CALIBRATING A HOUSEHOLD RELOCATION MODEL FOR LEICESTERSHIRE

Emma Revill and David Simmonds
David Simmonds Consultancy,
Suite 14, Millers Yard, Mill Lane
Cambridge CB2 1RQ England

Abstract

This paper reports on the calibration of the household relocation model within the land-use component of the Leicester and Leicestershire Integrated Transport Model (LLITM).

The land-use component of LLITM is a new implementation of the DELTA land-use modelling package. The treatment of residential location is very similar to that in previous DELTA applications, in that

- households are classified by composition/age/economic status (10 types) and by socio-economic level (4 levels based on occupation)
- households change over time (eg from young singles to young couples, or older couple to retired couple)
- only a proportion of households consider moving in each year – those that have changed type, and a proportion of others – giving the effect (consistent with observation) that younger households are generally more mobile
- households moves are a mixture of local moves, in which distance has a strong deterrent effect, dominated by housing issues, and longer-distance moves more related to employment issues.

The work here related to the choice of coefficients for the local moves model. Previous applications of this DELTA component have relied on limited, indirect calibration, drawing on previous research by others (including hedonic price analyses), matching the broad patterns of moves (eg average distance) reported in other surveys, and an important element of professional judgement informed by subjecting the resulting model to sensitivity testing. For LLITM we set out to supplement this with some direct calibration of a model of household location choice for moving households. A separate survey for this purpose was not affordable, so we attempted to obtain suitable data from additional questions in the Leicester and Leicestershire Travel Census, the major household survey conducted as part of the model development project.

The observed data provided a sample of moves by “from” and “to” zones. The main focus of the calibration was to explain these moves in a multinomial

logit model with the independent variables being our modelled values for accessibility, cost of location, floorspace per household, and distance from old to new location. Since only the chosen "new" location was reported, and since it would have been highly impractical to identify all of the other (>900) zones of the model as "rejected" locations, we resorted to the standard technique of specifying a random (or semi-random) sample of other zones as the rejected alternatives for each observation. Despite the small number of observations we succeeded in obtaining significant coefficients on each of the independent variables. These informed the final choice of coefficients for the working LLITM model.

Apart from its contribution to the development of LLITM, the exercise demonstrated the feasibility of calibrating a model of household relocation on data about such relocations, avoiding the limitations and pitfalls of building static models on cross-sectional data. Some lessons for future exercises of this type will be identified.

Introduction

The work described here was one small part of a project to implement the Leicester and Leicestershire Integrated Transport Model (LLITM) – which is a land-use/transport interaction model, not just a transport model. LLITM was commissioned by Leicestershire County Council and Leicester City Council from AECOM and David Simmonds Consultancy (DSC), working in collaboration with Scott Wilson and a number of specialist advisors. The land-use modelling work, carried out by DSC, involved setting up and testing a new application of the DELTA package (see previous ETC papers eg Dobson et al, 2007, 2009). The zone system for the model had to be decided in advance of the calibration work and is shown in Figure 1 below. It consists of a fully modelled area which extends some way beyond the Leicestershire boundary and which is in turn surrounded by a further buffer area.

Our paper provides a very brief outline of the DELTA package and then concentrates on the residential location component within DELTA, the issues of calibrating it, the data used in the present case and the results obtained.

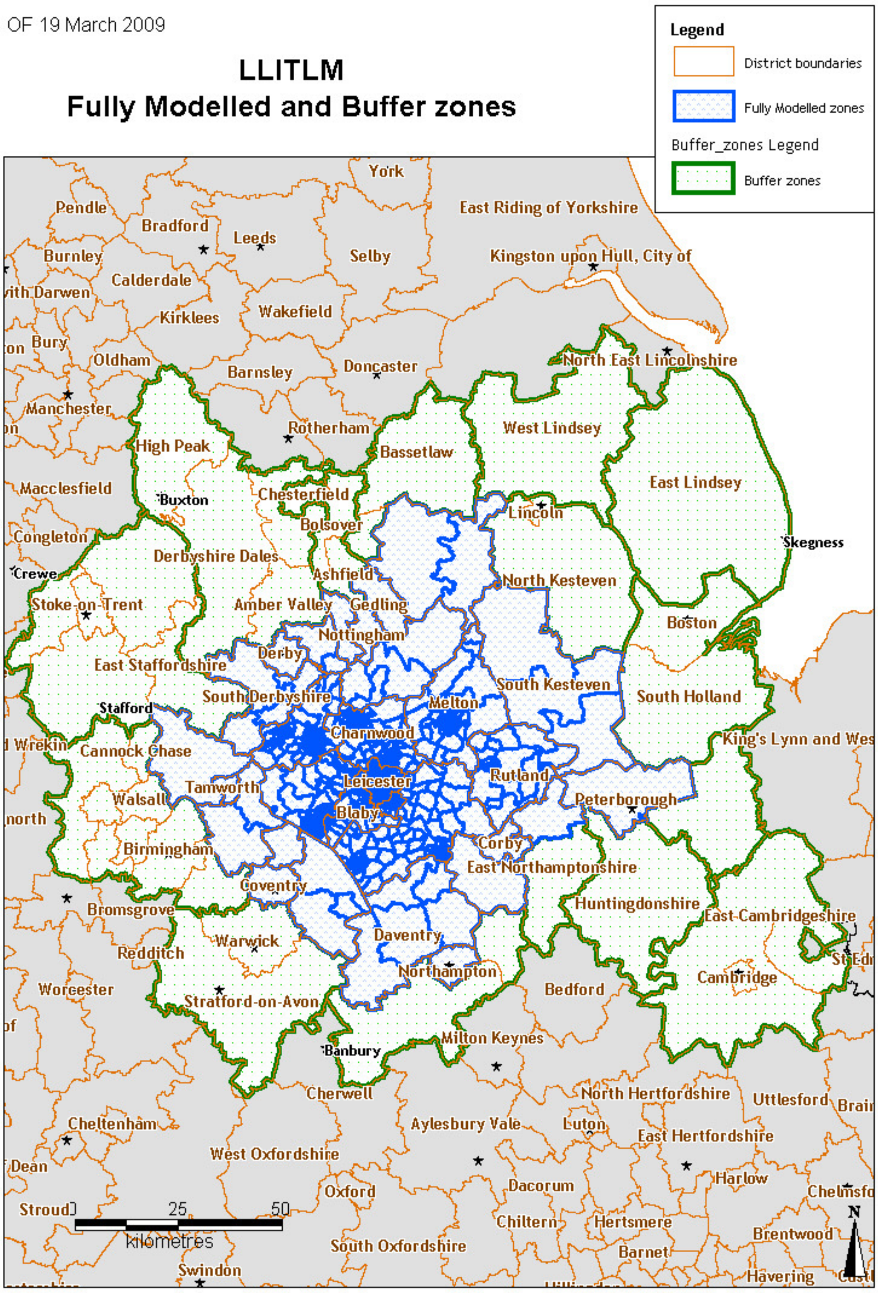


Figure 1 LLITM Modelled Area

The structure of the DELTA model

DELTA is a dynamic model of land-use change: it is run in one year steps, each step forecasting one year's worth of changes. The thinking behind the approach was originally published in Simmonds (1999). It works in terms of different processes of change; there are some links between these within each one-year period (eg so that people who leave the modelled region are removed from housing demand and labour supply) and many more links that work gradually, over time. The operation of the model is shown in Figure 2 and the main relationships in Figure 3.

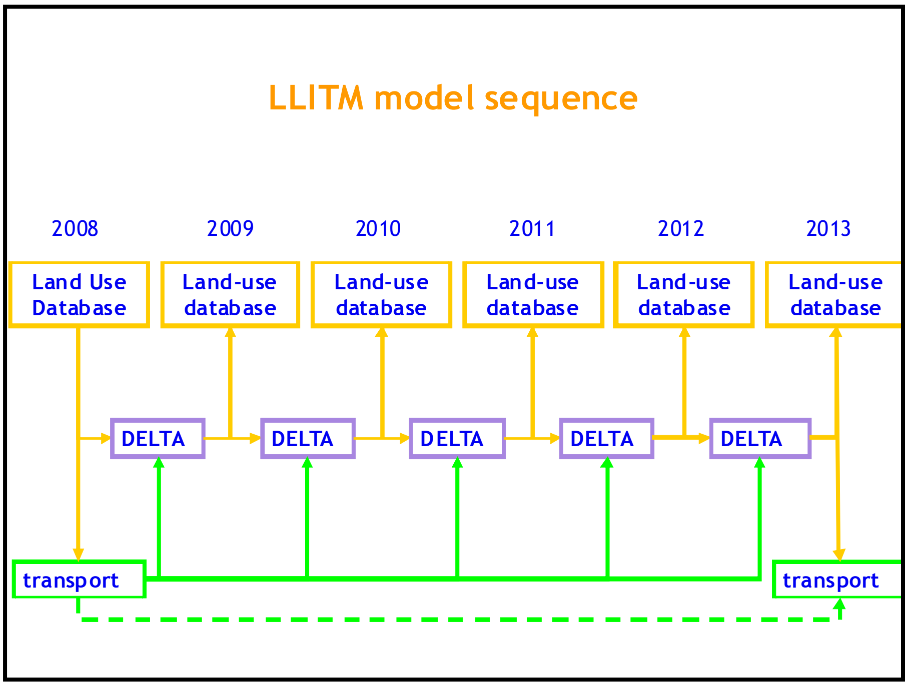
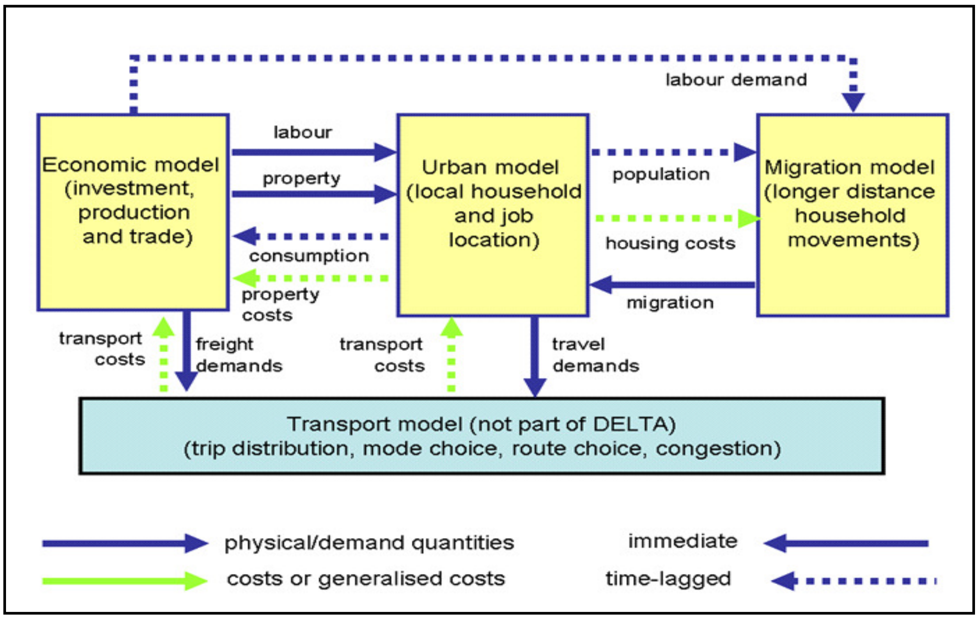


Figure2 Land use and transport model sequence (five-year example)

Figure 3 Overview of the Land Use model



The DELTA residential location model

Scope of the residential location sub-model

The residential location sub-model is both the "location and relocation choice sub-model", and the "residential property market sub-model". The model is solved iteratively: first households are located given an initial set of rent values. The amount of housing occupied by these households is then compared with the stock of housing available, and rents are adjusted to reconcile demand with supply. Changes in rents affect how many households will locate in each zone, how much space they will each occupy, and (in some circumstances) how much space is left vacant. The model has converged when a consistent solution has been found for all of these variables; this will ensure that all households are located, but not necessarily that all housing is occupied.

This paper is solely concerned with the choices made by households given a particular set of rents, and from here on deals only with that aspect of the model.

The households to be located

DELTA is intended to be applied with a detailed classification of households reflecting household size and composition. Like other recent applications in England, LLITM operates with 10 categories reflect the composition and age of household members cross-classified with 4 socio-economic levels based on the occupation of the household reference person.

An important characteristic of the model is that only a proportion of households make residential choices in any one period. It is assumed that the main reasons for making a new residential choice are linked to change in one of the household classification variables, eg a change in the household's composition or in its work status. All households which are forecast to move from one category to another – plus a proportion of unchanged households - are assumed to make such a choice, and hence enter into the location process.

The households in the location model fall into two groups: "pool" households, which have no previous location within the area, and "mobile" households, which do have a previous location within the area modelled. Newly formed households and households resulting from existing households merging (eg singles forming couples) are assumed to make new location decisions and are counted as "pool" households. "Mobile" households are those which are undergoing other changes (mainly from couple with children onwards). In addition, a proportion of non-changing households is assumed to be "mobile" in each period. The numbers of "mobile" and "pool" households are initially calculated in the household transition model (which also finds and subtracts the numbers of households which have dissolved or migrated out of the modelled area altogether). The inter-area migration model is then applied, before the location model. The migration model predicts moves of households between areas within the modelled system: these households are subtracted from the "mobile" and "pool" numbers for the areas they leave, and added to the "pool" numbers for the areas into which they migrate. Households migrating from the rest of the world are also added to the "pool" numbers.

The task for the residential location model in each area is therefore to locate two sets of households, each of which contains households in each of the 40 types:

- "pool" or "must locate" households which belong to an area

- “mobile” or “may relocate” households which are already located in a specific zone.

The focus of the calibration is on the latter; the resulting coefficients are assumed to apply to the former as well.

Influences on residential relocation

The DELTA residential relocation model is programmed to use various combinations of the following variables as influences on residential choice:

- availability of housing
- accessibility
- cost of location
- floorspace per household
- housing quality
- environmental quality
- distance moved.

The *amount of housing available* in one year is the amount that was left vacant in the previous year (if any), plus newly-completed development (if any), plus the amount which may be vacated by households which may choose to move out. An important characteristics of the model is that in any one year a large majority of households are “immobile” and their housing is not in any way in the housing market and hence is not available to locating households. The proportion of households that remain immobile varies by type – quite low for young singles and very high for wholly-retired households.

The *accessibility* variable is essential if the model is to work as a full land-use/transport interaction model – this is where the direct effects of transport on residential location enter the land-use model. Cost of location is also essential because this is the main variable through which rent influences location, ie this is the key variable in representing the competition between households for each location. Floorspace per household is also affected by rent (note that in earlier applications of DELTA, cost and floorspace per household were combined into a utility measure). Housing and environmental quality are not essential to the workings of the model, but both are included in the operational version of LLITM. Whilst this list of variables is relatively rigid, there is for most of them a lot of flexibility in how they are defined.

The *accessibility variables* are calculated from previous land-uses and from outputs of the transport model. Each component of household accessibility is a logsum measure of the expected generalised cost of accessing a particular kind of opportunities, eg jobs of a particular socio-economic level, or shopping. These component measures are weighted into household measures using the expected frequency of access to each kind of opportunity: hence retired households have zero weight on accessibility to work, whilst households with three or more working-age adults have a high weight. The result is that nearly every one of the 40 household types has a different accessibility value for each zone.

The *cost of location* and *floorspace per household* variables are both related to the rent per unit floorspace in each zone; as already noted, rents are calculated iteratively within the residential location modelling process. A variation on the Stone-Geary function is used to estimate households' consumption of floorspace, given their incomes and the prevailing rent. The function as used in LLITM includes a minimum requirement of floorspace per household of each type, and a minimum level of expenditure on other goods and services (ie other than housing).

The *environmental quality variable* in LLITM is a measure of "traffic density" provided by the transport model and taken as a proxy for the range of negative effects caused by traffic, including noise, emissions, and the risk or severance effects caused by moving vehicles themselves.

Housing quality is an index which is adjusted within the DELTA model: in the LLITM case it tends to rise if household incomes in a zone increase (reflecting increased ability to maintain and improve housing) and vice versa.

Residential relocation: model form

The model equations can be written in two different ways: as multinomial logit choice models based upon the absolute values of the "independent" variables, or as incremental logit models based on changes in those values. The former would typically include zonal constants or additional variables; in the latter form, the previous mixture of households is used as an explanatory variable, and the basis of the design is that the model responds to the changes that have occurred in the "independent" variables since those previous households themselves located. This is done separately for two groups of households: those with a previous location (who don't have to (re)locate), and those with no previous location in the area where they are going to locate (ie newly-formed households and arriving migrants, who do have to locate). The incremental form of the model implements the hypotheses that

- households with an existing location who don't have to relocate will tend to remain where they are unless the independent variables (or their budgets/requirements) change
- households with no existing location will tend to locate where similar households have located in the past, unless the independent variables change

The calibration is done on the non-incremental form of the model.

Calibrating the residential location model

An ideal calibration of the DELTA residential location model would require disaggregate data on a large sample of households moves, including sufficient detail to classify the household and "before" and "addresses" sufficiently precise to allocate them to zones (in order to add zonal data, whether observed or modelled) to the independent variables. Such data exists in the Census and some other surveys, but is essentially unavailable due to their confidentiality restrictions. In addition to such Revealed Preference (RP) data, Stated Preference (SP) data would also be highly desirable, for the standard reason of reducing the problems due to correlation among other variables.

Because of the limitations on using such data from standard, public surveys, and the costs of assembling new data, much of the calibration of DELTA models has been done as "secondary" calibration making use of results from other researchers' analyses. The main exceptions have been the study supported by the Engineering and Physical Sciences Research Council at the Institute for Transport Studies, University of Leeds, which carried out a Stated Preference study in connection with the development of the prototype DELTA application (Bristow et al, 1998) and the more recent work for Auckland Regional Council. The latter (reported by Feldman and Simmonds, 2009) was intended to obtain both RP and SP data from a sample of movers; the RP component was fatally compromised by the failure to collect full addresses during the fieldwork (which in that case was done by telephone), but the SP component was moderately encouraging.

LLITM survey and the data obtained

The LLITM household survey was a mail-out, mail-back self-completion survey sent to approximately 30000 households across the county, of whom 1880 sent back usable forms. The core of the data obtained was about household composition, car ownership and travel, including a one-day travel diary for each household member. The additional questions for modelling relocation asked how long households had been at their current address, and – if they had been at their current address for less than 5 years– where they had moved from. Some 223 households reported having moved within Leicestershire in the last 5 years, and these provided the sample on which the calibration of the relocation model was attempted.

The sample shows a distinct bias towards the older population. This could be entirely a matter of response bias – with older respondents perhaps being more likely to complete a paper questionnaire, and also perhaps being somewhat more likely to have regular travel patterns that would be easier to record in the travel diary. This must be kept in mind in all further analysis and discussion.

The proportion of households who have moved in the last year is very low, at 4.3%. This is much lower than the data we have from previous analysis of the Survey of English Housing (SEH), which gives an overall figure of 9%. Some of this lack of mobility could be attributed to the bias towards older households, who are less mobile; however the data shows less than average mobility in each of the 10 DELTA household age/composition types. This is particularly true for young single households; the SEH figure is 21% moved in the year before they were surveyed, LLITM has only 6%.

Some of the shortfall may be due to the particular conditions in the housing market over the year preceding the LLITM survey (in contrast with the SEH data which is pooled data from fieldwork over three years earlier in the decade). Some published data on the English housing market suggests that the number of housing market (purchase) transactions during the winter of 2009-10 was possibly one-third down from previous years; but the "last twelve months" in the LLITM data would also include the summer of 2009 which appears to have shown only a slight slow-down compared with previous years. Other sources suggest that the decline in sales transactions happened earlier. There is also the possibility that growth of the rental market has weakened the relationship between sales and moves – more households are (presumably) making rental-to-rental moves, and owner-occupier households who have to move may choose let their present home and rent their new one – temporarily if not permanently.

The data shows that most households have had more recent changes of job (or loss of job, or retirement) than of residence. This would tend to support the hypothesis that workplace choice is more often conditional on home location rather than vice versa, but since the data seems to be biased towards less-mobile households, even within each DELTA household category, we cannot draw any firm conclusions.

316 households reported a previous address. Of these 240 gave an address (postcode) within the Fully Modelled Area which we were able to use. These usable postcodes were converted to coordinates using the centroid for each unit postcode. Straight-line distances were then calculated from the coordinates. 187 households out of 240 moved less than 10Km on this definition.

For estimation, a new dataset was created containing the data on households who reported that they had moved in the last five years, plus modelled information. The sample contained 223 households. Their current and previous locations were converted to LLITM zones using the reported postcode information.

Estimation method and independent variables

The estimation method set out to account for each household's residential zone choice, given their previous location and the type of household involved. Their current zone (ie the one where they were living when surveyed) obviously represents the "chosen zone". In theory, we could include all the other zones in the model as "rejected alternatives". This is impractical, and we therefore adopted the standard method of randomly assigning other zones to each household to represent "rejected alternatives". Different methods of random assignment were tried as reported in discussion of the results.

The following explanatory variables were added to the dataset, for the chosen and rejected zones, from the LLITM land use model database:

- Accessibility (in minutes of generalised cost: logsum measure of accessibility for the household (ie expected generalised cost of a trip to a particular type of destination) weighted by average trips per purpose per household per week) (NB this is different depending on the household type)
- Floorspace per household (square metres) (discretionary floorspace used later – this is floorspace per household minus a minimum quantity, where the minimum depends on the household type)
- Quality (index)
- Zone size (floorspace based), the natural log of total housing floorspace

- Zone size (household based), the natural log of total households
- Distance from previous zone to current or rejected alternative zone (Km)
- Natural log of distance deterrence function ???to explain what the deterrence function is???

Estimation results

A series of models were then run of the form:

U(Choice)=βaccess.ChoiceAccess+βcostloc.ChoiceCostLocation+βFlspHHld.ChoiceFlspPerHHld+...

U(Alt1)=βaccess.Alt1Access+βcostloc.Alt1CostLocation+βFlspHHld.Alt1FlspPerHHld+...

U(Alt2)=βaccess.Alt2Access+βcostloc.Alt2CostLocation+βFlspHHld.Alt2FlspPerHHld+...

U(Alt3)=βaccess.Alt3Access+βcostloc.Alt3CostLocation+βFlspHHld.Alt3FlspPerHHld+...

..etc...

Table 1 shows what we hoped to see in the estimated coefficients.

The first reasonably successful model of residential choice which was estimated used a dataset including the chosen zone and five rejected alternatives. Previous, less successful, runs had already shown that floorspace per household and quality were highly insignificant factors in zone choice, that the natural logarithm of the distance deterrence function was a more significant variable than simply distance, and that either zone size term was significant with little variation in the size of the coefficient. The model therefore used the accessibility, cost of location, floorspace based zone size and natural logarithm of distance deterrence variables to estimate zone choice.

Table 2 (below) shows the results of this model. It is encouraging that the coefficients on accessibility and cost of location were of the correct sign and were both significant at the 90% level, these coefficients imply that increased costs of location and worsening accessibilities deter households from choosing one zone over another. The coefficients on the natural logarithm of distance deterrence and on zone size were both significant at the 95% level. However the coefficient on the natural logarithm of distance deterrence was further from 1 than desired, suggesting that the distance deterrence function in the LLITM land use model may need some adjustment, and the coefficient on zone size was of the wrong sign. The fact that the coefficient on zone size is negative suggests that households are choosing zones which are smaller than the rejected alternatives.

Table 1 Multinomial Model of Residential Zone Choice: Ideal Coefficients

Variable	Expected Coefficient	Explanation
Zone size (either floorspace or household based)	Close to 1	Theoretical, if the sample is large enough and the correct zone size variable is used, we would expect the number of choices to be proportional to the number of opportunities to move there.
Accessibility	Negative	Similar work in Bristol and Merseyside found coefficients in the range -0.01 to -0.07 using a minutes per trip accessibility measure. We would expect smaller values here as the DELTA accessibility measure is based on minutes per week.
Cost of location	Negative	Theoretical, households will chose to locate in more affordable locations.
Floorspace per household	Positive	Similar work in Sweden suggests a value of approximately 0.04.
Quality	Positive	Theoretical, households are more likely to choose to locate in an area where housing quality is high.
Distance	Negative	Theoretical, households are likely to choose a new location close to their old one.
Ln(Distance Deterrence)	Close to 1	Assuming that the distance deterrence function within the LLITM land use model is well calibrated the coefficient on the natural logarithm of distance deterrence should be close to 1.

Table 2 Results for residential zone choice model 1

Variable	Variable name	Coefficient	T-ratio
Accessibility	access	-0.000314	-1.81
Cost of Location	costloc	-0.00422	-1.92
Natural logarithm of distance deterrence	logdd	1.61	8.03
Floorspace based zone size	znsizeflsp	-0.265	-2.90

Further analysis of the LLITM household survey data suggested that there is a bias in the dataset towards households moving to small zones, the coefficient on zone size is reflecting this bias rather than anything else. We therefore decided to exclude the zone size variable from future model runs. The bias within the data is probably unsurprising given that the sample of movers represents only about one third of the number of zones in the zone choice set.

The next round of development of a multinomial model of residential zone choice focused on the distance term. Rather than include the natural logarithm of distance deterrence variable, the absolute distance between the previous zone and the alternatives was reintroduced, though this time the variable was split into two separate variables: distance up to 5km, and distance above 5km.

Table 3 shows the results of this model. The coefficients on accessibility and cost of location remain significant and of the correct sign. The coefficients on the two distance variables are also highly significant and both have a negative coefficient implying that households are more likely to choose to move to a zone close to their present location. The coefficients on the two distance variables are very different, once a household has moved over 5km each additional km has less impact on their zone choice. In other words for longer distance moved distance from current location is less important whilst for short distance moves households are highly influenced by distance from their current address.

Table 3 Results for residential zone choice model 2

Variable	Variable name	Coefficient	T-ratio
Accessibility	access	-0.000554	-1.99
Cost of Location	costloc	-0.00512	-1.96
Distance up to 5km	distTo5	-1.36	-4.46
Distance above 5km	Dist5Plus	-0.0895	-8.25

As a final stage in the model development it was decided that we should reintroduce a floorspace per household variable. As mentioned previously we had already found this variable to be insignificant however, when asked for their reasons for moving, LLITM household survey respondents reported housing needs to be a key factor. Within the LLITM land use model floorspace per household is the most reasonably proxy for housing needs. Rather than reintroduce the simple floorspace per household variable we have introduced a new variable, the natural logarithm of discretionary floorspace per household. That is the natural logarithm of floorspace per household less the minimum quantity of housing a household may occupy within the model.

Table 4 shows the results of this model. The coefficients on accessibility, cost of location and the two distance variables remain significant and of the correct sign. The coefficient on the natural logarithm of discretionary floorspace per household is significant and of the correct sign, increased discretionary floorspace per household attracts households to a zone.

Table 4 Results for residential zone choice model 3

Variable	Variable name	Coefficient	T-ratio
Accessibility	access	-0.000287	-1.68
Cost of Location	costloc	-0.00563	-2.25
Distance up to 5km	distTo5	-0.793	-8.36
Distance above 5km	Dist5Plus	-0.0794	-7.69
ln (discretionary floorspace per household)	FlspPerHHld	1.82	3.21

We consider this to be the most successful model of residential zone choice which we have estimated using the LLITM household survey data. It is possible to calculate odds ratios from these coefficients; this process leads to the following findings:

- A household is 0.03% less likely to choose a zone for every one minute worsening in accessibility relative to the alternative.
- A household is 0.56% less likely to choose a zone for every £1 per week increase in cost of location relative to the alternative.

- A household is 54.8% less likely to choose a zone for every additional 1km it is away from its current address, up to 5km. Beyond 5km a household is 7.63% less likely to choose a zone for every additional 1km.
- A household is 3.5 times more likely to choose a zone where the discretionary floorspace it could occupy is twice that in another zone.

The ratio of the accessibility coefficient (utility per minute) to that on cost (utility per £) implies a value of time of about £0.05/minute or £3.06/hour. This is low compared with standard values of time – though we find it reassuring that it is of the expected magnitude. The low value could be due to the characteristics of our sample (this is almost certainly a contributory factor), the way in which our accessibility measures are built up, the way our rents are estimated, and/ or differences in the way accessibility is perceived in location choices as distinct from the travel choices in which standard values of time are estimated.

The coefficients in Table 4 provided the starting point for further adjustments and the addition of non-zero coefficients on quality and environment in the process of setting up the working version of the LLITM model.

Conclusions and lessons for future work

Our conclusions are that

- we were able to do some useful calibration on data obtained from a simple add-one to a conventional travel diary survey, which has helped to design a small improvement to the function used in the model and (at the very least) to provide some confirmation for the relationships and coefficients used in the working model;
- the main limitation on the analysis was the size and composition of the sample of movers.

For any comparable future work, the first priority would seem to be to seek to obtain a larger and more representative sample. Depending on the survey approach chosen, the possibility of adding Stated Preference questions, as in the earlier Auckland study, would again be of interest.

Acknowledgements

We would like to thank the client organizations for the opportunity to carry out this work; colleagues at AECOM for their work on the design of the survey (including lengthy discussion of the questions intended to assist the land-use modelling) and its execution; and our colleagues both at AECOM and DSC for the modelling work which contributed the independent variables to the analysis. The authors remain responsible for any errors of commission or omission. Our thanks too to all the Leicestershire residents who provided the data used.

References

Bristow, A, A D May, S Shepherd (1998): Land use - transport interaction models: the role of environment and accessibility in location choice. *Proceedings of the 8th World Conference on Transport Research*, Antwerp.

Dobson, A, D Simmonds, O Capon, A Potter and A Ash (2007): Developing a sub-regional transport strategy for regeneration and growth. Paper presented to the *European Transport Conference*, 2007. Available at www.etcproceedings.org.

Dobson, A C, E C Richmond, D C Simmonds, I Palmer and N Benbow (2009): Design and use of the new Greater Manchester Land-Use/Transport Interaction Model (GMSPM2). Paper presented to the *European Transport Conference*, 2009. Available at www.etcproceedings.org.

Feldman, O, Simmonds, D (2009): Use of the Auckland Land-Use Survey in Calibrating the Household Location Model. Paper presented to The 12th International Conference on Travel Behaviour Research, Jaipur, India, 13-18 December 2009. Available at www.davidsimmonds.com.

Simmonds D.C. (1999). The design of the DELTA land-use modelling package. *Environment and Planning B: Planning and Design*, pp665-684.