} \qquad (12)$$

given the useful shorthand definition:

$$\left(e^{U_{a*j}}\right)^{\frac{1}{\mu}} = \sum_k \left(e^{U_{akj}}\right)^{\frac{1}{\mu}} \qquad (13)$$

where:

U_{aij} is the utility of travelling from zone a by rail via origin station i to destination station j

U_{akj} is the utility of travelling from zone a by rail via competing station k to destination station j (the set of competing stations k includes origin station station i)

U_{a^*j} is the composite utility of travelling from zone a by rail via any station to destination station j

μ is the dissimilarity parameter (to be estimated) between choices of origin station

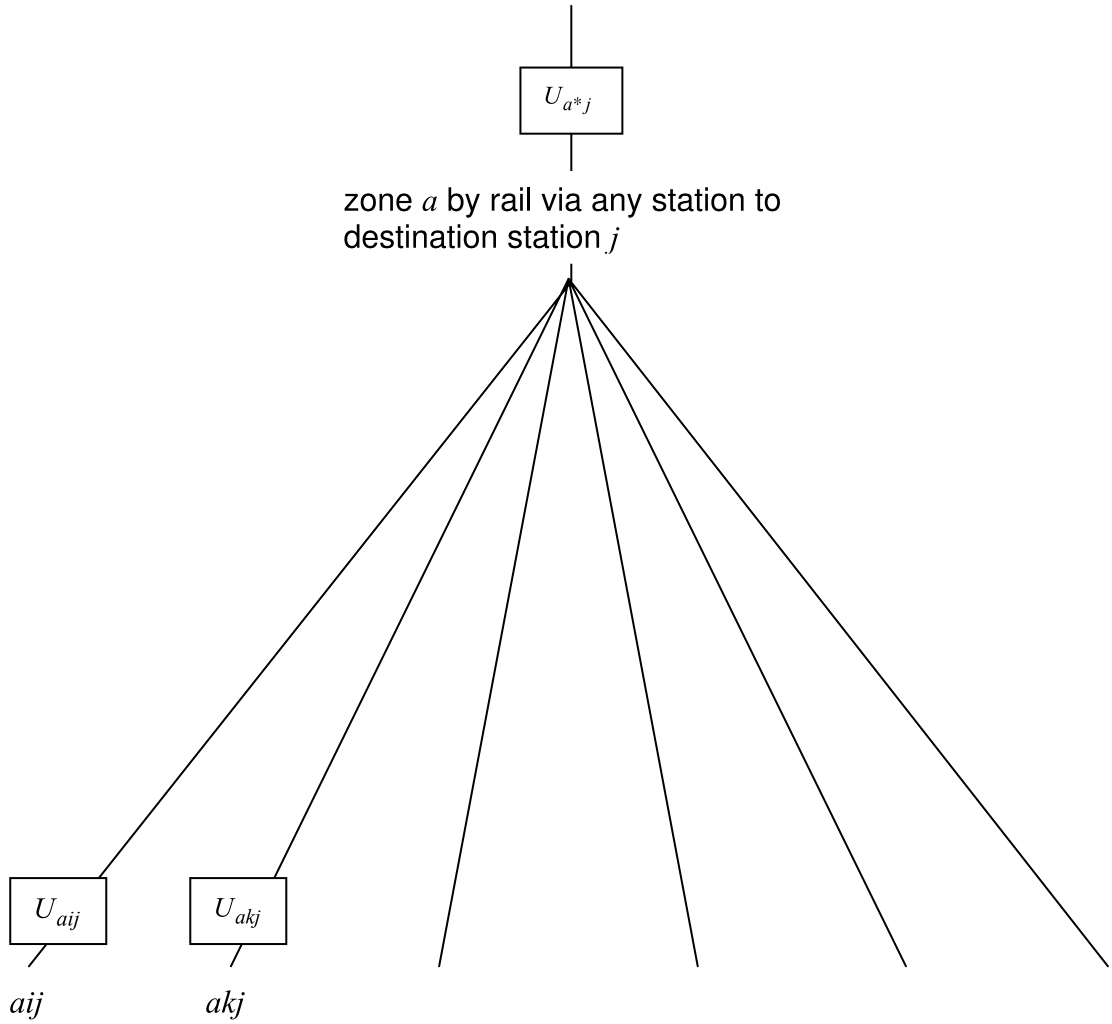

Figure 4 Multinomial logit origin station choice showing utility notation

Substitute Equation (12) into Equation (11):

$$Q_{ij} = KA_L{}^{O_L}\left(\sum_a p_a F_{ai}\right)^{\delta_i - 1}\left(\sum_a p_a \frac{\left(e^{U_{aij}}\right)^{\frac{1}{\mu}}}{\left(e^{U_{a^*j}}\right)^{\frac{1}{\mu}}} F_{\overline{a^*j}}\right) B_j \qquad (14)$$

A 'generation ratio' is defined as the proportion of newly generated journeys to the total increase in journeys. An 'incremental generation ratio' ρ can be defined as the ratio of the incremental change in the probability P_{a*j} to the incremental change in the probability P_{aij}, at a given value of P_{aij}. The incremental generation ratio is given by:

$$\rho = \frac{\eta_{P_{a*j},U_{aij}}}{\eta_{P_{aij},U_{aij}}} \frac{P_{a*j}}{P_{aij}} \tag{15}$$

where:

$\eta_{P_{a*j},U_{aij}}$ is the elasticity of P_{a*j} with respect to U_{aij}

$\eta_{P_{aij},U_{aij}}$ is the elasticity of P_{aij} with respect to U_{aij}

When, for all a, $P_{aij} = 0$, then $Q_{ij} = 0$, and the incremental generation ratio is given by:

$$(\rho)_{Q_{ij}=0} = \mu P_{\overline{a*j}} \tag{16}$$

However, it can be assumed that:

$$P_{\overline{a*j}} \approx 1 \tag{17}$$

so that:

$$(\rho)_{Q_{ij}=0} \approx \mu \tag{18}$$

5 CROSS-NESTED LOGIT STATION CHOICE

The MNL station choice form discussed immediately above leads to a constant generation ratio ρ which is independent of the proximity of station i to any of its competitors. This was considered to be a deficiency. In order to overcome this problem a particular form of cross-nested logit station choice form, as represented in Figure 5, was devised. This cross-nested logit station choice form is:

$$P_{aij|a*j} = \sum_{k \neq i} \frac{\left(\alpha_{ik} e^{U_{aij}}\right)^{\frac{1}{\nu_{ik(a)}}}}{\left(e^{U_{a\left[\substack{i\\k}\right]j}}\right)^{\frac{1}{\nu_{ik(a)}}}} \frac{\left(e^{U_{a\left[\substack{i\\k}\right]j}}\right)^{\frac{1}{\mu}}}{\left(e^{U_{a*j}}\right)^{\frac{1}{\mu}}} \tag{19}$$

given the useful shorthand definitions:

$$\left(e^{U_{a\left[\substack{i\\k}\right]j}}\right)^{\frac{1}{\nu_{ik(a)}}} = \left(\alpha_{ik} e^{U_{aij}}\right)^{\frac{1}{\nu_{ik(a)}}} + \left(e^{U_{akj}}\right)^{\frac{1}{\nu_{ik(a)}}} \tag{20}$$

$$\left(e^{U_{a*j}}\right)^{\frac{1}{\mu}} = \sum_{k \neq i} \left(e^{U_{a\left[\substack{i\\k}\right]j}}\right)^{\frac{1}{\mu}} \tag{21}$$

where:

$U_{a\left[\substack{i\\k}\right]j}$	is the utility of travelling from origin a using rail from either station i, or station k (a 'choice pair'), to station j
$\nu_{ik(a)}$	is the dissimilarity parameter between station i and station k, (the choice pair), given that the journey starts at origin a
μ	is the dissimilarity parameter (to be estimated) between choices of choice pairs
α_{ik}	is an allocation parameter applied to the use of origin station i, which is used to distribute its probability to each of the choice pairs

Note that the square brackets ($[\ ]$) used in the subscripts in Equations (19) to (21) are used to indicate journeys via either of two stations.

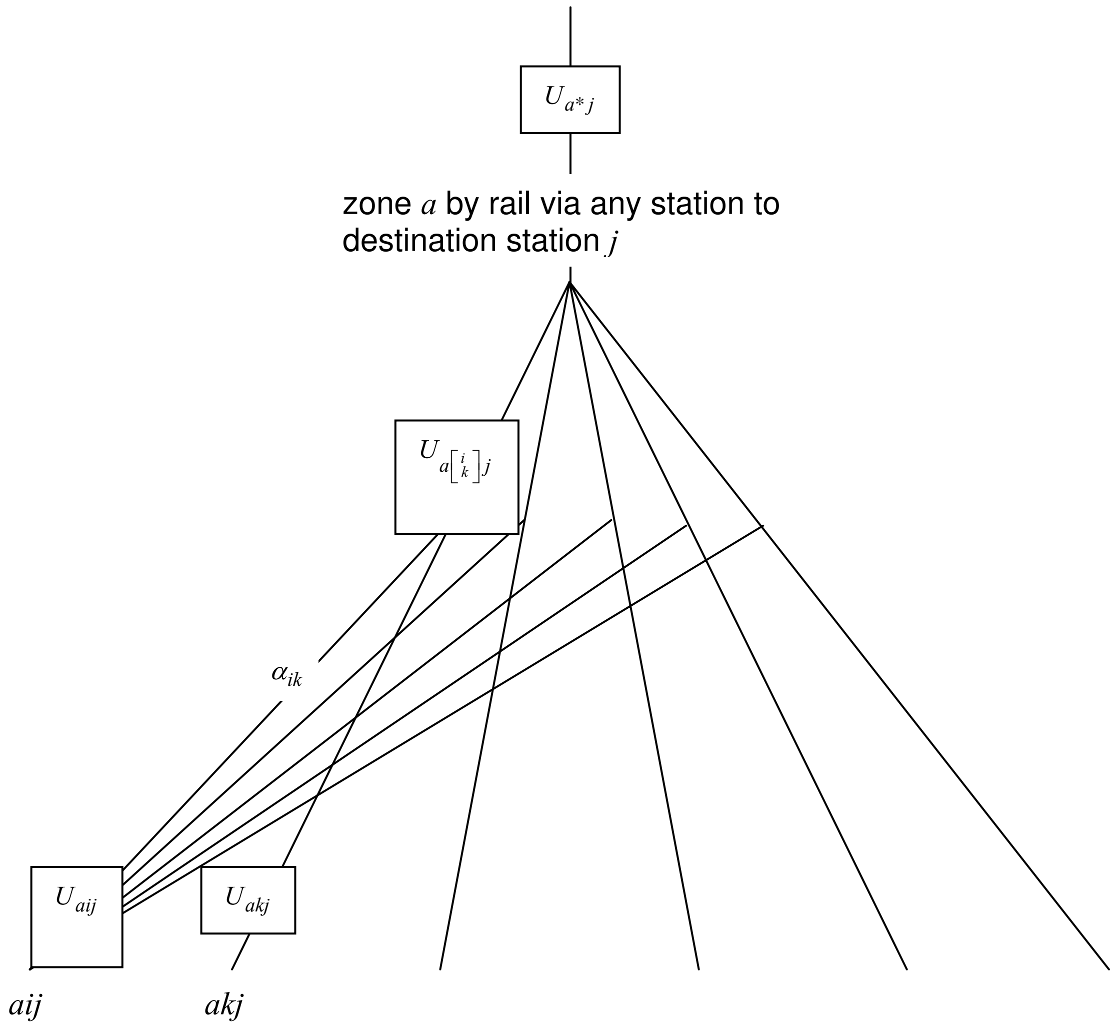

Figure 5 Cross-nested logit origin station choice showing utility notation (α_{ik} is an example allocation parameter)

In a more general context, this can be regarded as a GNL model (Koppelman and Sethi, 2000) with a nest for each competing station $k \neq i$, such that:

$$0 < \alpha_{ik} \leq 1 \tag{22}$$

$$\sum_{k \neq i} (\alpha_{ik})^{\frac{1}{\mu}} = 1 \tag{23}$$

$$\alpha_{kk} = 1 \tag{24}$$

$$\alpha_{k'k} = 0 \qquad \text{(if } k' \neq i,\ k' \neq k\text{)} \tag{25}$$

where:

$\alpha_{k'k}$ apportions the probability of the use of station k' to the nest for station k (in general, the set of k' includes both k and i)

Substitute Equation (19) into Equation (11):

$$Q_{ij} = KA_L{}^{O_L} \left(\sum_a p_a F_{ai} \right)^{\delta_i - 1} \left(\sum_a p_a \sum_{k \neq i} \frac{\left(\alpha_{ik} e^{U_{aij}} \right)^{\frac{1}{\nu_{ik(a)}}}}{\left(e^{U_{a\left[\substack{i\\k}\right]j}} \right)^{\frac{1}{\nu_{ik(a)}}}} \frac{\left(e^{U_{a\left[\substack{i\\k}\right]j}} \right)^{\frac{1}{\mu}}}{\left(e^{U_{a*j}} \right)^{\frac{1}{\mu}}} F_{\overline{a*j}} \right) B_j \tag{26}$$

When, for all zones a, $P_{aij} = 0$, then $Q_{ij} = 0$, and the incremental generation ratio is given by:

$$(\rho)_{Q_{ij}=0} = \nu P_{\overline{a*j}} \tag{27}$$

where:

ν is 'a composite dissimilarity parameter' (this is a weighted harmonic mean of the $\nu_{ik(a)}$ across all k and all a)

However, again it can be assumed that:

$$P_{\overline{a*j}} \approx 1 \tag{28}$$

so that:

$$(\rho)_{Q_{ij}=0} \approx \nu \leq \mu \tag{29}$$

Although μ is to be estimated by regression, Figure 6 shows how $\nu_{ik(a)}$ might be part calculated prior to estimation using costs C_{ai}, C_{ak} and C_{ik} as separation measures between a, i and k, so that:

$$\nu_{ik(a)} = f\left(C_{ai}, C_{ak}, C_{ik} \right) \tag{30}$$

where $f\left(C_{ai}, C_{ak}, C_{ik}\right)$ increases monotonically with C_{ik} and where:

$$\left.\begin{aligned} & f\left(C_{ai}, C_{ak}, 0\right) = f\left(C_{ai}, C_{ai}, 0\right) = 0 \\ & f\left(C_{ai}, C_{ak}, C_{ai} + C_{ak}\right) = \mu \end{aligned}\right\} \tag{31}$$

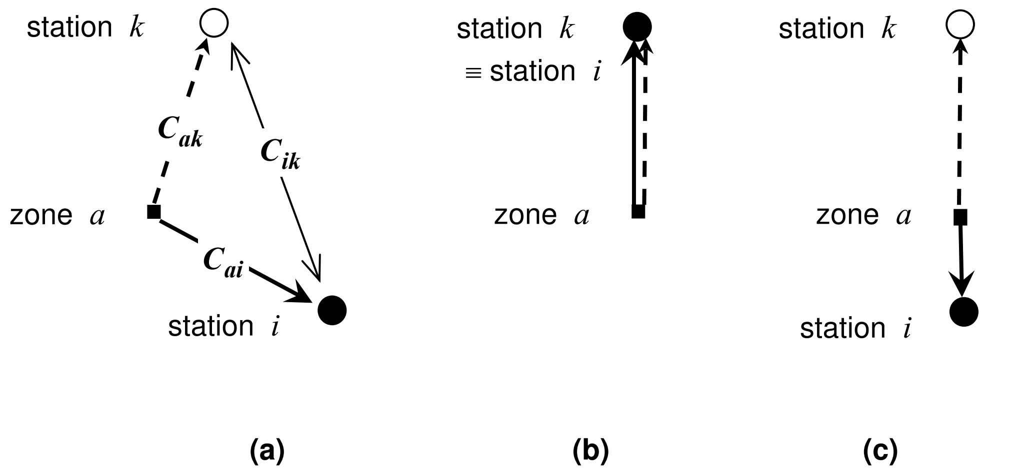

Figure 6 **Minimum and maximum values of dissimilarity parameter $\nu_{ik(a)} = f\left(C_{ai}, C_{ak}, C_{ik}\right)$:**
(a) illustration of journey costs C_{ai}, C_{ak} and C_{ik} between a, i and k.
(b) arrangement of a, i and k when $\nu_{ik(a)} = 0$ ($C_{ai} = C_{ak}$, $C_{ik} = 0$)
(c) arrangement of a, i and k when $\nu_{ik(a)} = \mu$ ($C_{ik} = C_{ai} + C_{ak}$)

6 DATA SOURCES AND PREPARATION

In Great Britain, the rail industry's CAPRI (Computer Analysis of Passenger Receipts and Revenue) system provides the numbers of journeys between pairs of stations and has been used for many years for the development of robust rail demand models. This study used CAPRI data for the numbers of journeys and the revenues for full, reduced and season tickets for all 1.4 million GB rail passenger flows in financial year 1999/2000. In this exercise, season tickets were excluded.

The stations were chosen for this study by adding revenues for journeys originating and terminating at each station, then sorting the stations in descending order of these total revenues. For the top 438 stations sorted in this way, Generalised Journey Times (GJTs) for 1999/2000 for flows between pairs of stations from this subset were obtained from another system known as MOIRA.

Populations from the 1991 census were obtained through MIMAS/ Casweb at enumeration district (ED) level for England and Wales, and output area (OA) level for Scotland, together with the Ordnance Survey grid references of their 'centroids'. There is an average of 181 households in an ED, and 53 households in an OA (Leventhal et al, 1993).

The road network for the whole of Great Britain was downloaded in the form of 1:250,000 Ordnance Survey 'Strategi' tiles from EDINA/ Digimap, and converted to a MapInfo Geographical System (GIS) compatible format.

The ticket sales data from CAPRI for all station pairs were merged with the GJT data from MOIRA for journeys between pairs of stations from the station subset explained above to produce a 'data superset' with 152,949 observations. These data cover 63% of all full and reduced ticket revenues, and 40% of all full and reduced ticket journeys. When journeys less than 40km are excluded, the 'data superset' is reduced to 146,292 observations.

A FORTRAN program was used to create a zonal structure for the 20km around each station, and to generate zonal populations. The program also produced an ordered list of competing origin stations for each origin station. As well as providing data for model estimation, this program can also be used to generate data for forecasting purposes. Further details of this process are reported in Lythgoe (2004).

Potential competing stations are defined as those within 20km of at least one origin zone. Thus a competing station may be up to 40km from the origin station and, in consequence, destinations must be at least 40km away from the origin station. This is, of course, in line with the existing criterion that journeys less than 40km are excluded. Candidate competitor stations are ordered by criteria calculated from the product of the total number of journeys originating at the candidate station and the population of a zone, divided by the distance to the centre of population for that zone, then summing across all zones for the origin station. The candidate stations are sorted in decreasing order of these criteria, and the top 15 form an ordered list of competing stations.

Road distance and time matrices between two sets of locations have been estimated using road network data and road speeds for a series of road types. The road speeds that were used in this study are listed in Table 1. These are based on default values used in Autostreet Navigator (ISYS Systems Ltd). Software developed at the Institute for Transport Studies for the purpose was used to calculate distances and times from zonal centres of population to the origin station and all competing stations, from the origin station to all competing stations, and from zonal centres of population to the destination station.

Table 1 Road types and corresponding speeds used in calculating access times and costs from population zones to stations

Road type	Description	Road speed (kph)
1	Motorway	112
2	Primary dual carriageway	96
3	Primary single carriageway	80
4	Primary narrow	72
5	A road dual carriageway	80
6	A road single carriageway	72
7	A road narrow	64
8	B road dual carriageway	80
9	B road single carriageway	64
10	B road narrow	56
11	Minor road > 4m	48
12	Minor road < 4m	48

(based on default values used in Autostreet Navigator)

Building the observations on origin stations and up to 15 competitors in this way should provide 146,292 observations available for estimation, but the inclusion of populations, and road distance and time matrices, would have led to potentially unmanageable data sets and excessive processing times in order to estimate the more complex models. Therefore, it was decided that a subset of observations would be created by excluding flows with small numbers of journeys. The top 10% of the original 152,949 flows gives 15,295 flows with at least 462 journeys per year. When flows of less than 40km are excluded from this top decile, the number of observations is further reduced to 12,252.

Finally, only 334 of the 438 stations for which the observations had been generated were used due to further considerations which are not elaborated here. This reduced the number of flows in the final subset of observations to 10,324.

Generalised costs were calculated within the SAS programs (SAS Institute Inc., 1999). The value of time formula from the Passenger Demand Forecasting Handbook (PDFH) (ATOC, 2002), which cites Wardman (2001), was used to calculate the money costs of both rail GJTs and road journey times. This value of time varies with journey distance and, in this study, it was calculated using total journey distance; in other words the sum of the access distance to the origin station and the rail journey distance. The fare is added to the GJT multiplied by the value of time to give the generalised cost for the rail journey. For road journeys, the distances are multiplied by a notional but plausible car cost of 7 pence per km and added to the time multiplied by the value of time to give their generalised costs.

7 MODEL ESTIMATION

The models are estimated by using non-linear least squares on the logarithm of the model form. The following empirical forms are found to provide a good fit:

$$F_{ai} = e^{\theta_{a*}C_{ai}} \tag{32}$$

$$U_{aij} = \theta_{a*}C_{ai} + \gamma_{*j} \log C_{ij} \tag{33}$$

$$U_{akj} = \theta_{a*}C_{ak} + \gamma_{*j} \log C_{kj} \tag{34}$$

$$F_{\overline{a*j}} = e^{\theta_{aj}L_{aj}} \tag{35}$$

where:

C_{ai} is the road journey generalised cost between zone a and station i (access cost)

C_{ak} is the road journey generalised cost between zone a and station k (access cost)

C_{ij} is the rail journey generalised cost between station i and station j

C_{kj} is the rail journey generalised cost between station k and station j

L_{aj} is the road distance between zone a and station j

$\left.\begin{matrix}\gamma_{*j}\\ \theta_{a*}\\ \theta_{aj}\end{matrix}\right\}$ are model parameters (to be estimated)

The following empirical form was used for a part calculation of the cross-nested logit allocation parameters:

$$\alpha_{ik} = \left(\frac{e^{\theta_{ik}L_{ik}}}{\sum_{k'} e^{\theta_{ik}L_{ik'}}} \right)^{\mu} \tag{36}$$

where:

L_{ik}	is the road distance between station i and station j
θ_{ik}	is a model parameter (to be estimated)

The following empirical form was used for a part calculation of the cross-nested logit dissimilarity parameters:

$$\nu_{ik(a)} = \left(\frac{2T_{ik}}{T_{ai} + T_{ak} + T_{ik}} \right)^{\phi} \mu \tag{37}$$

where:

T_{ai}	is the road journey time between zone a and station i
T_{ak}	is the road journey time between zone a and station k
T_{ik}	is the road journey time between station i and station k
ϕ	is a model parameter (to be estimated)

Tables 2 and 3 show the parameters estimated for both models.

In Table 2, γ_{*j} is a rail journey generalised cost elasticity, and its estimated value is in line with expectations. For the cross-nested model, μ is the maximum dissimilarity between the origin station and any competing station while, for the MNL model, it would be an average dissimilarity. Therefore, the higher μ estimate for the cross-nested model versus the lower μ estimate for the MNL model appears to be intuitively correct, as this is what would be expected. It is interesting to note that $\theta_{ik} = 0$ in the cross-nested model provided the best fit, indicating that the allocation parameters α_{ik} are constant.

Table 2 MNL and Cross-nested origin choice forms: parameter estimates (10,324 observations)

	MNL origin choice		Cross-nested origin choice	
		t Value		**t Value**
K	18.14	93.45	17.31	89.74
μ	0.505	78.87	0.566	63.80
θ_{a*}	-0.00395	-31.01	-0.00362	-26.94
γ_{*j}	-1.82	-75.33	-1.83	-75.91
θ_{aj}	0.00315	29.29	0.00310	28.85
δ_i	0.45	*	0.50	*
θ_{ik}	------	------	0	*
ϕ	------	------	0.754	13.16
Adj. R^2	**0.6087**		**0.6108**	

t Value * indicates that a series of models have been tested with this parameter set to different values

In Table 3, the B_j values are broadly correlated with the populations around the destination stations although it can be seen that London, Edinburgh, York, Brighton and Norwich, all tourist destinations, attract more journeys than would be expected, were population to be the only factor. Another destination attracting more journeys than would be expected is Gatwick, which obviously attracts air travellers.

Table 3 MNL and Cross-nested origin choice forms: estimates of $\log A_L$, and of $\log B_j$ for a selection of stations (10,324 observations)

	population within 10km of station	MNL origin choice			Cross-nested origin choice		
		$\log A_L$	t Value	A_L	$\log A_L$	t Value	A_L
London	2,082,964	2.10	31.50	8.16	2.11	31.28	8.28
Station (j)		$\log B_j$	t Value	B_j	$\log B_j$	t Value	B_j
London	2,082,964	2.36	26.42	10.61	2.35	26.41	10.54
Manchester	969,343	0.00	*	1.00	0.00	*	1.00
Newcastle	619,302	-0.14	-2.09	0.87	-0.14	-2.06	0.87
Leeds	549,410	-0.17	-2.50	0.84	-0.17	-2.42	0.85
Edinburgh	406,781	0.10	1.42	1.10	0.10	1.42	1.10
Gatwick	166,229	-0.42	-6.24	0.65	-0.41	-6.10	0.66
Birmingham	1,123,502	-0.08	-1.27	0.92	-0.08	-1.22	0.92
York	157,688	-0.31	-4.51	0.73	-0.31	-4.52	0.73
Glasgow	792,335	-0.08	-1.17	0.92	-0.09	-1.20	0.92
Liverpool	855,898	-0.23	-3.40	0.79	-0.23	-3.35	0.80
Sheffield	559,787	-0.52	-7.18	0.59	-0.53	-7.32	0.59
Bristol	514,611	-0.30	-4.24	0.74	-0.30	-4.27	0.74
Nottingham	495,390	-0.48	-6.64	0.62	-0.48	-6.63	0.62
Leicester	441,754	-0.86	-11.15	0.42	-0.86	-11.25	0.42
Brighton	252,085	-0.59	-8.38	0.55	-0.56	-7.97	0.57
Reading	271,012	0.30	0.46	1.35	0.17	0.26	1.19
Norwich	219,941	-0.64	-7.77	0.53	-0.64	-7.80	0.53
Southampton	358,240	-1.24	-13.38	0.29	-0.64	-8.51	0.53

(Manchester is the redundant destination dummy variable)
t Value * indicates that value is constrained

8 CONCLUSIONS

This paper describes a new new cross-sectional direct demand aggregate model of rail passenger demand between UK stations, and is innovative in making extensive use of geographical information in the form of spatial distribution of populations and road network data.

The model can forecast demand for inter-urban rail journeys of over 40 km in Great Britain. Because it is based on observed demand levels and other readily available information, it is more straightforward to apply than previous techniques which could include, for instance, extensive survey work. Since it is based on observed demand, it also provides consistent results.

The models focus on travel choices made by residents located around the origin stations, while dummy variables are used to represent the destination station effects.

These residents, should they opt to travel, can choose between competing origin stations for a rail journey, or make their journey by road. Initially the origin station choice was modelled using a Multinomial logit (MNL) form but it was found that abstraction from competing stations took no account of their proximity to the origin station, and this was obviously a limitation.

Work on a model variant examined whether a station choice form with overlapping nests could be used. A particular cross-nested logit form recognised the proximity of a station to each of its competitor stations, and the model used this proximity to vary the proportion of journeys that are abstracted from those stations. The two theoretical models provide the foundations for the models that were subsequently estimated.

A number of potential weaknesses in the study can be identified which could be addressed in further work. For instance, the model developed here used 1991 census populations to explain demand in 1999/2000. The reason for this is that most of the work in preparing the data for the model was carried out prior to the availability of the 2001 census data. Models could be re-estimated again with population data that is more contemporary with the rail flow, revenue and service quality data, and it is likely that the fit would be improved. Also, demographic and socio-economic variables have been excluded. The authors suggest that the model might be refined by including variables such as those for car ownership or social class.

The model assumes that all access is by car, even though some passengers would arrive using other modes. This may be acceptable in most instances, given that road access times and distances could generally be effective proxies for all access modes, although the assumption should be recognised as a potential limitation. It may be possible to avoid this assumption by further work on modelling choice of access mode.

The origin population elasticities were introduced in order to overcome specification errors, discussed for instance by Gujarati (1995), in the theoretical functional form. These elasticities were less than one and appeared to provide a proxy for intervening opportunities around the origin station.

Competition with road journeys between the origin zone and the destination station has been modelled implicitly by including terms which recognise the utility of not travelling by rail. Further work might introduce the explicit effect of road competition into the model.

Finally, the model could also be modified in order to better understand traveller behaviour at destination stations, which were only estimated with dummy variables in this thesis. In the first instance, destination only models with origin station dummy variables could be modelled. The models could then be further extended to include population zones and competitor stations at both origin and destination stations.

ACKNOWLEDGEMENTS

The authors are grateful to Jonathan Pugh of ATOC for arranging the provision of CAPRI and MOIRA data. Thanks also go to Jonathan Tyler for providing station Ordnance Survey grid references, and to Peter Balmforth for software which calculates road distance and time matrices.

REFERENCES

ATOC (2002) *Passenger Demand Forecasting Handbook Version 4*, London

Gujarati, D. N. (1995) *Basic Econometrics, Third Edition* McGraw-Hill, Inc.

Hensher, D. A. and Button, K. J. (eds) (2000) *Handbook of Transport Modelling*, Pergamon

Holt, S. R. and White, P. R. (1981) *Modelling of Interurban Passenger Demand, Research Report No.5*, Polytechnic of Central London

Jones, I. S. and Nichols, A. J. (1983) *The Demand for Inter-City Rail Travel in the United Kingdom,* Journal of Transport Economics and Policy Vol XVII, No 2 pp.133-153

Jones, D. E. S. and White, P. R. (1994) *Modelling of cross-country rail services,* Journal of Transport Geography 2(2), pp.111-121

Koppelman, F. S. and Sethi, V. (2000) *Closed-form Discrete-choice models* Chapter 13 in Hensher and Button (2000)

Leventhal, B. J., Moy, C. and Griffin, J. (1993) *An Introductory Guide to the 1991 Census*, Henley-on-Thames, Oxfordshire: NTC Publications Ltd

Lythgoe, W. F. (2004) *Enhancing Cross-sectional Rail Passenger Demand Models*, PhD thesis, Institute for Transport Studies, University of Leeds

Lythgoe, W.F and Wardman, M. (2002) *Estimating Passenger Demand for Parkway Stations*, Paper Presented at European Transport Conference, PTRC, London

Lythgoe, W.F and Wardman, M. (2004) *Modelling Passenger Demand For Parkway Rail Stations*, Transportation 31, pp.125-151

Owen, A. D. and Phillips, G. D. A. (1987) *The Characteristics of Railway Passenger Demand,* Journal of Transport Economics and Policy Vol XXI, pp.231-253

SAS Institute Inc. (1999) *SAS System Version 8*. Cary, North Carolina, USA
Shilton, D. (1982) *Modelling the Demand for High Speed Train Services* Journal of the Operational Research Society Vol. 33, pp. 713-722

Tyler, J. and Hassard, R. (1973) *Gravity/ Elasticity Models for the Planning of the Inter Urban Rail Passenger Business* Proceedings of Seminar G, PTRC Summer Annual Meeting Warwick

Wardman, M. (1996) *Cross-sectional modelling of Inter Urban Rail Demand, Technical Note 403,* Leeds, Institute for Transport Studies, University of Leeds

Wardman, M. (2001) *Public Transport Values of Time. Working Paper 564*, Institute for Transport Studies, University of Leeds