Wider Factors affecting the long-term growth in Rail Travel

Technical Appendix

Ian Williams and Kaveh Jahanshahi

November 2018

The Independent Transport Commission
Registered Charity: 1080134
www.theitc.org.uk

Contents

List of Figures

List of Tables

1. Introduction

The aim of the modelling methodology that has been adopted is to maximise the quality of the evidence that can be derived from the NTS dataset in order to identify unambiguously the primary influences on rail passenger travel demand for each of four individual travel purposes.

An overview of the main steps in the development of the model has already been presented in Section 3.6 of the Report, while tabulations and interpretations of the estimated results for individual journey purposes are presented in its Chapters 7 to 9. In this Technical Appendix below we provide more detailed technical information on the individual modelling steps and then present indicators of the statistical significance of each of the estimated parameters.

The full modelling methodology that was applied to commuter rail travel is initially presented, together with its estimated results. The comprehensive modelling approach adopted for the commuter travel purpose needed to be substantially simplified when applied to the other travel purposes. These simplifications are briefly summarised and the statistical significance results are then presented in turn for each of these travel purposes: business; shopping and personal business; and social and holiday. Finally, future model development tasks are outlined that would further increase the precision of the estimated influences on rail travel demand for individual travel purposes.

2. The Modelling Methodology

The analysis of spatial and temporal trends in commuter rail travel within the main report has indicated:

a) that the rate of rail usage differs greatly between different types of individuals as well as between workers in different types of jobs;

b) that the rate of rail usage differs greatly depending on the particular type of area in which an individual is resident or is working;

c) that there are major differences between area types in their typical cross-section of types of resident individuals and of types of workplace jobs;

d) that the above relationships are not necessarily fixed immutably but may evolve somewhat over time.

The complexity of these interrelationships implies that simple statistical methods are inadequate for rail demand analysis, due to the internal correlations between the explanatory factors that are being analysed. A "self-selection effect" arises because those who locate in any particular area will often tend to have many socio-demographic characteristics in common with the other local residents there. Accordingly, it is difficult within conventional statistical analysis to distinguish the relative importance between:

- those influences that are related specifically to the built form of that locality (e.g. the range and frequency of rail services tends to be greatest in dense urban areas, whereas due to congestion and parking charges the car convenience there may be low);
- those influences that are related specifically to the types of residents or to the types of workplaces located within that area.

Moreover, as explained in section 3.3 of the Report, when analysing rail demand growth it is important to distinguish those influences that are due to **behavioural** changes, from those that are due simply to changes in the spatial or temporal **incidence** of the population or workforce. Accordingly, our analysis methods need to support the identification of such distinctions.

Advanced Structural Equation Modelling (SEM) in combination with Conditional Latent Categorical Analysis (LCA) and Zero Inflated Negative Binomial (ZINB) have been developed to disentangle interlocking systems of influences on behaviour and so to resolve the estimation complexities that such contexts generate. Accordingly, our statistical estimations of the influences on rail commuter demand are carried out in two stages:

- firstly, we use LCA to cluster individuals who adopt different geographical patterns for commuting by estimating simultaneously individuals' built form cluster membership and their socioeconomic and demographic profile;
- secondly, ZINB model is used to estimate the number of commuter rail trips made, while taking account of the structural zeros (i.e. to separate out the subset of observations originating from a subpopulation that can only have zero rail trips - the non-rail commuters). This second stage of estimation is performed conditional on the cluster membership that has been estimated in the first stage.

The first stage addresses the question of incidence by clustering together those with similar geographic characteristics. Socio-demographic characteristics affect the utility function (decision boundary) which is used to cluster the geography. The second stage then identifies the direct influences on rail commuting trip rate behaviour within each such homogeneous group.

Identifying built form latent clusters

To model potential non linearity in built form influences and to account for self-selection and spatial sorting effects; we use LCA for clustering commuters based on their homeplace and workplace characteristics. LCA is a machine learning technique to reduce the dimensionality of highly correlated variables into a tangible list of distinct clusters. Recent improvements in statistical methods have made it feasible to characterise heterogeneity among individuals or choices through *latent clusters*.

Using data on all persons in employment[1] from 2002 to 2012 in the NTS, we first run an LCA to identify a set of distinct clusters as market segments. LCA classifies a population into the set of distinct clusters that exhibit the largest difference across them and the most similarities within them.

The clustering indicators used are the built form characteristics at both homeplace and workplace, while controlling for demographic and socioeconomic characteristics. Based on the available NTS survey variables, the built form at the homeplace end is described based on the following set of characteristics [and their set of classes]:

- resident area type [9 classes: ranging from Central London through to those rural areas with perhaps towns of at most 25,000 population];

[1] The modelling of commuting and of business trip purposes is carried out using the NTS sub-sample of all persons in employment, whereas the modelling of other trip purposes uses the full NTS sample of adults.

- population density: in persons per hectare [5],
- frequency of bus service [5];
- walk time to bus stop: minutes [7];
- walk time to rail station: minutes [7].

At the workplace end, the NTS only reports information on the area type, without further information on built form or socio-economic characteristics. Accordingly, the six area types at the workplace end are assumed to be fixed as:

1) Central and Inner London;
2) Outer London;
3) Metropolitan Area;
4) Big Urban Area;
5) Medium Urban Area;
6) Small Urban and Rural Area.

The identification of the latent clusters is also conditional on the socioeconomic characteristics of the individuals. This cluster identification is **not** conditional on the actual modal travel behaviour of its inhabitants.

LCA has identified four distinct latent clusters for each of the six workplace area types in the NTS data, forming 24 commuter clusters in total. Table A-1 presents the probability of being a member of each commuting cluster when travelling from each homeplace area type[2]. For instance, the first block provides the cluster compositions for those who commute to the Central and Inner London workplace area type.

The clusters C1-1 and C1-2, which respectively form 8% and 47% of all commuters to Central and Inner London, have similar profiles. They: both largely comprise (i.e. 82% for C1-1 and 78% for C1-2) those residing in the same area type as this workplace (i.e. Central and Inner London); but also contain some of those (i.e. 15% for C1-1 and 22% for C1-2) who commute from Outer London. However, C1-1 contains some commuters from Central London (i.e. 28%) while the majority of C1-2 members (i.e. 77%) are commuters from Inner London.

Likewise, cluster C 1-3 (i.e. 27% of all commuters to Central and Inner London) is largely formed by commuters resident in Outer London, and cluster C 1-4 (i.e. 19% of all commuters to Central and Inner London) are those who mainly commute from Rural areas but it also includes commuters from Big, Medium and Small Urban area types.

---

[2] Similar statistics can be provided for other residential built form characteristics (see Section 3 below) which were the indicators of estimated latent clusters. Here we have provided those for residential area types so as to assist labelling the commuter clusters.

Table A-1 Proportion of people in each of the 24 clusters, based on LCA

	1-To Central and Inner London				2-To Outer London			
	C1-1	C1-2	C1-3	C1-4	C2-1	C2-2	C2-3	C2-4
1-Central London	28%	1%	1%	0%	0%	2%	0%	0%
2-Inner London	54%	77%	0%	2%	4%	70%	0%	0%
3-Outer London	15%	22%	81%	2%	94%	28%	4%	7%
4-Metropolitan	0%	0%	0%	0%	0%	0%	0%	0%
5-Large Urban	0%	0%	0%	1%	0%	0%	1%	0%
6-Big Urban	2%	1%	8%	25%	1%	0%	56%	11%
7-Med Urban	0%	0%	6%	9%	1%	0%	16%	2%
8-Small Urban	1%	0%	4%	21%	0%	0%	19%	9%
9-Rural	0%	0%	0%	41%	0%	0%	4%	71%
Grand Total	**8%**	**47%**	**27%**	**19%**	**67%**	**8%**	**17%**	**7%**

	3-To Metropolitan				4-To Big Urban			
	C3-1	C3-2	C3-3	C3-4	C4-1	C4-2	C4-3	C4-4
1-Central London	0%	0%	0%	0%	0%	0%	0%	0%
2-Inner London	0%	0%	0%	0%	0%	2%	0%	0%
3-Outer London	1%	0%	0%	0%	1%	14%	0%	0%
4-Metropolitan	93%	35%	12%	1%	9%	1%	2%	0%
5-Large Urban	0%	62%	61%	4%	2%	1%	4%	0%
6-Big Urban	4%	0%	2%	0%	87%	80%	40%	9%
7-Med Urban	0%	3%	3%	1%	2%	0%	8%	0%
8-Small Urban	1%	0%	16%	12%	0%	1%	26%	0%
9-Rural	0%	0%	7%	82%	0%	0%	19%	90%
Grand Total	**12%**	**43%**	**36%**	**10%**	**50%**	**12%**	**25%**	**14%**

	5-To Medium Urban				6-To Small Urban and Rural			
	C5-1	C5-2	C5-3	C5-4	C6- 1	C6- 2	C6- 3	C6- 4
1-Central London	0%	0%	0%	0%	0%	0%	0%	3%
2-Inner London	0%	2%	0%	0%	0%	0%	0%	22%
3-Outer London	0%	11%	0%	0%	0%	0%	3%	27%
4-Metropolitan	0%	9%	0%	0%	0%	1%	6%	12%
5-Large Urban	2%	10%	3%	0%	0%	2%	15%	3%
6-Big Urban	2%	26%	4%	0%	0%	5%	22%	28%
7-Med Urban	50%	39%	14%	3%	0%	6%	25%	5%
8-Small Urban	47%	0%	46%	2%	0%	15%	28%	2%
9-Rural	0%	3%	33%	95%	100%	71%	0%	0%
Grand Total	**38%**	**10%**	**31%**	**21%**	**38%**	**43%**	**16%**	**2%**

Studying all six blocks, each containing four latent clusters, prompts us to adopt a common overall numbering convention. For each block, commuters are classified by clusters that range from: the most internal commuting pattern (i.e. having the same common area type for residence and workplace) on the left; to the most external ones on the right of the block. For instance, for the commuters to Central and Inner London, cluster C 1-1 and to a lesser extent C1-2 are formed from those who reside in the same area type as this workplace, cluster C1-3 are those living in adjacent Outer London, and cluster C1-4 are mainly commuters resident in the furthest away, less populated areas. A similar presentation sequence is adopted in Table A-1 for the commuters to each of the other five workplace area type blocks.

Having identified 24 relatively homogenous latent clusters, the next section examines how each cluster differs from other clusters with respect to some of the socio-economic characteristics of its members.

3. Socio-demographic Profile of Latent Cluster Members

There are striking differences in socio-economic profiles across the latent clusters as illustrated for the examples: of two of the employment sectors in

Figure A-1; for three household income classes in Figure A-2; and for four occupation groups in Figure A-3. Based on 2002 to 2012 NTS data[3], each figure illustrates within each of the 24 individual latent clusters, its proportional split across the set of classes for the specified socio-economic variable. For example, the leftmost column C1-1 in Figure A-2 denotes that among the set of commuters to Central and Inner London within this cluster C1-1, 10% have household incomes below £25k; 25% are between £25k and £50k; whereas 65% are over £50k / annum.

The main findings from examining the set of socio-demographic characteristics of the commuters within each of the clusters, through combining the proportions from Table A-1 with those in

Figure A-1 to Figure A-3 and with charts for other socio-demographic characteristics, are as follows.

a) Commuters working in the financial sector comprise between 10% (C1-2) and 23% (C1-4) across the total commuters to Central and Inner London (C1-1 to C1-4). This proportion working in finance is significantly smaller (less than 6%) for those commuters to all other workplace clusters.

b) The majority of manufacturing jobs are in the Medium Urban and Small Urban and Rural area types, followed by Metropolitan and Big Urban areas but only with very few in London. Commuters to manufacturing in Metropolitan areas are more likely to be from further away large urban areas, while a large proportion of those who commute to Big, Medium and Small Urban and Rural areas are internal to these area types.

c) As expected, a large proportion of those who commute to Central/ Inner London (between 40% and 65%), to Outer London (between 40% and 57%) and to a lesser extent those to Metropolitan and Large Urban areas (between 20% and 42%) are members of high income households. Also, with the exception of commuters from Central/Inner London to Central/Inner London (C1-1), the general patterns show that those who commute from less populated Small Urban and Rural areas to

[3] The latent cluster analysis was based only on 2002 to 2012 NTS data. We were not able to include 2013 to 2015 data because the accessibility information is unavailable after 2012 within it.

London, Metropolitan and Big Urban areas are more likely to be in high income households while those in the low income band tend to shorter distance commuting, making trips internal to their area type (particularly when they reside in less populated areas).

d) A large proportion of commuters (around 40%) in commuting clusters from less populated areas to more populated ones (e.g. C1-4, C2-4, C3-4, C4-4) are in professional/managerial households, this trend is the reverse of that for manual workers who tend to reside closer to their workplace.

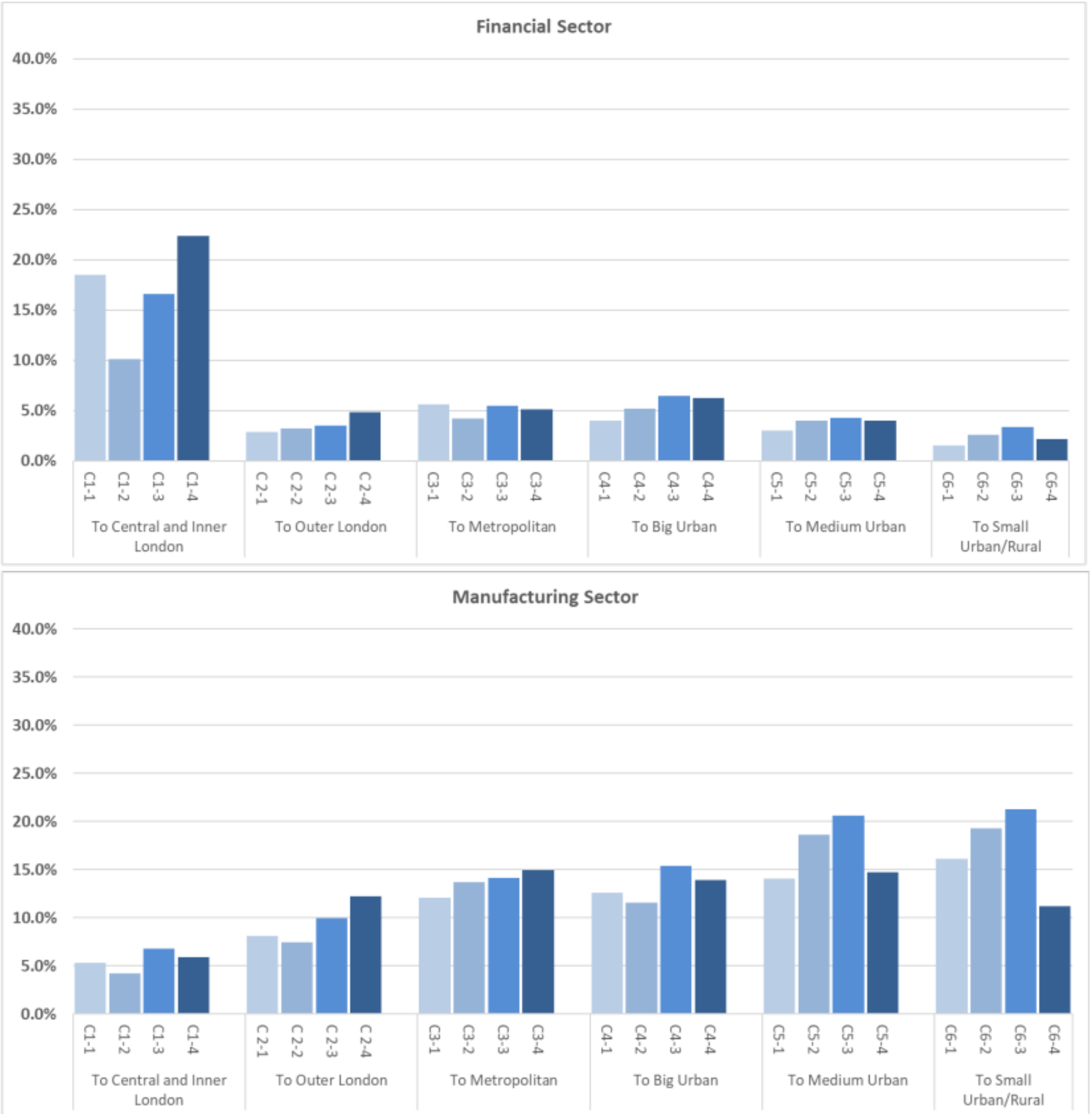

Figure A-1 Socio-demographic profile of latent clusters - selected employment

sectors

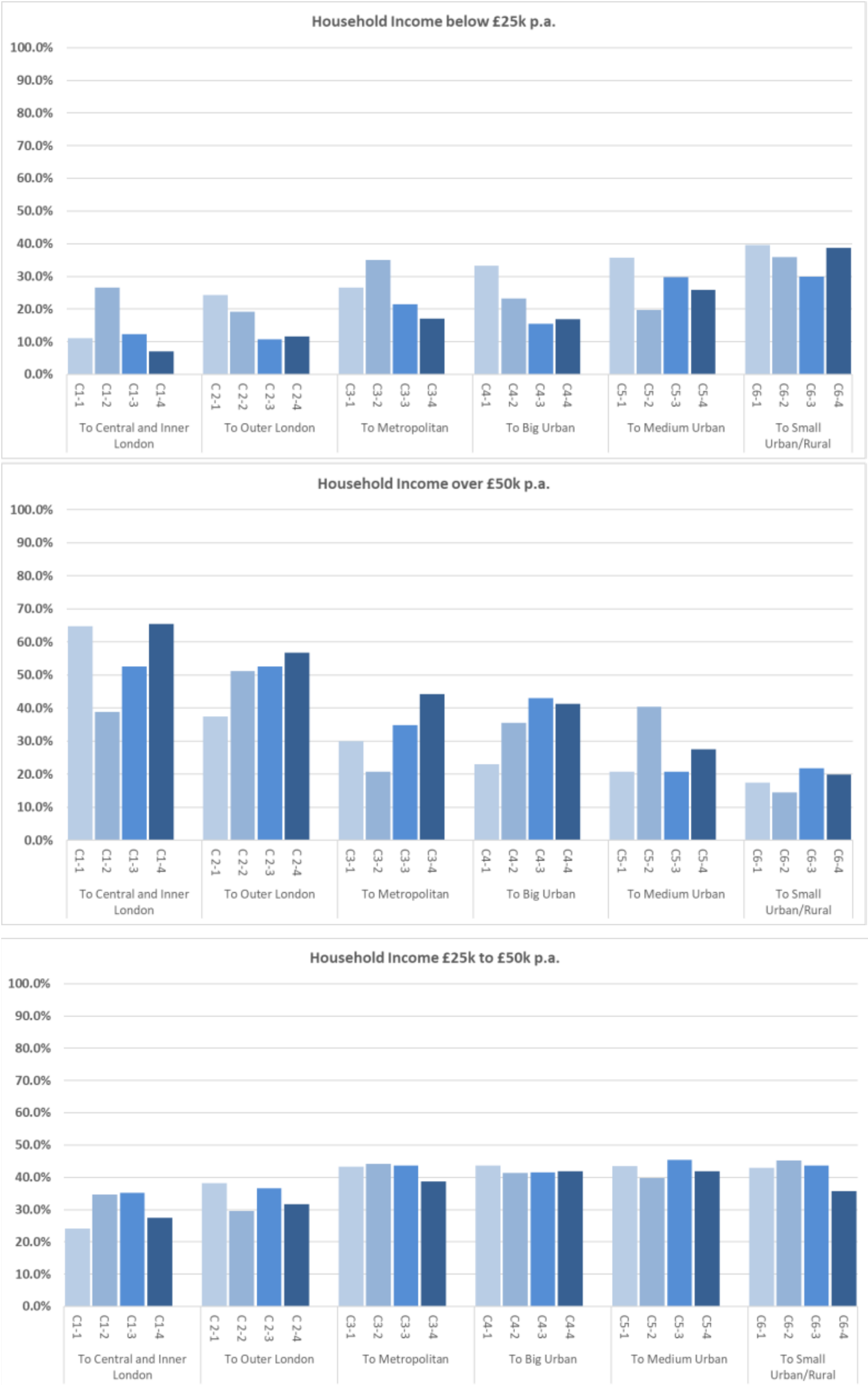

Figure A-2 Socio-demographic profile of latent clusters - household income bands

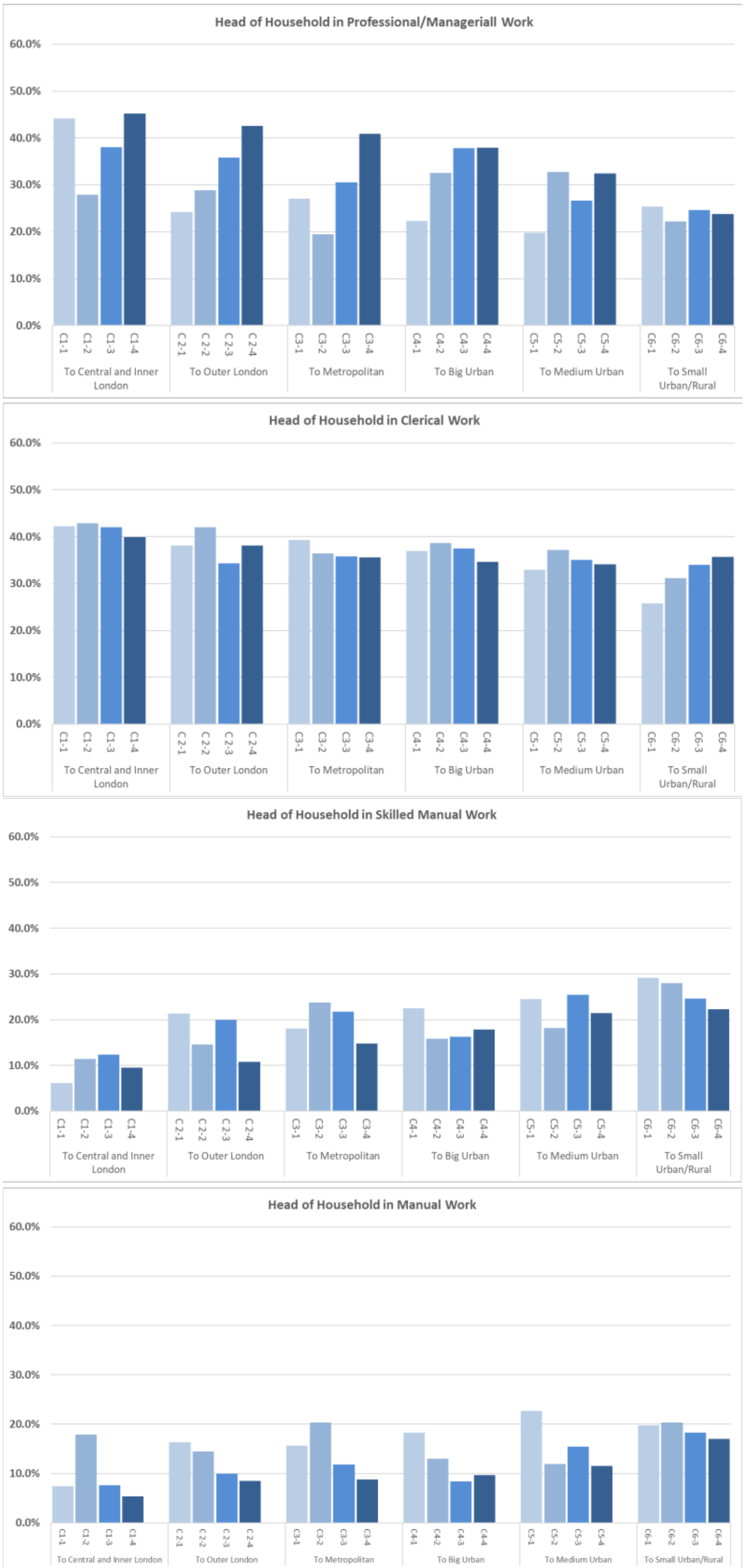

Figure A-3 Socio-demographic profile of latent clusters - Head of household occupation

Analysis of rail trips across latent clusters

Having identified the clusters without any reference to rail trip making, we then examined the commuting rail trips within each of these clusters. This helps in identifying the major market segments (clusters) for commuting. Figure A-4 compares weekly commuting trip rates across all modes versus rail trip rates, while Figure A-5 shows the rail share (the bar chart with the left axis) and the absolute number (the dotted line with the right axis) of commuting rail trips (from the NTS data from 2002 to 2012).

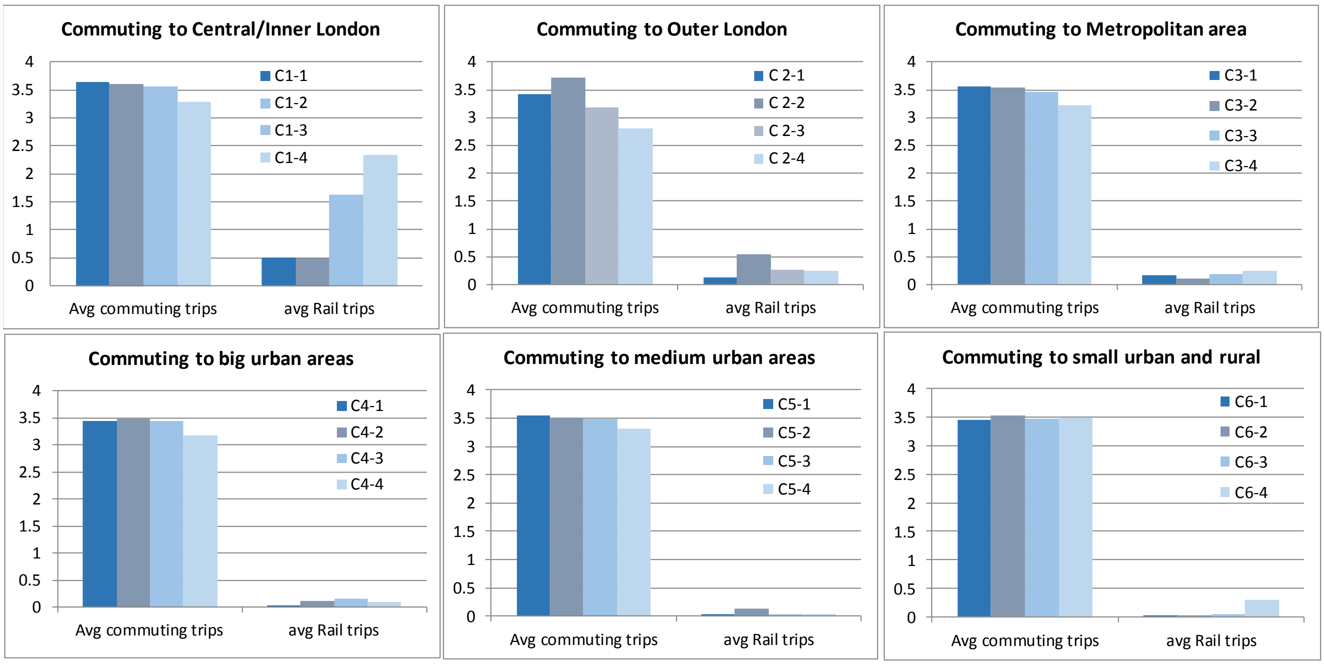

Figure A-4 Average commuting trip rates (per person per week) across clusters

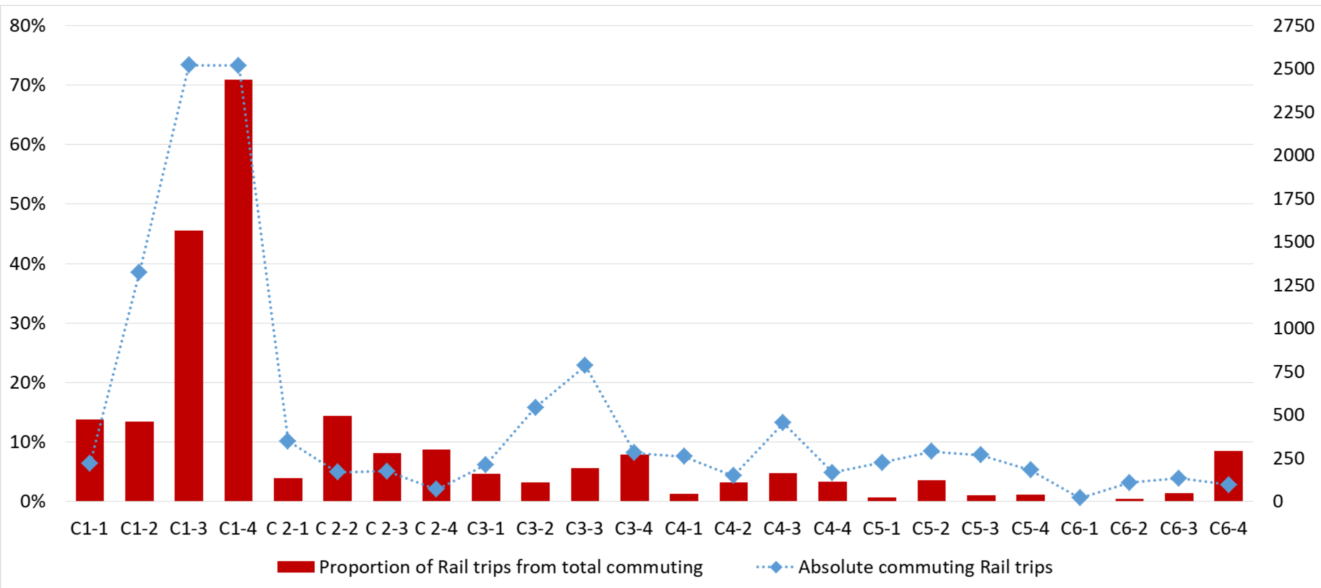

Figure A-5 Proportion of rail trips within total commuting (Red columns), Absolute number of rail commuting trips in NTS sample (blue dotted line), by cluster

It can be seen that there is a striking difference in rail commuting trip totals across the clusters. Clusters C1-4 and C1-3 (i.e. those who commute to Central/Inner London from Rural and surrounding Big to Small Urban areas, and those who mainly commute from Outer London) have the largest absolute number of rail trips as well as the largest proportion of rail use within their total commuting (respectively 70% and 45%). The next largest rail trip clusters are: C1-2: commuting within Inner/Central London; then C3-3 and C3-2 from Large Urban areas to Metropolitan areas.

These results highlight that for a large part of the total commuter market, the rail mode has a minimal share. Rail's main usage for commuting is instead highly concentrated into a few specific workplace markets, primarily longer distance movements to London and to a lesser extent to the centre of metropolitan areas. Accordingly, the statistical modelling of rail commuting has focused on these rail intensive clusters.

4. Estimating Influences on Rail Commuting Using ZINB

The analysis of rail trip rates is implemented using a ZINB model that is applied to individual *fixed clusters*, rather than to the probabilistic latent clusters that had been derived from LCA. This is chosen as the second best approach because we are not able to use the probabilistic modelling based on the likelihood of clusters' membership[4]. This occurs because:

- Consistent data for all built form characteristics, in particular data for accessibility measures, were not available from the NTS for the years 2013 to 2015. These years are particularly important to us, specifically because of the declines observed in aggregate rail trip numbers in the most recent years.
- Fixing the clusters to be based just on Area Type pairs provides a better understanding of the variations in commuting trips across the main rail market segments. This form of segmentation is supported by our analysis which shows a strong link between: area type pairs; other built form characteristics; and socio-economic variables.

The relative importance of area types in identifying latent clusters (as indicated in Table A-1) and the striking differences in rail trip numbers across latent clusters (as indicated in Figure A-5) suggest that this use of fixed clusters will provide a good approximation to the probabilistic ones, when analysing changes in rail trips over time.

The following summarises the approach used to estimate the influences on commuting rail trips,

a) Firstly, we defined the appropriate fixed clusters that are based on area type pairs at homeplace and workplace. These fixed clusters approximate closely to the latent clusters presented above in Table A-1.

[4] The probabilistic modelling uses the probabilities of being a member of each of the identified clusters. The advantage of probabilistic modelling is that we benefit from using the clusters identified by the model which are based on a range of built form characteristics and socioeconomic variables. However, Area Types are shown to be the most important identifier of these clusters; hence as the second best approach these area types can be used for clustering when analysing trends over time.

b) Secondly, we analyse the average rail usage within the selected clusters in terms of both: the weekly rail trip rates per rail user; and the propensity for an individual to make at least one rail trip within the survey week – termed “rail propensity”.

c) Finally, we examine how the incidence of rail trip making has changed over time, within a cluster.

This estimation is carried out only for those area type pairs that have sizeable numbers of rail commuting trips so as to ensure that there is an adequate sample size for examining the trends over 14 years. The selection is made based on Figure A-4 and Figure A-5, which indicate that the latent clusters C1-3 and C1-4 have much larger trip numbers than any of the other clusters. Accordingly, the two major rail fixed clusters: C1-3*[5]: commuters from Outer London, and C1-4* those from Big/Medium/Small Urban and Rural areas, both to Central/Inner London are the clusters that are used for the detailed statistical analysis of the influences on rail commuting, as reported below in Table A-2.

It is encouraging to note from the estimates within Table A-2 that:

- the same subset of variables (with only one exception) is significant for both clusters;
- the direction of their effects are always common across clusters (e.g. both odds-ratio values are either greater than 1 or else both are less than 1);
- in many cases the magnitudes of both odds-ratio values are reasonably similar.

The exception is that the manual head of household variable is not significant for the longer distance travellers who commute from outside London to Central/Inner London (cluster C1-4*). This lack of significance does not automatically indicate a deviation in behaviour from that in cluster C1-3*. Instead it may relate more to the high generalised costs of commuting over the long distances to Central/Inner London, which dissuades manual workers from such long distance commuting by any mode. Figure A-3 indicates that only 4% of commuters in cluster C1-4 are manual, a lower proportion than in any of the other clusters. This implies there will only be a minimal sample of such travellers within the NTS dataset which then militates against extracting any statistically significant influences for manual workers within this cluster C1-4*.

The results in Table A-3 indicate for the cluster C1-4* that there is a systematic reduction over time in the rail trip rate for those full-time workers travelling in by rail to Central/Inner London from outside London. The high significance level (p-value =.0.000 in all cases) indicates that this reduction in effect over time has been well estimated within this NTS cluster sample.

[5] The notation C1-3* is used to distinguish this fixed cluster that is based on all those commuting between the area type pair, from C1-3 the latent cluster as defined above in Table A-1.

Table A-2 Significant influences on: generating at least one rail commuting trip; and on number of rail trips, conditional on generating at least one rail commuting trip

a) **Influences on making at least one rail commuting trip in the week**			
Fixed Cluster		**c1-3***	**c1-4***
Influence	**Reference group**	**Odds ratio (p-value)**	**Odds ratio (p-value)**
Head of household: manual	… in clerical work	0.59 (.005)	not significant
Head of household: skilled manual	… in clerical work	0.67 (.020)	0.40 (.006)
High income households: > £50k	Medium income: £25-50k	1.52 (.000)	1.64 (.038)
Self employed	Employed	0.53 (.001)	0.48 (.022)
Work in public admin, etc. – gSIC4	gSIC Ref.	1.37 (.049)	2.04 (.039)
Work in financial sector – gSIC6	gSIC Ref.	4.35 (.000)	3.45 (.005)
Work in business services – gSIC7	gSIC Ref.	2.08 (.000)	2.04 (.027)

b) **Influences on number of commuting rail trips, conditional on making at least one such trip**			
Influence	**Reference group**	**Odds ratio (p-value)**	**Odds ratio (p-value)**
Full time work	Part time work	1.5 (.000)	2.0 (.000)
Head of household: skilled manual	… in clerical work	1.10 (.084)	1.17 (.030)
High income households: > £50k	Medium income: £25-50k	not significant	0.91 (.027)

Table A-3 Variation over time in the odds-ratio (and p-value) for work status in influencing the weekly rail commuting trip rate of rail commuters from outside to Central/Inner London (cluster C1-4*)

Influence	**2002 to 2004**	**2005 to 2008**	**2009 to 2012**	**2013 to 2015**
Full time work (ref group: Part Time)	2.52 (.000)	2.51 (.000)	2.04 (.000)	1.72 (.000)

5. Statistical Results for Non-commuting Purposes

In an ideal world, an SEM type of approach, analogous to that used above for commuting, would also have been implemented within this study for each of the other trip purposes. However, because the main focus of this study was on the general analysis of rail trends, rather than on the development of modelling methodologies, sufficient resources were not available to carry out the exploratory research that would have been required to complete this task satisfactorily, including the challenges of the relatively small sample sizes for rail travel within some of these purposes.

Accordingly, the rest of the travel purposes listed below were modelled using a ZINB model for rail trip making and a multivariate regression model for trip length. This is similar to the stage two task for commuting but without first carrying out its LCA identification of homogeneous clusters. A brief overview is presented below in Section 6, of how SEM methods could potentially be introduced in the future to refine the estimates within the various models for the non-commuting travel purposes.

The tables below correspond to the set of tables 13 to 18 of the Report. They cover in turn the: business; shopping and personal business; and social and holiday trip purposes. They present the significance level (p-value) for each of the influencing variables that was estimated to be statistically significant for the specified model. Overall they indicate that most of these listed influences are significant at the 1% level or better.

Table A-4 Significant influences on: generating at least one business rail trip; and on the number of rail trips, conditional on generating at least one business rail trip

a) Influences on making at least one business rail trip in the week		
Influence	**Reference group**	**Odds ratio (p-value)**
Head of household: manual	… in clerical work	0.16 (.000)
Head of household: skilled manual	… in clerical work	0.22 (.000)
High income households: > £50k	Medium income: £25-50k	2.38 (.005)
Self employed	Employed	0.48 (.043)
Work in financial sector – gSIC6	gSIC Ref.	20.0 (.006)
Work in business services – gSIC7	gSIC Ref.	2.70 (.003)
1+ car in household	No car in household	0.33 (.003)
b) Influences on number of business rail trips, conditional on making at least one such trip		
Influence	**Reference group**	**Odds ratio (p-value)**
Head of household: manual	… in clerical work	2.0 (.017)
Head of household: skilled manual	… in clerical work	2.2 (.006)
1+ car in household	No car in household	0.48 (.000)

Table A-5 Significant influences on average rail trip length for those making business rail trips

Influence	Reference group	Estimate (p-value)
Head of household: managerial/profess.	… in clerical work	8.3 (.036)
1+ car in household	No car in household	22.7 (.000)

Table A-6 Significant influences on: generating at least one rail trip; and on the number of rail trips, conditional on generating at least one rail trip for shopping and personal business

a) Influences on making at least one S&PB rail trip in the week		
Influence	**Reference group**	**Odds ratio (p-value)**
Full Time workers	Economically Inactive	over 100 (.000)
Male	Female	2.08 (.000)
1+ car in household	No car in household	1.96 (.018)
b) Influences on number of S&PB rail trips, conditional on making at least one such trip		
Influence	**Reference group**	**Odds ratio (p-value)**
Full Time workers	Economically Inactive	0.59 (.000)
Male	Female	0.79 (.007)
High income households	Medium income: £25-50k	1.46 (.000)
Age below 24 years	Age 35 to 49	1.8 (.000)
Age 25 to 34 years	Age 35 to 49	1.28 (.001)
1+ car in household	No car in household	0.32 (.000)

Table A-7 Significant influences on average rail trip length for those making rail trips for shopping and personal business

Influence	Reference group	Estimate (p-value)
Single Adult households	2+ adult households	3.7 (.024)
1+ car in household	No car in household	8.6 (.000)

Table A-8 Significant influences on: generating at least one rail trip; and on the number of rail trips, conditional on generating at least one rail trip for social and holiday purpose

a) Influences on making at least one S&H rail trip in the week		
Influence	**Reference group**	**Odds ratio (p-value)**
Full Time workers	Economically inactive	0.67 (.002)
Male	Female	0.74 (.000)
Single Adult households	2+ adult households	1.41 (.004)
Head of household: skilled manual	Economically inactive HoH	0.48 (.000)
Head of household: manual	Economically inactive HoH	0.66 (.015)
High income households	Medium income: £25-50k	2.27 (.000)
Age below 24 years	Age 35 to 49	3.87 (.000)
Age 25 to 34 years	Age 35 to 49	1.48 (.0001
1+ car in household	No car in household	below 0.01 (.000)
Year	Continuous variable	1.04 (.000)
b) Influences on number of S&H rail trips, conditional on making at least one such trip		
Influence	**Reference group**	**Odds ratio (p-value)**
Male	Female	1.13 (.001)
Head of household: managerial/ professional	Economically inactive HoH	1.41 (.000)
Head of household: clerical	Economically inactive HoH	1.43 (.000)
Low income households	Medium income: £25-50k	0.67 (.000)
Age below 24 years	Age 35 to 49	1.65 (.000)
Age 25 to 34 years	Age 35 to 49	1.40 (.000)
Age over 65 years	Age 35 to 49	0.82 (.018)
1+ car in household	No car in household	0.86 (.027)
Year	Continuous variable	1.019 (.000)

Table A-9 Significant influences on average rail trip length for those making rail trips for social and holiday purpose

Influence	Reference group	Estimate (p-value)
Full Time workers	Economically inactive	4.2 (.044)
Part Time workers	Economically inactive	5.0 (.047)
Male	Female	-4.9 (.001)
Age over 65 years	Age 35 to 49	10.8 (.001)
1+ car in household	No car in household	4.8 (.020)
Year	Continuous variable	-0.63 (.004)

6. Further Potential Development of these Rail Travel Models

The small scale together with the particular focus of this study has meant that the statistical model development tasks included within it needed to be quite limited in their resource requirements. The research has however, identified a variety of future activities that would provide greater precision in the estimates of the relative strengths of many of the influences on rail passenger growth that this study has identified.

For commuting, we used a form of SEM through the combination of LCA with ZINB. This could be improved by further using SEM to represent car ownership indirectly as an intervening variable, rather than as an explicit independent variable within the current model. We have found from other research (Jahanshahi et al., 2015) that this approach provides a better behavioural understanding of how land use patterns, socio-demographic characteristics and car ownership all interrelate to influence travel demand.

It would also be instructive to experiment to develop the LCA approach for use for each of the other travel purposes so as to generate homogenous clusters to underpin their modelling analysis and estimation. The area type of the address of the workplace of an individual is known within the NTS, which is how the commuter destination could be used as a variable within the commuting analysis. However, the NTS does not generally provide information on trip destination locations so that the geographic component of the LCA for any of the other travel purposes could only use the residential area type of an individual. As discussed above already, there would also be advantages to changing the representation of car ownership to act instead as an intervening variable. Finally, care would be needed within the analysis to balance the sample size of rail trips that is available within the NTS for the travel purpose being analysed, against the inherent data requirements for the estimation procedure that is being developed.

A further worthwhile development would be to integrate these various models of rail demand by trip purpose, together with population enumeration based projections of socio-demographic change, segmented by residential density band. The resulting comprehensive model could then be used to provide improved projections of future rail demand under a variety of scenarios.

An improved ideal would then be to further integrate this approach, together with the strengths of the existing PDFH system in representing the influence of actions within the rail sector. The resulting combined model should provide a more reliable, evidence-based approach to rail demand forecasting. Moreover, it would be able to assess the potential impacts on rail of a wide range of policy measures both those within and those outside the rail industry itself.

7. References

Jahanshahi, K., Y. Jin, and I. Williams, *Direct and indirect influences on employed adults' travel in the UK: New insights from the National Travel Survey data 2002–2010.* Transportation Research Part A: Policy and Practice, 2015. **80**: p. 288 - 306.