Modelling residential location and travel decisions using detailed revealed preference data

Md Bashirul Haque

Submitted in accordance with the requirements for the degree of Doctor of Philosophy

The University of Leeds

The Institute for Transport Studies

November, 2020

Intellectual property and publications

The candidate confirms that the work submitted is his own, except where work which has formed part of jointly-authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

The work in chapters 3, 4, 5 and 6 consists of original research, where the candidate was the lead researcher. The main ideas in these chapters were developed by the candidate during the PhD with the support and guidance of the supervisors (Charisma Choudhury and Stephane Hess) and other researchers, who have been listed as co-authors where appropriate.

The candidate performed the data processing, data analysis, modelling work and wrote the manuscript. The other authors provided recommendations on the modelling and comments on the results. The manuscript was improved by comments from all the co-authors. The manuscripts are summarised below;

The work in Chapter 3 and 4 of this thesis has been published as follows;

> **Haque, M. B.,** Choudhury, C., Hess, S. and Crastes dit Sourd R. (2019). Modelling residential mobility decision and its impact on car ownership and travel mode. *Travel Behaviour and Society.* 17, pp. 104-119.

The work in Chapter 5 of this thesis has been published as follows;

> **Haque, M. B.**, Choudhury, C. and Hess, S. (2020). Understanding differences in residential location preferences between ownership and renting: a case-study of London. *Journal of Transport Geography*. 88, pp. 102866.

The work in Chapter 6 of this thesis has been published as follows;

> **Haque, M. B.,** Choudhury C. and Hess, S. (2019). Modelling residential location choices with implicit availability of alternatives. *Journal of Transport and Land Use.* 12(1), pp. 597-618.

Acknowledgements

I am grateful to many people for their generous support, advice and encouragements during the journey of my PhD. At first, I would like to express my sincere gratitude to my supervisors Charisma Choudhury and Stephane Hess for their valuable guidance, assistance and insightful thoughts and comments on my work. They were always very keen on finding opportunities for me to attend training, seminar, conferences for improving my skills and knowledge.

I would also like to thank my two examiners Brett Day and Thijs Dekker who made my viva an interesting discussion from which I learnt to look at my work from a different perspective. The critical comments of my examiners have helped me greatly to improve the thesis. Thijs had assessed my work at the transfer stage and I thank him for his useful comments and suggestions that helped to shape my thesis.

I would like to thanks those who helped through useful discussions and suggestions for shaping my PhD. Special thanks to Andrew Daly, Romain Crastes dit Sourd and other staff members and PhD students of the Institute for Transport Studies, particularly those in the Choice Modelling Centre.

I am grateful to the Economic and Social Research Council (ESRC) for providing financial support through the ESRC White Rose Doctoral Training Centre and the Institute for transport studies for partially funding my tuition fees.

Data used in this research was obtained from different institutes and sources such as the Greater London Authority, Transport for London, Office for National Statistics and Institute for Social and Economic Research, University of Essex. I am very thankful to them for making the data available for me.

I would like to mention my employer Shahjalal University of Science and Technology for granting leave for pursuing my dream.

I would like to thanks to my mother, parents in law and my siblings for their moral support. Finally, I am very grateful to my wife Irin and my son Ihan for their countless support.

Abstract

Choices regarding residential location are closely linked with travel behaviour. Mathematical models of residential location choice and travel decisions can be used to quantify how these interdependent decisions are influenced by the location and transport attributes and the socio-demographic characteristics of the decision making household or individual. While, revealed preference (RP) data is the most dependable and unbiased source of data to capture the interdependencies among the residential and travel decisions – missing information and coarse spatial and temporal resolution of such data makes it very challenging to use it for developing detailed residential choice and travel behaviour models. This study aims to model household residential and travel decisions and their interdependencies capturing some of the crucial behavioural modelling issues.

The residential decision of a household is typically a two-step process: residential mobility decision and residential location choice. Existing models have weaknesses in terms of capturing the geographical scale of the residential mobility decision (i.e. whether to move local, regional or national level) and its impact on household travel decisions. Models to predict the geographic scale of the residential mobility have been developed in this research using the British Household Panel Survey (BHPS) dataset. Further, while capturing the role of residential mobility on car ownership and mode choice decisions, existing studies have considered each direction of shift in car ownership change (e.g. gaining first car, gaining additional car, etc.) and mode choice (e.g. switching from car to public transport, car to active travel, etc.) in separate models. To fill in this research gap, this study attempts to jointly explore the multiple dimensions of changes in a single econometric model.

On the location choice aspect, this work also provides important behavioural insight into how the residential location preferences of two major housing markets (ownership and renting) are different from each other. The London Household Survey Data (LHSD) is combined with the Ward Atlas Data (WAD) of Greater London area and travel distance data from the London Transport Studies Model (LTSM) to get a comprehensive set of factors influencing the zonal level choice of residential locations.

The residential location preferences modelled in this work are complex due to unobserved choice set for individuals and the large size of the universal choice set.

The probabilistic approach and heuristic based methods available in the literature are likely to have weaknesses in terms of capturing behaviourally realistic choice sets in the context of residential location choice. This research makes advancement in the context of choice set generation by proposing an improvement of the state-of-the-art= semi-compensatory choice set construction technique. The proposed technique has better performance over other available semi-compensatory techniques.

The empirical results using the RP data provide insights for urban and transport planners by enabling them to better predict the residential and travel decisions in alternative policy scenarios.

Table of contents

List of Tables

List of Figures

Chapter 1
Introduction

1.1 Background and motivation

The global urban population increased by more than ten-fold during the 20th century. In 1900, the global urban population was less than 15% which increased substantially to 50% in 2005 (Satterthwaite, 2005). In the UK, 63 cities constituted more than 55% share of the country population in 2011. By the year of 2036, major cities in the UK will receive around 21% more people than in 2011 whereas non-urban areas will receive only around 12% more in the same time period (Champion, 2015). The high population growth in urban centres is creating substantial additional demand for transport, housing, employment and other facilities leading to traffic congestion and putting pressure on the public transport services. Due to inadequate spaces in the city core areas, new developments specifically housing, are being placed on the fringe side of the city and pushing the boundaries of developments further from the city centre (Rhoads and Shogren, 2006). People living in the fringe areas are more likely to accept long travel distance for work, education and leisure trips. Long travel distance also increases auto dependency and/or increased usage of public transport. These warrant a substantial increase in transport supply (e.g. increase in road capacity, new public transport lines, etc.). However, the supply-side solutions for transport system management have been criticized for yielding only short-term benefits as the congestion tends to return to its original state in the long run (Fulton et al., 2000; Noland, 2001). Demand-side management is hence advocated in the literature for easing traffic congestion. Efficient land development can be an effective tool for transport demand-side management. For example, promoting mixed land development where people have easy access to facilities such as job, shopping, schooling from their home can reduce the trip length and car dependency (Ewing and Cervero, 2010). On the other side, transportation planning decisions also influence land use patterns and land development. For instance, improving urban highways for faster mobility encourages development of more detached neighbourhoods and automobile-oriented urban sprawl, while walking, cycling and public transit-oriented transport planning encourages compact and mixed development (Litman, 2016). Thus, the transport and land use developments are interdependent and integrated urban modelling can be a useful tool for sustainable urban development. The connection between these two elements is demonstrated in Figure 1.1.

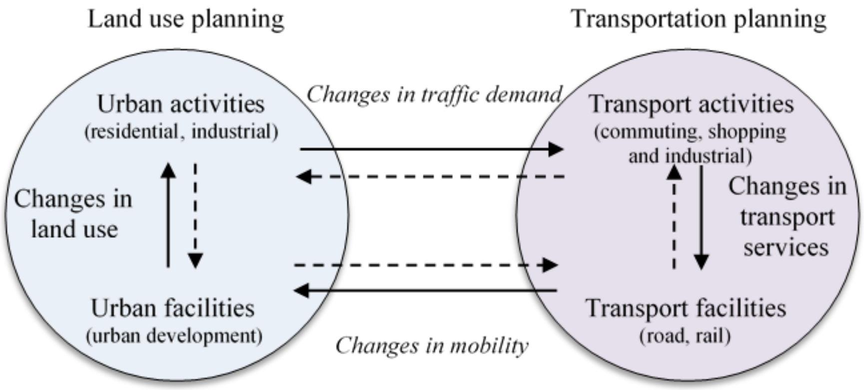


Figure 1-1 Relation between land use and transport planning (Morimoto, 2015)

A significant level of progress has been achieved in integrated modelling of the urban system, however, the proposed models in the literature have limitations in terms of capturing the full range of behavioural dynamics and multi-dimensional decision interactions at the level of agents (individual or household) (Fatmi, 2017). Residential change, job change and car ownership are interconnected life trajectory decisions and have a significant consequence on daily travel behaviour. If a change of workplace of a household member leads to a long commute, (s)he may consider a change in daily travel mode and/or the household may consider changing their residential location. Changes in life events also incur complex dynamics at the household level. For example, birth of a child may prompt a household to move to a bigger dwelling. Ignoring the dynamics and interdependencies at the household or individual level decisions may lead to inaccurate prediction and forecasting of urban model components.

Residential change (moving home) is a special biographical moment, in which familiar routines are likely to be broken (Scheiner, 2006). It can be assumed as a two-tier individual or household level decision. In the upper tier, households decide whether they want to move (decision of residential mobility) due to changes in their circumstances such as getting married, changing employers, an increase in the number of members in the household, etc. The new circumstances may also determine the geographical scale of residential mobility (how far households need to move). For instance, getting or switching job to another metropolitan area may lead to an inter-city move whereas the requirement of extra space can be met by a relocation within a short distance from the previous location. Therefore, both the decisions of residential mobility and geographical scale of residential mobility are driven by the changes in household circumstances. In the second tier, households choose neighbourhoods or locations or dwellings (usually called residential location choice) from a set of possible options in the area or city they need to move. Therefore, the choice of a residential location is mostly affected by the attributes of the alternatives (e.g. quality of the schools, transport accessibility, dwelling price, etc.).

Geographical scale of residential mobility can be at local level (moved within the same ward or zone) or regional level (moved within the metropolitan area or region) or national level (moved in another metropolitan area or region) and different scale of residential mobility can affect the household social network, travel choices and other circumstances differently. If a household moves at the national level, the social capital that the household has built over the years is likely to be lost (Aditjandra et al., 2012; Lin et al., 2018). Moreover, regional and national level residential mobility can considerably affect household transport and other forms of accessibilities leading to changes in travel patterns and car ownership levels[1]. For example, if households moved into an area with poorer public transport accessibility, car owing propensity is most likely to increase and, in turn, owning a car will change travel behaviour. Thus, it is observed that the intensity of changes in household car ownership and travel behaviour after relocation are necessarily linked with the geographical scale of residential mobility. Interlinkage between household residential location, car ownership and travel decisions are presented in Figure 1-2.

As mentioned in the previous section, the choice of a residential area or a neighbourhood can also affect household travel choices. Within a metropolitan area or region, nighbourhood characteristics and land use patterns can vary significantly. Disaggregate level residential location choice model can capture the variability at a micro-level such as dwelling, parcel or zone level (Zolfaghari, 2013). For selecting an alternative in disaggregate level approach (the neighbourhood to move), households consider a pool of potential alternatives (dwelling, parcel or zone) and finally select one from the pool which they perceived to be the best.

---

[1] Individual travel pattern includes the number of daily trips, length of each trip, mode of travel, etc. and car ownership level indicates how many cars a household owns for private use.

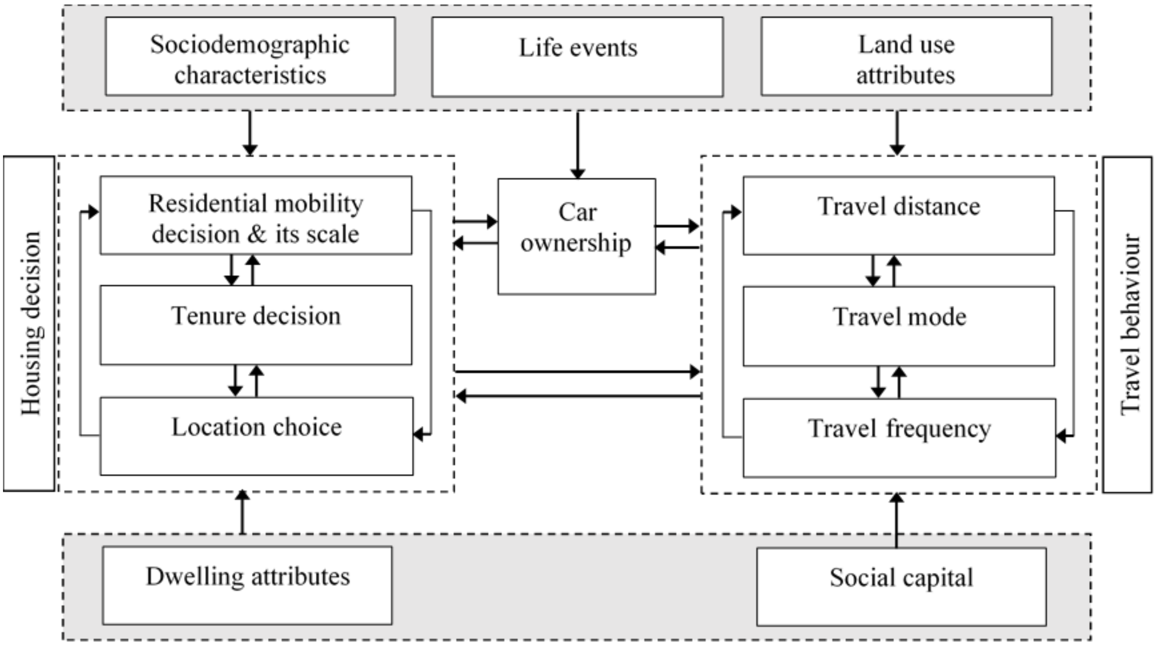


Figure 1-2 Interdependencies between residential decisions and travel choices

For example, based on the circumstances, dwelling and neighbourhood preferences of one group of households can be different from other groups. In developed countries, wealthy and older people prefer to live in the low-density suburban areas for increased status, a higher level of safety and like-minded neighbours (Aero, 2006; Pisman et al., 2011) and commute a long distance daily for working in the city area (Rivera and Tiglao, 2005; Bill et al., 2006; Watts, 2009). However, in the developing countries, life in the city is better in terms of facilities like educational, healthcare, recreational, overall quality of life, therefore, wealthy people prefer to live in the city area (Choudhury and Ayaz, 2015; Lin et al., 2018). Residential preferences could be different among other demographic groups such as working male and female (White, 1977), single and multiple working member households (Curran et al., 1982; Van Ommeren et al., 1998), different household types such as young married couple, married couple with parents, old couple, couple with child, couple with parents and child (Jiao and Harata, 2007), knowledge-workers and unskilled laboures (Florida, 2002; Frenkel et al., 2013), etc.

On the other hand, there could be potential heterogeneities in the preferences of different housing markets such as ownership and renting. Renters have a higher rate of residential mobility with shorter tenure length in most of the cases due to low initial investment, flexible contact and low housing search cost. On the other hand, owners are likely to be more stable in their place due to high investment resulting in longer tenure length. Therefore, the priorities and compromises of these two groups are likely to be different. Due to the higher flexibility of moving home, renters put less effort into searching houses in some cases and made an irrational decision (e.g. accept long commute distance, less shopping access, unwelcoming neighbourhood, etc.). On the other hand, owners invest more time and money to find the best option and very unlikely to compromise in their basic criteria. Anticipation can play a significant role in long-term ownership decisions. Before buying a house, households are more likely to consider possible events they are anticipating in the next few years such as having a baby, switching job, owning cars, etc. Sociodemographic characteristics and travel behaviour of these two groups are also different. In general, high-income households are more likely to own houses while low-income households prefer social or private renting (Yates and Mackay, 2006). Due to the higher rate of car ownership, owners are more car-dependent and more likely to make a higher number of trips compared to the renters. Based on the discussion in the preceding sections, a significant level of preference heterogeneities is anticipated in the residential location choice of owners and renters.

People's behaviour and preferences are also changing over time. Over the last few decades, car use has increased enormously across the world. In the UK, household access to car shifted from 14% to 75% since 1950 (Goodwin et al., 2004). A higher level of car ownership increases household car-dependency, trip making propensity and trip length. Over the last 50 years, the average travel distance in the UK has increased from 4000 km to 12000 km per year without air travel (Goodwin et al., 2004). An increase in car ownership level over time changes household preferences in their residential location choice. People started living far from their workplace and the city area and using the car more for work, shopping and other trips. For example, from the year 2001 to 2011, the average commute distance in the UK has increased by 12% (Gower, 2014) which indicates that household commute distance sensitivity has decreased over time. Technological advancement is also contributing to changing household lifestyle and preferences. Telecommuting (home-based employment or working from home) is being increasingly popular last few decades. It increases the likelihood of living far from the workstation (Tayyaran et al., 2003) because people don't need to travel for work daily. Future advancements in technology also have the potentiality to profoundly change the transportation system (Fagnant and Kockelman, 2015) eventually leading changes in household preferences of residential and other choices. If people don’t need to drive and can work while in the car, they might accept a longer commute distance. People's preferences on land use patterns and characteristics in the residential neighbourhood are also changing. Transit-oriented mixed land use pattern has been observed to gain popularity over car-dependent suburban areas in the last few decades (Burda, 2014). Although, many suburbanites do not like this for fear of decreasing property values, having strange neighbours (Baar, 1992) and neighbourhood racial succession (Nelson, 1997).

Despite of capturing several behavioural aspects, residential location choice modelling based on revealed preference (RP) data has numerous methodological and empirical challenges. Choice set construction is one of them. In RP based residential location choice models, the number of alternatives in the universal choice set is very high and the researcher does not have information about the actual choice set of individuals. In zone level models (where small geographical area such as zone is considered as a location alternative), the number of alternatives in the universal choice set can be hundreds to thousands that can be hundreds of thousands in case of dwelling level models (where each dwelling is considered as an alternative). However, stated preference (SP) based residential location choice model has very limited choice alternatives and dataset consists of the true choice set of individual respondents (Walker and Li, 2007; MALAITHAM et al., 2013; Choudhury and Ayaz, 2015). Due to the lack of information about the actual choice set of individuals, several RP data based studies in the literature have assigned universal choice set as individual choice set (Zondag and Pieters, 2005; Chen et al., 2008). Estimation of the models considering the full choice set for each individual is computationally challenging and behaviourally non-representative. It is very unlikely in the context of residential location choice that households consider all possible alternatives during the decision-making process. Moreover, the massive universal choice set flatten the choice probability distribution and reduces the predictive power of the estimated model (Wegener, 2011). In reality, households consider a small set of credible alternatives from a dynamically changing choice set for making a final decision (Zolfaghari, 2013). In literature, different approaches have been proposed for behavioural choice set construction capturing the underlying search mechanism (Martínez et al., 2009; Farooq and Miller, 2012; Rashidi et al., 2012). Most of these methods are based on heuristics, therefore, have different levels of challenges and limitations alongside potential strength.

1.2 Research gaps

The previous sections highlighted the connection between the integrated urban system components and behavioural issues and the challenges of modelling different components. Numerous studies have tried to capture the behavioural aspects in residential mobility decision, residential location choice and travel decision (as detailed in Chapter 2). However, still, there are several gaps in the literature and this research is aimed to fill a few of them.

RG1. The existing literature of residential mobility decision mostly covered the role of life events (Bartel, 1979; Clark and Davies Withers, 1999; Beige and Axhausen, 2012), tenure type (Clark et al., 1986; Van der Vlist et al., 2002), social capital (David et al., 2010) and neighbourhood characteristics(McCulloch, 2010) on residential mobility decision. However, the decision regarding the geographical scale of residential mobility decision has not been investigated, yet it has significant importance as discussed in section 1.1.

RG2. The geographical scale of residential mobility is likely to have varying impacts on mid-term (e.g. car or transit pass ownership) and day-to-day mobility decisions (e.g. mode choice for a specific trip for example). However the focus of existing studies has only been on capturing the impact of residential mobility decision (irrespective of its scale) on travel decision (Prillwitz et al., 2006; Oakil, 2013; Clark et al., 2014; Clark et al., 2016b; Lin et al., 2018).

RG3. As discussed in section 1.1, residential location choice also has several behavioural issues to capture. Many of them have already been investigated in the literature such as gender role on residential location choice (White, 1977), residential location choice of multiple working member households (Curran et al., 1982; Van Ommeren et al., 1998), residential location choice of knowledge-workers (Florida, 2002; Frenkel et al., 2013), the role of ethnic segregation on residential location choice (Ibraimovic and Hess, 2016) However, the potential behavioural differences in the housing markets (owners and renters) and the time-varying nature of household preferences remain as a research gap.

RG4. Choice set formation in modelling residential location choice is also very challenging. Most of the existing studies have considered heuristic-based approaches (Martínez et al., 2009; Farooq and Miller, 2012; Rashidi et al., 2012). However, these approaches have limitations in terms of capturing the true choice set (Bierlaire et al., 2010; Zolfaghari, 2013). In the literature, it is not clear which method is best for capturing behavioural choice set for disaggregate level residential location choice modelling.

1.3 Research objectives

This research aims to develop a framework of modelling residential decision (residential mobility decision and choice of residential location) and how it relates to household car ownership and travel behaviour using large-scale RP data. To fill the research gaps mentioned in section 1.2, the specific objectives of this research are presented below

a. To develop econometric models for residential mobility decision and its geographical scale (RG1).
b. To investigate the role of the geographical scale of residential mobility on changes in household car ownership and commute mode (RG2).
c. To develop models for residential location choice capturing the behavioural differences between residential ownership and renting (RG3).
d. To propose an improved technique for constructing choice sets with a better behavioural underpinning in the context of residential location choice (RG4).

This study uses RP data for achieving the research objectives in the subsequent chapters. RP data has the potential to better capture the true behaviour with accurate parameter estimation avoiding the potential bias associated with hypothetical responses in stated preference data. Moreover, the panel nature of RP data also enables capturing the dynamics in the household decisions over time. However, RP data has challenges in terms of capturing a comprehensive range of decisions using a single dataset due to missing information, an inadequate number of observations, etc. Therefore, this research required to use two different RP datasets to fill the research gaps: British household panel survey (BHPS) data and London household survey (LHS) data (detailed in Section 3.2 of Chapter 3 and Section 5.2 of Chapter 5). The BHPS dataset is used to investigate the research objectives a and b. In the BHPS dataset, the number of households who moved in a single housing market or region (e.g. London) is very few, therefore, the LHS dataset is used for investigation of the research objectives c and d.

1.4 Research approach

Households or individuals make many short-term and long-term decisions in their daily life and over their life course. Since most of the decisions are interconnected, the total structure of the decision interdependencies is likely to be very complex (an example is presented in Figure 1-2). However, all the units of this complex decision structure and all the directions of interdependencies may not have the same level of importance from a behavioural, modelling and policy point of view. Moreover, it is very complicated and challenging to model the entire structure of decision interdependencies, not to mention the lack of suitable data that has the full set of information to capture all the interdependencies. Therefore, a simplified modelling framework is adopted for this study (Figure 1-3). The justifications regarding the assumptions made for simplification are discussed below.

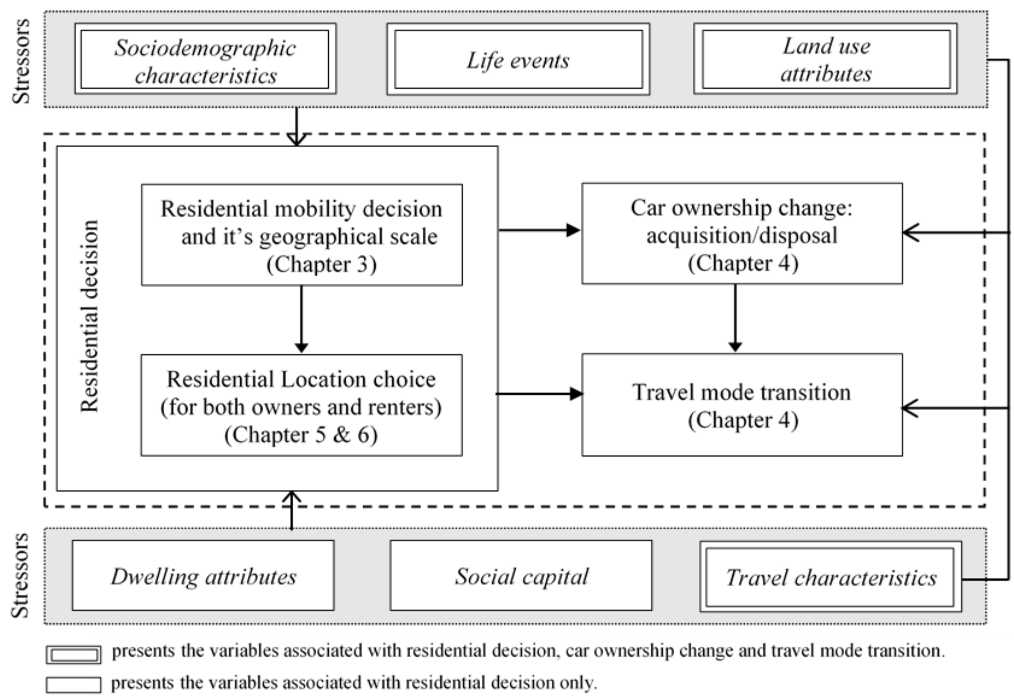


Figure 1-3 Research framework

The components of the residential decision presented in Figure 1-2 are sequential is most cases (Harold and Leonard, 1991). Therefore, this study attempts to model the residential mobility and residential location decisions sequentially where the tenure decision is assumed to be exogenous (Figure 1-3). However, tenure is considered as an explanatory variable in the model to capture the role of tenure in residential mobility behaviour. In addition, the location choice behaviours of two major tenure groups are modelled separately to investigate the potential differences in the preferences of these two groups. In the case of residential mobility and location choice, households are generally assumed to first consider to move (upper layer) then they search for the best location to move (Habib 2009). Intuitively, the reverse causality is very rare where household mobility decision is conditional on the choice of location and this is thus ignored in this study. Therefore, this study is aimed at modelling residential mobility behaviour of the households first, followed by the choice of locations of the households who moved (the same approach has also been used elsewhere in the literature, e.g. Habib, 2009). Finally, other than the two possible sequential orderings, it should be acknowledged that it is not impossible that for some decision makers, the choice of tenure and location can be simultaneous. This study focuses on the most obvious sequence, but for future work, a latent class model can be used to capture the heterogeneity in decision rules[2].

As mentioned in Figure 1-2, the residential decision may have bidirectional relation with car ownership and travel choices. However, both directions of the relationship may not have the same strength and consequences. For example, residential relocation can change household travel behaviour (e.g. Clark et al. 2016b, Fatmi and Habib 2017, Lin et al., 2018) and car ownership level (e.g. Oakil et al. 2014; Clark et al., 2016a) significantly, whereas, it is less likely that household change residential location due to changes in their travel behaviour and car ownership. Similarly, the changes in household travel behaviour specifically switching to or from car depends on the car ownership status change (Fatmi and Habib 2017). However, it is less likely that households lose their cars due to switching from the car for commute trips because households may need the car for non-commute trips and it is also rare that households switch to commuting by car (through renting for instance) before owning a car (the reverse causality). The other two components of travel behaviour (travel distance and travel frequency) mentioned in Figure 1-2 are not considered in this study due to data limitations.

Therefore, based on the research gaps identified and the discussion in the previous section, this study aims to model a few key components of household-level decisions such as residential choice (chapter 3,5 and 6), car ownership change (chapter 4), travel mode transition (chapter 4) and their dominant causality directions such as the impact of residential decision on car ownership and travel mode transition (chapter 4). The

[2] This is however beyond the scope of the current research and can be promising direction of future work if suitable data is available.

model components are estimated sequentially. It may be noted that given the rare events nature of these long-term decisions, some of the choice combinations (i.e. households who have moved at different geographical scales, changed car ownership in different levels and changed travel mode in different directions) do not have high representation in the data. For instance, there are no observations for a combination of households that have moved at the regional level, gained car(s) and switched from public transport to car. This posed challenges in simultaneous or joint estimation of decision components. It is acknowledged that this sequential decision estimation can under/overestimate the correlations among the decisions neglecting the inherent trade-offs and simultaneity in choice (Habib and Kockelman 2008).

A disaggregate level modelling approach is used in this research where each household is considered as a decision-making unit. Random utility maximization (RUM) theory is used for estimating the model parameters. Different formulations of RUM theory are considered in different chapters for specific aspects. For example, multinomial logit (MNL) and mixed multinomial logit (MMNL) models are estimated in chapters 3, 4 and 5 whereas constrained multinomial logit (CMNL) models are estimated in chapter 6. While different model components in different chapters offer important behavioural insights relevant to the respective modelling issues, they also contribute to large-scale integrated land use and transportation modelling, planning and policy implication.

1.5 Structure of the thesis

This thesis is organised into seven chapters. This chapter presents the background and motivation of this study, research gaps, specific research objectives and research approach. The contents of the rest of the chapters are outlined next.

Chapter 2 reviews the literature on the integrated urban model, modelling residential decisions, association of residential decision with travel decisions, modelling challenges and modelling approaches.

Chapter 3 investigates the geographical scale of household residential mobility decision using the British Household Panel urvey (BHPS) data. Residential mobility decision and the joint decision of residential mobility and its scales are captured in two separate models. Then the results are compared to investigate the contribution of geographical scales on household residential mobility decision. Sociodemographic characteristics, life events and travel behaviour of households moved at the local, regional and national levels are added as explanatory variables. Mixed multinomial logit models are estimated to capture the correlations in the repeated choices in panel observations and random taste heterogeneities across the households observed. A validation test is also performed to check whether the improvement in the MMNL model still holds in prediction.

Chapter 4 investigates whether the different geographical scales of residential mobility influence household travel decisions differently. Models are estimated to capture household car ownership and commute mode transition behaviour in two consecutive years. Models consider each direction of switching (e.g. car to public transport, car to active travel, etc.) as an alternative and alternative specific parameter sensitivities of households moved in different geographical scales are estimated along with other explanatory parameters such as changes in household sociodemographic, life events, etc.

Chapter 5 focuses on the modelling of residential location choices with attention to capturing the existence of preference heterogeneities in the different housing markets. London household survey data (LHSD) is used in this chapter as a primary data source. Residential location choices of owners and renters are modelled jointly but separate parameters are estimated for these two groups to investigate the differences in their preferences. Mixed multinomial logit modes are estimated to capture the household random taste heterogeneities.

Chapter 6 captures choice set construction for modelling residential location choices. This chapter critically discusses the limitations of existing semi-compensatory approaches and proposes an improvement of an existing technique. The proposed method is tested on the pooled model estimated in chapter 5 and compare the goodness of fit of this model with the models estimated using existing semi-compensatory approaches. The predictive power of the proposed improved model is also investigated using holdout sample validation.

Chapter 7 provides a summary of the key findings of this research, contributions of this research and presents some avenues of future research.

Chapter 2
Literature review

2.1 Introduction

Household level decisions can be long-term (e.g. residential decision, employment decision), medium-term (e.g. car ownership) or short-term decision (e.g. daily travel behaviour). Although these decisions are interdependent, long-term decisions have significant consequences on medium-term decisions and day to day travel behaviours. Several studies in literature attempted to address these issues. This chapter summarizes state-of-the-art knowledge in this context presenting a brief outline of the integrated urban model followed by the discussion on residential decision, the impact of residential decision on travel decision and modelling challenges.

2.2 Integrated urban land use and transport model

Integrated urban model (IUM) is a large-scale modelling system that simulates agent's (household or firm) decisions to predict the urban development and transport system. IUM consists of four interconnected components: location choice (residential, firm and job), auto ownership, activity/travel and land development (Miller et al., 1998). Each component consists of several submodels. For example, residential choice consists of residential mobility, tenure choice, location choice, etc. Model components are affected by temporal and long-term dynamics at individual and household level and the corresponding market dynamics. For instance, one life event (e.g. changing job, moving home) can bring changes to other domains of life by altering the daily routine and travel behaviour. On the other hand, changes in the housing market and labour market due to government intervention are most likely to affect household residential choices and employment related decisions respectively. An integrated urban model should capture all components of urban system and their interdependencies.

Several integrated urban models have been developed over the last few decades. The most notable integrated models can be split into five categories (Fatmi, 2017) which are presented in Table 2-1. The developed models have limitations in terms of capturing the drivers of the integrated urban system and behavioural representation. For instance, many of the advanced models did not consider vehicle ownership, although, it has a significant influence on household travel behaviour (Waddell, 2002; Ettema et al., 2007). In literature, many studies have found to address the several behavioural issues and empirical challenges of modelling integrated model components. The subsequent sections are focused on discussing the state of art knowledge on modelling household residential, car ownership, travel decisions and their interdependencies.

Table 2-1 Integrated urban models (source: Fatmi, 2017)

Categories	Integrated urban models
Economic activity-based model	• PECAS (Hunt, 2003)
	• MEPLAN (Echenique et al., 1990)
	• TRANUS (De La Barra et al., 1984)
Market principle model	• SelfSim (Zhuge et al., 2016)
	• ILUTE (Salvini and Miller, 2005)
	• MUSSA (Martínez and Donoso, 2004)
Quasi market-based model	• SimTRAVE (Pendyala et al., 2012)
	• ILUMASS (Wagner and Wegener, 2007)
	• UrbanSim (Waddell, 2002)
Hybrid model of Heuristic	• SILO (Moeckel, 2017)
	• PUMA (Ettema et al., 2007)
	• TRESIS (Hensher and Ton, 2002)
Emerging complex system Models	• SimMobility (Adnan et al., 2016)
	• POLARIS (Auld et al., 2016)
	• SynCity (Keirstead et al., 2010)

2.3 Residential decision

Household residential choice can be assumed as a two-step decision process. The first step consists of decision to move and its geographical scale that is driven by household sociodemographic characteristics and changes in their conditions and circumstances (Clark and Huang, 2004). Although geographical scale captures the implicit value of moving to a specific distance or a larger area, this decision is mostly independent of the characteristics of the new are to move and people may not have a choice in many cases. For instance, if an individual needs extra space for a growing family, (s)he might consider moving in the same area, but if someone enters university in another city, (s)he will consider an intercity move. In literature, many studies have investigated the residential mobility at different geographical scales such as residential mobility across cities (e.g. Bell, 2002; Bell and Rees, 2006; .Clark, 2013; Klinger and Lanzendorf, 2016), residential mobility within the city (e.g. Clark, Deurloo and Dieleman, 1984; Clark and Ledwith, 2006), residential mobility at local level (e.g. Speare, 1970; Chevan, 1971; Pickvance, 1974, Stillwell and Thomas, 2016).

In the second step, households evaluate a finite set of disaggregate level alternatives (e.g. dwelling/location/neighbourhood) within a geographical area they desired to move (same area or same city or another city) and select one they perceive to be the best. This process is mainly rolled by the characteristics of disaggregate level dwelling or location alternatives although the driving factor of residential mobility decision can influence the choice of location in some cases (Kim et al., 2005).

2.3.1 Residential mobility decision

Within the life course paradigm, changes in one dimension are necessarily linked with the changes in other dimensions (Dieleman, 2017). Therefore, changes in household composition, gaining jobs, getting married and changes in other circumstances influence household relocation decision that in turn influences car ownership level and travel behaviour. Many studies in literature tried to investigate the connection between life events and residential mobility decisions which are discussed in the following sections. Table 2-2 also presents a summary of a few key studies highlighting their limitations in terms of capturing the geographical scales of mobility, behavioural dynamics, etc.

The existing literature reveals that individual socio-demographic characteristics have a strong influence on individual and household level mobility decision. Employed people have a higher propensity to move house compared to unemployed people (Eluru et al., 2009). Changing job also increases the likelihood of residential move (Clark and Withers, 1999) while moving home often triggers the changing job (Bartel, 1979). The education level also influences the residential mobility decision. Highly educated people are more likely to have frequent residential moves due to having higher access to opportunities (Kortum et al., 2012). The role of marital status, gender, age of individuals on residential mobility were also found significant. For example, females have been found more likely to move house than males due to personal family-related issues (Eluru et al., 2009) and young people were found to move more frequently than older people (Clark et al., 1986; Van der Vlist et al., 2002).

Household level sociodemographic characteristics largely drive the household residential mobility decision. The rate of residential mobility was found to decrease with the increase of the household size (Eluru et al., 2009; Kortum et al., 2012). An increase in household income increases the propensity of a residential move. Middle-income people have a higher tendency of moving home whereas high-income people are less likely to do so (Capuano, 2011; Kortum et al., 2012). The presence of senior adults in the household reduces the likelihood of moving (Kortum et al., 2012). Immigrants have a high mobility rate but their likelihood of moving decreases with the time spent in the country (Kortum et al., 2012). The social network of household members was also found to affect the mobility rate negatively (David et al., 2010). Residential mobility decision is also linked with housing characteristics like housing cost, tenure type, etc. For example, owners are less likely to move and accept longer tenure length whereas renters move more frequently, which results in shorter tenure length (Eluru et al., 2009; Tatsiramos, 2009).

Neighbourhood characteristics also have a strong association with household residential mobility decisions. Better neighbourhood quality and school facilities decrease the likelihood of moving to a new place (Fack and Grenet, 2010; Kortum et al., 2012). There is a strong connection between travel behaviour and residential mobility decision. Car-dependent commuters are less likely to move houses compared to public transport or active travel mode oriented commuters (Eluru et al., 2009). Having a long commute distance was found to push people to relocate to save daily travel distance to work (Eluru et al., 2009).

The above discussion presents a picture of the relationship between the household characteristics and decision to move where it is assumed that the residential mobility decision is independent of the characteristics of the new location and other external factors such as housing market characteristics. However, housing market characteristics can play an important role in residential mobility decision (Van der Vlist et al., 2002). High tax rate, high mortgage rate, inadequate housing supply and other strict regulations can increase the cost of moving and living that can potentially decrease the mobility rate in a given area. Therefore, the distinct characteristics of individual housing markets can shape the residential mobility process differently from place to place (Pawson and Bramley, 2000). For example, in the UK, the highest loss of internal migrants has been observed in London over several years mainly due to the housing unaffordability which has been boosted up by the shortage of housing supply (Figure 2-1).

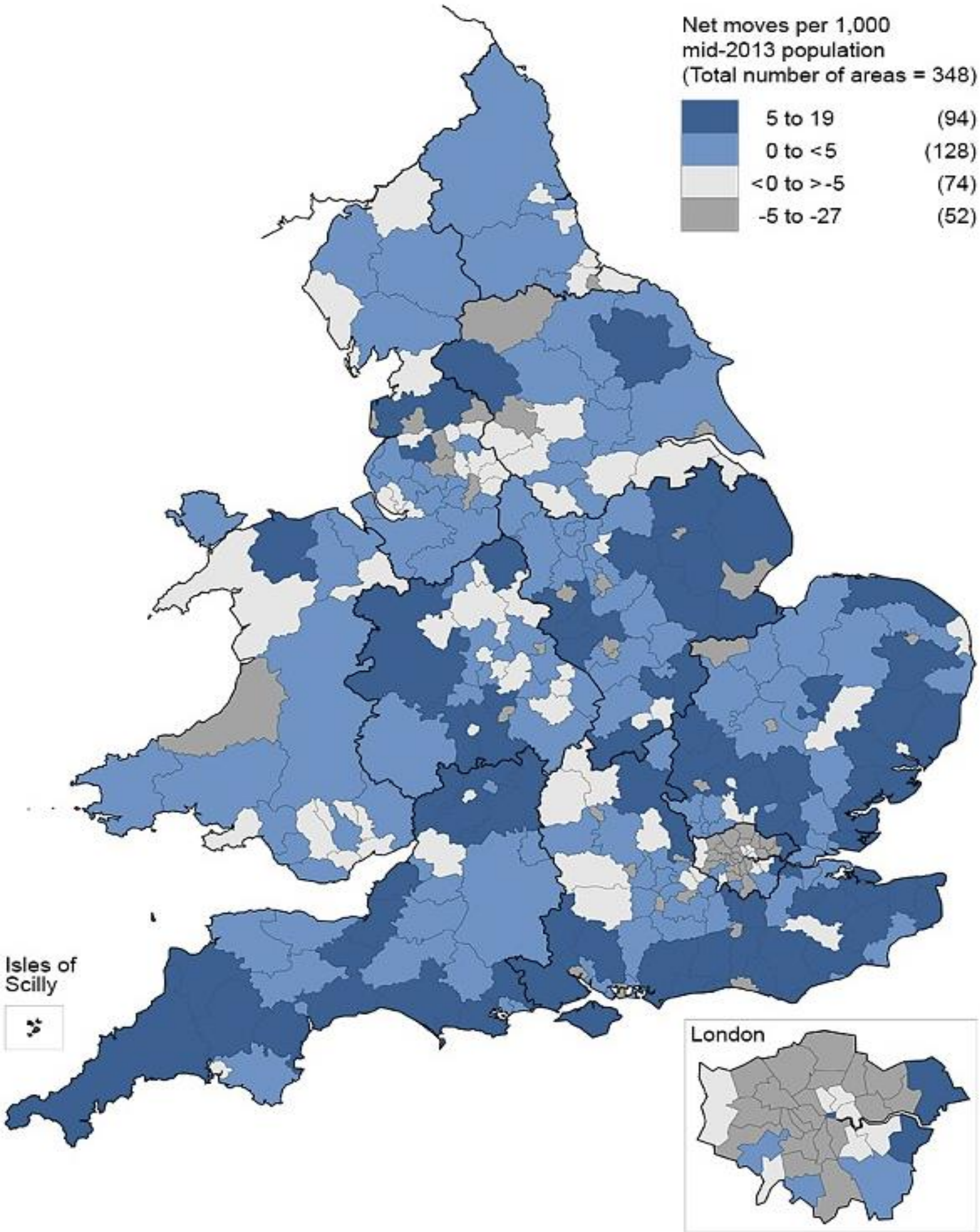


Figure 2-1 Net internal residential mobility in England local authorities (Source: Office for National Statistics)

Table 2-2 Summary of relevant studies on residential mobility

Author	Description
Bartel (1979); Clark and Withers (1999); Coulter and Scott (2015)	These studies have captured the association of job mobility with residential mobility along with other drives. Although all these studies have used longitudinal survey data, simple logit technique used in the Bartel (1979); Coulter and Scott (2015) studies did not allow to capture the dynamics and correlations in the household behaviour. However, hazard-based models estimated in Clark and Withers (1999) study allowed to capture the association between job change and residential mobility by providing a comparison of the timing of the behavioural changes and dynamic interpretation of the relations. None of these studies have considered or acknowledged the geographical scales of residential mobility, although, the geographical scales of residential mobility might have a different level of association with job mobility and other stressors.
Clark et al. (1986)	The core aim of this study was to observe the differences in the residential mobility behaviour of owners, public renters and private renters. Dutch National Housing Survey data has been used for estimating the model parameters using the logit technique which is less suitable for modelling time-dependent household behaviours. Moreover, the residential mobility behaviour of these three-tenure groups captured in this study could be different in different geographical scales. For example, due to the limited scope of work and ability, the social or public renters can have less propensity to move into other metropolitan areas compared to the owners and private renters. This potential behaviour difference has largely been ignored in this study.
Van der Vlist et al. (2002)	This study has explored that residential mobility is not only driven by household-level characteristics, but housing market characteristics can also play a vital role. However, the housing market parameters tested in this study were very few such as the size of the housing market and share of the socially rented houses. Only the intra housing marker mobilities have been investigated here, although, inter housing market mobility can also be affected by the characteristics of housing markets before and after the relocation.
Clark and Huang (2003)	Household residential mobility behaviour has been observed by Clark and Huang (2003) using British Household Panel Survey data (1991 to 1999). Discrete-time logit models have been estimated. Although this study has explored whether the residential mobility behaviour in the London housing market is different from the whole country, the potential differences in the drivers of inter and intra metropolitan relocations have not been investigated. The findings suggested that residential mobility is a demographically driven process, but it has a connection with neighbourhood contexts. They also observed that residential mobility in the London housing market shares similarities with the rest of the country but also reveals some differences.

Table 2-2 Summary of relevant studies on residential mobility (cont.)

Author	Description
David et al. (2010)	This study has explored the association between social capital and inter-regional residential mobility. Probit models have been estimated using the European Community Household Panel (ECHP) survey data. It has been observed that moving in the same neighbourhood is unlikely to affect the social capital that results higher rate of residential mobility locally, on contrary, moving in another city affects the social capital significantly resulting in lowering the likelihood of moving. This study findings support the necessity of investigating the geographical scale of residential mobility.
Tatsiramos (2009)	This study has captured the effect of unemployed benefit on residential mobility behaviour. Random effect profit models have been estimated using the European Community Household Panel (ECHP) survey data. The study has explored that the unemployment benefit does not have an adverse effect on the residential mobility behaviour in UK, French and Spain. Since the inter-regional mobility rate was very low, this result is more likely to be dominated by the behaviour of intra-regional movers. However, policymakers may need to know how unemployment benefit affects both intra and inter regional mobility behaviour.
McCulloch (2010); Rabe and Taylor (2010)	The role of neighbourhood quality on residential mobility decision has been addressed in these studies. British Household Panel Survey data has been used in both cases where McCulloch (2010) has used a multinomial logit regression model and Rabe and Taylor (2010) has used random effect logit model for estimating the parameters. The neighbourhood quality is characterized as whether the area is deprived or not. Households living in the deprived area have been found to have a higher propensity to move into the non-deprived area close to the current location without affecting the commute and other facilities. The relation between neighbourhood quality and residential mobility could be different in case of long-distance relocation due to job change or other big events. None of these studies have addressed nor acknowledged this issue.

2.3.2 Geographical scale of residential mobility

The geographical scale of residential mobility means where and how far a household relocates. This is one of the key considerations of residential mobility decision. In most of the cases, households move within a short distance or the same neighbourhood (Dieleman, 2017). This trend supports the basic concept of migration behaviour which is that the intensity of migration decreases as distance increases (Stillwell and Thomas, 2016). Households moved in the same neighbourhood or close to their past home is mostly for adjustment with dwelling needs (e.g. bigger house, better quality housing, end of the contract, etc.). The benefit of moving locally is that households do not lose the proximity of friends and families in many of the cases and also attain the benefit of community facilities and social capital they have built up over time. In this case, household travel and commute characteristics are unlikely to be affected significantly all else being equal. On the other hand, life events such as entering university, getting or switching job encourage long-distance mobility (Duke-Williams, 2009; Stillwell and Thomas, 2016) such as moving to a different neighbourhood in the same metropolitan area (or region) or moving to other metropolitan areas (national level mobility) (Figure 2-2). Long-distance mobility is constrained by job, kids schooling, social ties and other circumstances and is therefore considerably low in number. The low rate of long-distance mobility is also attributed to the lack of knowledge and information about the new location (Ritchey, 1976), moving cost if properties are being brought or sold (Stillwell and Thomas, 2016). Long-distance movers lose social capital and also experience changes in their travel and commute behaviour depending on the transport accessibility, neighbourhood characteristics of the area they moved to. As a result, the scale of residential mobility is an important factor in terms of investigating residential mobility decision and its consequences on car ownership and travel behaviour.

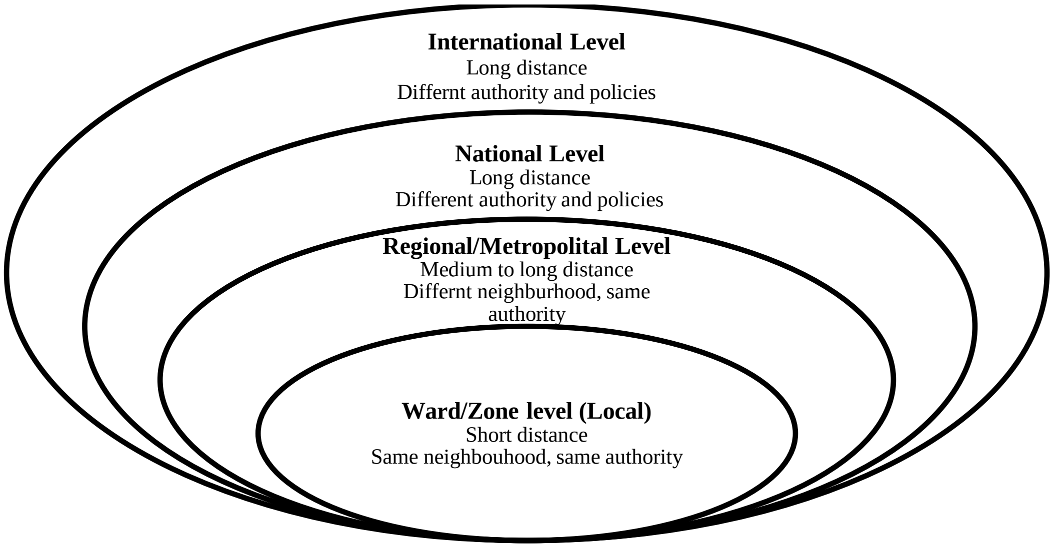


Figure 2-2 Geographical scales of residential mobility (Source: Dieleman et al., 2000)

2.3.3 Residential location choice

Residential location choice is a household decision about their dwelling and neighbourhood for relocation. Households typically consider a dynamically changing set of alternatives before choosing a residential location. Each alternative has unique characteristics in terms of accessibility, housing price, neighbourhood quality, etc. Therefore, to choose the best alternative, households make a trade-off between the attributes of available alternatives. Numerous studies in the literature have investigated household residential location choice behaviour. For example, Schirmer et al. (2013) has examined the impact of location attributes on residential location choice and other studies (Zondag and Pieters, 2005; Chen et al., 2008; Lee et al., 2010) have captured the role of accessibility on residential location choice. Most of the previous studies have however looked at residential ownership decisions (e.g. Zondag and Pieters, 2005; Zhou and Kockelman, 2008) or renting decision (e.g. Lee et al. 2010; Ibraimovic and Hess, 2017) in isolation or both (e.g. Waddell, 2006; Chen et al. 2008). Details of a few important studies are summarized in Table 2-3 and also discussed the limitations of these studies in terms of capturing the differences in the residential location choice behaviour of owners and renters.

Table 2-3 Summary of relevant studies on residential location choice

Authors	Description
Weisbrod et al. (1980)	The study has focused on the trade-off between transportation attributes (e.g. travel time, travel cost, accessibility, etc.) and non-transportation attributes (e.g. dwelling characteristics, neighbourhood quality and demographic factors, etc.) in the decision to move and choice of location. Multinomial and nested logit models have been estimated using Minnesota, metropolitan area survey data. This study has observed a significant level of trade-off between transport and non-transport attributes. For instance, the effect of a small change in housing cost on the likelihood of choosing a residential location is equivalent to the effect of a larger change in travel time. This impact is twice in the case of renters compared to owners. Although, this study estimated owner and renter specific housing cost sensitivities but did not test potential differences and similarities in other parameters. Testing the differences in the preferences of owners and renters to the full set of attributes may give a more comprehensive picture of how they trade-off in their residential location choice attributes.
Zondag and Pieters (2005); Lee et al. (2010)	Both studies have focused on capturing the role of accessibility on residential location choice. Accessibility has been measured as the composite indicators of regional accessibility such as accessibility of employment, shopping, education, transport, etc. Zondag and Pieters (2005) has modelled the owners choice using the nested logit technique whereas Lee et al. (2010) has modelled the renters choice using the multinomial logit technique. Both studies have found an association between accessibility and residential location choice. However, it is difficult to explore from these study outcomes whether the association between accessibility and location choice is different for two different groups (owners and renters) since the data and model specifications used in these two studies were different.
Waddell (2006); Ibraimovic and Hess (2017)	These studies have investigated the preferences for ethnic neighbourhood structure during residential location choice. Ibraimovic and Hess (2017) study used stated preference data for modelling renters behaviour whereas Waddell (2006) modelled both owners and renters behaviour using revealed preference data. Although both studies have found that the individual dislikes the decrease of the co-nation (or increase of minority) in their residential neighbourhood but none of these studies have investigated whether there are any differences in the preference of the owners and renters. Since ownership is a long-term decision, owners are likely to be more sensitive to the ethnic composition in their neighbourhood compare to the renters. If differences in the behaviour of the subgroups existed but did not capture in the model, estimation results may give poor forecasting leading to inappropriate policy analysis.

Table 2-3 Summary of relevant studies on residential location choice (cont.)

Authors	Description
Walker and Li (2007)	Latent class modes have been estimated to investigate the impact of lifestyle on residential location choice. Lifestyle has been captured as a latent preference of living in different urban forms and facilities such as living in suburban, urban or transit-oriented development areas. Sociodemographic characteristics have been considered as predictors of latent preference. From the model outcome, lifestyle preferences have been found to influence residential location choices significantly. Lifestyle preferences are likely to be different for owners and renters and can also change over time. For instance, transit-oriented mixed land development has been found to gain popularity over car-dependent suburban areas in the last few decades (Burda, 2014). This study did not acknowledge these issues.
Chen et al. (2008)	Chen et al. (2008) has investigated the role of past location on the choice of a new residential location using Puget sound panel data. Multinomial logit (MNL) models have been estimated to model the behaviour of both owners and renters although the MNL technique has limitations in terms of capturing the unobserved heterogeneities. This study has found that previous experience plays a significant role in shaping the new choice where only the trade-off between the location attributes has been assessed. However, the trade-off between dwelling level attributes (dwelling cost, size, etc.) of the past and new homes can also play a significant role in residential location choice (Habib and Miller, 2009). Although this study has captured the differences in the preferences of the households having school going chid and households don't have school-going children, the preferences are likely to be differences between long term ownership decision and medium-term renting decision.
Zhou and Kockelman (2008)	This study aimed to elucidate the differences in the residential location choice of households who have worker(s) and who do not have worker(s) while owning. Mixed logit models have been estimated using Austin metropolitan area data. This study has found that workers are highly sensitive to commute time during residential relocation, but this sensitivity can vary depending on the reasons for making a residential move. It can be anticipated that the sensitivities to commute time and other variables can also vary depending on the nature of the decision such as ownership or renting.
Schirmer et al. (2013)	This study has focused on evaluating the impact of neighbourhood characteristics on residential location choice. The neighbourhood is characterised by variables of the built environment, the social environment and points of interest. Dwelling level models have been estimated using a random subset of 50 alternatives from 3890 alternatives. It is unlikely that the households have considered the full choice set (considering a random subset is equivalent to considering the full choice set) for selecting one they perceived to be best. Choice set misspecification might produce bias parameter estimation. Moreover, the multinomial logit models estimated in this study were restricted to capture the variation in the taste of individuals.

2.4 Impact of residential decision on car ownership and travel choice

Household car ownership level and travel decisions are associated with sociodemographic characteristics, life events (e.g. residential relocation, getting married, changing job, etc.) and the neighbourhood characteristics. Residential change, an important life event, can alter household circumstances resulting in a significant level of changes in daily routine and travel behaviour. Several studies in the literature have attempted to model the association between household relocation decision, changes in car ownership and travel behaviour (mode choice behaviour, commute behaviour, driving behaviour, vehicle miles travelled, etc.). One group of studies has investigated the changes in household car ownership and travel behaviour due to changes in their circumstances after residential relocation. These studies only have considered households that have moved in the recent past. Another group of studies has captured car ownership and changes in the travel behaviour of all households (irrespective of they have moved or have not moved recently). In the latter case, residential mobility is considered as an independent variable in the model to investigate its impact on the behaviour of car ownership and travel behaviour changes. Most importantly most of these studies only have investigated the binary choice of changed or did not change. Few studies captured the directions of transition, for example, changes from active travel to non-active travel and vice versa (Clark et al., 2016b) or transition from 0 car to 1 car, 1 car to 2 cars, etc. (Clark et al., 2016a) as a binary choice. The limitations of existing studies in terms of capturing the directions of changes in car ownership level and travel behaviour, the role of geographical scale of relocation on car ownership level and travel changes, dynamics in household behaviour, etc are presented in Table 2-4. However, the factors that influence the changes in car ownership level and travel behaviour are further discussed in the following sections as well.

Table 2-4 Summary of relevant studies on car ownership and travel behaviour changes

Author	Description
Krizek (2003)	This study has investigated the effect of neighbourhood change after residential relocation on household travel behaviour such as vehicle miles travelled, number of trips, etc. However, the neighbourhood change can have a significant impact on household car ownership level which, in turn, may influence travel mode choice, travel distance and travel length. This study did not consider or acknowledged car ownership change behaviour due to residential relocation and its influence on travel behaviour. Although this study has used longitudinal data, estimation technique (regression analysis) did not allow to capture the dynamics in the life course and correlations in the repeated choices overtime.
Prillwitz et al. (2006); Prillwitz et al. (2007)	These studies have emphasised on how life events (e.g. residential mobility) influences car ownership level (Prillwitz et al. 2006) and changes in commute distance (Prillwitz et al. 2007) using German socioeconomic panel survey data. The modelling techniques used in these studies (probit model and regression analysis) were less flexible to capture the benefit of panel data. In addition, levels of car ownership change (e.g. zero to one car, one to two cars, etc.) and directions of travel distance change (increasing or decreasing.) may involve different decision-making processes. However, these studies did not focus on these issues. For instance, Prillwitz et al. (2006) investigated whether the households are gaining car(s) or not although the behaviour of gaining the first car is unlikely to be same as the behaviour of gaining the second or more cars.
Cao et al. (2007); Aditjandra et al. (2012)	Both studies have looked at a similar issue (effect of neighbourhood change on car ownership and travel behaviour) in the context of two different geographical locations: Northern Californian (Cao et al., 2007) and Tyne and Wear, the UK, (Aditjandra et al. 2012). The structural equation modelling technique has been used to investigate the research questions using quasi longitudinal data. A strong association of the neighbourhood change with changes in car ownership level and travel behaviour have been observed in both cases. These studies have investigated the role of neighbourhood change on car ownership level and travel behaviour after relocating within a city or region. However, the scale of neighbourhood change when relocating in another region is likely to be largely different when relocating within a region (Milakis et al., 2015). Because of the land use pattern, transport and other facilities, policies can differ significantly from city to city. Ignoring this issue can under or overestimate the neighbourhood effect on car ownership level and travel behaviour. In addition, policymakers may want to know the impact of different types of neighbourhoods on the specific directions of behavioural change (e.g. gaining of first car, gaining of additional car(s)). However, these studies have not investigated or acknowledged this issue.

Table 2-4 Summary of relevant studies on car ownership and travel behaviour changes (cont.)

Author	Description
Oakil (2013)	Oakil (2013) has estimated mixed logit models to investigate the changes in car ownership level and travel mode using retrospective survey data from Utrecht, Netherlands. Although this study has found several useful insights about the role of life events and demographics on changes in household car ownership level and travel behaviour but rather limited in terms of capturing the direction of changes. For example, binary logit models are estimated to capture the behaviour of switching to the car or not switching. However, switching from public transport to car may involve different decision-making process compared to the switching from active travel to car. In addition, the policymaker might be interested to know about the attributes of each direction of switching because switching to car from active travel may have a different impact on the transport network compare to the switching to the car from public transport.
Clark et al. (2014); Clark et al. (2016a)	Changes in household car ownership level over the years has been investigated in these studies using the British household panel survey data. Clark et al. (2014) has investigated the factors driving the changes in car ownership level whether Clark et al. (2016a) investigated the direction of changes. Although the second study (Clark et al., 2016a) has looked at the differences in the different directions of behavioural changes, modelling technique used here (logit model) was limited in terms of capturing the dynamics in the life course and correlations in the repeated choices. In addition, geographical scales of residential relocation might have a different level of influence on the directions of behavioural change which need to be investigated.
Clark et al. (2016b)	This paper addressed the association between life events and commute mode changing behaviour using British household panel survey data. Random-effects binary logit models have been estimated for modelling the behaviour of switching to the car and switching from the car. The specific directions of switching (e.g. switching from car to bus/rail/cycling/walking) that may have policy importance have not been investigated. This study only used two out of eighteen waves of the panel data. Although the commute mode switching behaviour has been found to be associated with household life events, use of the longer or full panel data might give more robust estimation.
Klinger and Lanzendorf (2016); Lin et al. (2018)	These studies have focused on the changes in car ownership (Lin et al., 2018) and travel behaviour (Klinger and Lanzendorf, 2016) after residential relocation. The role of residential mobility scales on car ownership change and travel behaviour has not been investigated. However, the car ownership change and travel behaviour of the households who have moved in the same neighbourhood (or city) are more likely to be different from the behaviour of the households who have moved in a different neighbourhood (or city). In addition, cross-section data used in Klinger and Lanzendorf (2016) study and only two waves of longitudinal data used in the Lin et al. (2018) study may not be suitable to model these time-dependent household behaviours.

2.4.1 Changes in car ownership

Car ownership is an important determinant of household travel behaviour. Although income is an indicator of household car ownership level (Dargay, 2001; Van Acker and Witlox, 2010), acquiring a driving licence (Van Acker and Witlox, 2010; Clark et al., 2016a), neighbourhood characteristics (Clark et al., 2016a), life events (e.g. moving house, changing job, having baby, etc.) (Clark et al., 2014) also have a close association with the likelihood of changing car ownership level. Dargay and Hanly (2007) revealed that the probability of changing car ownership levels of households experiencing a life event recently (e.g. residential relocation, changes in household composition) is higher compared to households who do not experience it. Lin et al. (2018) found that non-car owning households having well-educated and employed member(s), high level of income and large family size are more likely to acquire a car after relocating. Moving house has been observed to have a strong to moderate influence on car ownership level changes in several other studies (Prillwitz et al., 2006; Yamamoto, 2008; Rashidi et al., 2011; Oakil et al., 2014; Zhang et al., 2014). However, Clark et al. (2016a) observed a weak association of car ownership change with residential mobility and change in urban form. They found a strong connection of car ownership level changes with the life events, employment and acquiring a driving licence, etc. Changes in the built and social environment (in terms of safety, neighbourhood cohesion) were also found to have a strong influence on gaining and losing cars. In case of Beijing, China, Lin et al. (2018) found that moving out from the city centre is associated with a higher chance of car disposal and moving in the city centre from a suburban area increases the chances of car acquisition. This result is intuitive because wealthy people in China live in the city centre and having a high level of car ownership. However, other studies in the context of European and American cities claimed that households who live in the low-density suburban area are more car-dependent (Alexander and Tomalty, 2002; Naess, 2009) whereas households who live in the high-density urban area are less car-dependent due to having better transport access (Masnavi, 2000; Ewing and Cervero, 2010). The above discussion confirms the strong association between life events, urban form and car ownership changes.

2.4.2 Changes in travel behaviour

2.4.2.1 Travel model transition

Accessibility levels of household (e.g. transport, shopping, social, etc.) are likely to be affected after moving in a different neighbourhood and changes in neighbourhood accessibilities incur changes in the travel behaviour (Krizek, 2003). The daily commute mode of an individual is likely to be affected by the changes of commute trip length due to moving home or changing job (Oakil et al., 2011; Clark et al., 2016b). An increase in commute distance has been found to increase the propensity of switching from non-car (transit and active travel) to car use while high-quality public transport links to employment centres have been found to increase the likelihood of switching from car to non-car use (Clark et al., 2016b). Importantly, mixed land use has been found to encourage switching to active travel (walking and cycling). Klinger and Lanzendorf (2016) found that the travel mode choice behaviour after an intercity relocation is determined by urban mobility cultures, spatial characteristics of the neighbourhood, household preferences and underlying self-selection processes. Few studies found that moving to a neighbourhood with a higher level of public transport accessibility reduces car use (Kockelman, 1997; Cao et al., 2007; Aditjandra et al., 2012). However, few other studies found a positive effect on car use with an increase in regional and neighbourhood accessibility (Rajamani et al., 2003; Lin et al., 2018). Different specifications of accessibility term in different studies might lead to different outcomes. Contradictory results have been observed in terms of the influence of social networks on travel mode use. Aditjandra et al. (2012) found that an increase in the social network decreases the likelihood of using car whereas Lin et al. (2018) found that an increase in the social network decreases the use of non-motorized travel mode. However, an improvement in the social environment (in terms of safety and neighbourhood cohesion) helps to drop car dependency (Lin et al., 2018). The impact of household sociodemographic characteristics on mode choice behaviour was found significant in several studies (Klinger and Lanzendorf, 2016; Lin et al., 2018). For example, elderly people are found more inclined to non-motorized travel on the other hand male, highly educated people, newly employed persons, car owners and wealthy people tend to have a higher tendency of car use after the residential move (Klinger and Lanzendorf, 2016).

2.4.2.2 Changes in commute distance

The length of the commute trip of people is linked with where they live and where they work. Therefore, a change in home location alters the commute distance but the direction of alternation depends on the purpose of moving. Moving home is aimed for saving commute distance in many cases, however, households may also accept longer commute distance after relocation to meet other requirements (get a bigger house, find a cheaper house, proximity to better schools, etc.). Prillwitz et al. (2007) found that moving from core (urban) to noncore (suburban) area, job change, increase in car ownership and move to the single-family house are associated with an increase in individual commute distance. Sometimes, easy access to public transport increases commute distance. Krizek (2000) observed that moving from a medium to low LADUF (Low Auto-Dependent Urban Form) neighbourhood increases the trip and tour distance of household members.

2.5 Modelling issues

Modelling residential choice has several methodological and estimation challenges. In recent years, significant improvements have been achieved in the context of choice set formation (Zolfaghari, 2013; Bhat, 2015), sampling of alternatives (Guevara, 2010), level of aggregation of alternatives (Zhou and Kockelman, 2008; Zolfaghari, 2013), treatment of complex correlation structures (Bhat and Guo, 2004; Sener et al., 2011), and endogeneity correction (Guevara, 2010), etc. These issues can be critical in case of modelling car ownership choice and travel mode choice. Given the relevance of this research, the literature on the level of aggregation of alternatives, choice set construction, sampling of alternatives and endogeneity correction are explained in more detail in the following sections.

2.5.1 Level of aggregation

In case of residential location choice modelling, the alternatives could be zones or parcels or dwellings. Typically, traffic analysis zones or census output areas or electoral ward areas are considered as zones and individual land areas or buildings are considered as parcels. An example of parcels and zones are given in Figure 2-3. The assigning of alternatives in the model whether zone, parcel or dwelling level depends on the level of spatial granularity of the alternatives in the datasets. Due to the unavailability of dwelling level information or dwelling supply data, most of the previous studies have considered zones as location alternatives (Zondag and Pieters, 2005; Walker and Li, 2007; Chen et al., 2008; Sener et al., 2011). Zone level aggregated information of dwellings (average dwelling price, average dwelling size, etc.), land use characteristics (land use density, transport accessibility, etc.) and demographics (average household income, household size, etc.) are used as independent variables and each zone is considered as location alternative. In the literature, several advanced modelling techniques have been proposed to capture the spatial issues in zone level residential location choice models (Bhat and Guo, 2004; Pinjari et al., 2011). Despite of using more advanced estimation techniques, zone level models have limitations to capture the household sensitivities towards dwelling attributes.

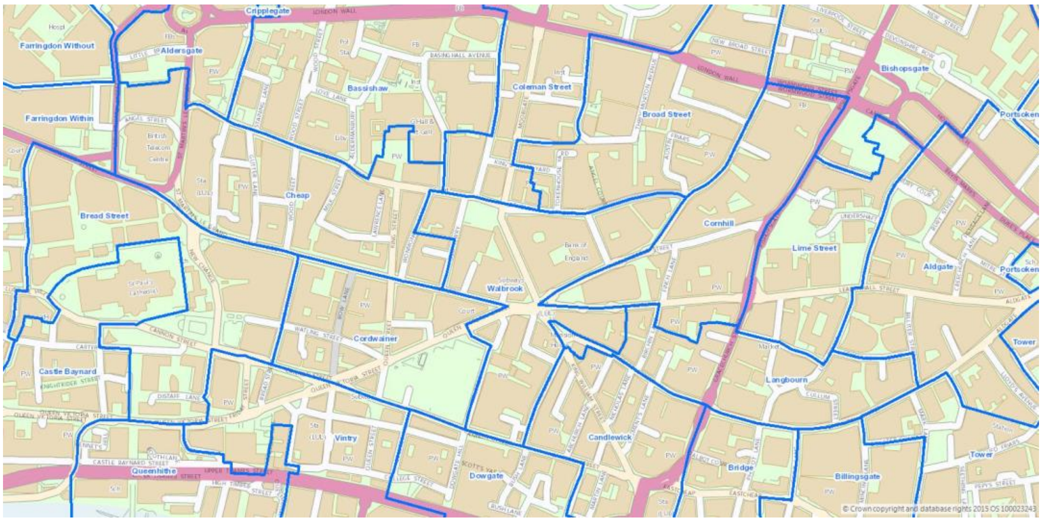


Figure 2-3 An example of zone (blue lines) and parcel (light brown blocks inside zones) (Source: Zolfaghari, 2013)

Due to the availability of high-resolution data (parcel level), several studies have estimated parcel-level residential location choice models (Waddell, 2006; Zhou and Kockelman, 2008; Lee et al., 2010). This approach is better than the zone level approach because it can capture the variabilities in the neighborhood characteristics and other facilities within a zone. For example, an alternative having an unrestricted parking facility is likely to be more attractive compared to some other alternatives within the same zone having restricted parking access. However, Parcel-level models also have limitations to capture the variation of different dwellings within a parcel/building. For example, the basement floor of a multi-story building might be less attractive than other floors and parcel level models cannot capture this dissimilarity.

The dwelling level approach is the finest level of disaggregation in residential location choice modelling. This approach can capture the variations at the dwelling level. In this approach, alternative dwelling information such as dwelling size, dwelling cost is considered as an independent variable whereas zone and parcel level approaches use aggregate level dwelling information. Several attempts have been made in the literature for estimating zone-based dwelling level models (Habib and Kockelman, 2008; Zolfaghari, 2013). However, estimation of the dwelling level model is challenging due to a large number of alternatives, lack of dwelling supply data in many cases and complex correlation structure.

2.5.2 Choice set generation

Defining the individual specific choice set in a discrete choice model is critical and substantially influences the model outcomes (Swait, 2001; Bell, 2007). Many alternatives in the choice set and unavailability of information about individual choice set in some cases aggravate the level of complexity. For instance, the total number of alternatives in the zone level residential location choice model ranges from hundreds to thousands which could be hundreds of thousands in case of dwelling level model. Under choice set construction, the analyst may consider a reduced set but the reason to do it is behavioural. In this context, the analyst aims to ignore the unrealistic alternatives that decision makers did not consider during the decision process. Moreover, it is unrealistic that an individual considers a very large number of alternatives for choosing one or have perfect knowledge about all the alternatives (a basic assumption of discrete choice analysis).

Different techniques have been used in literature to capture behavioural choice sets. The two-stage probabilistic approach proposed by Manski (1977) is a classical solution to model individual choice sets. This method requires the estimation of probabilities of all possible choice sets in the first stage and the conditional probabilities of alternatives across all choice sets in the second stage. Both stages are estimated simultaneously. The number of possible choice sets explodes with the number of alternatives in the universal choice set. For *J* alternatives, the number of possible choice sets is 2^J-1. Therefore, this method is computationally infeasible for a medium to large choice set (e.g. residential location choice). The unconditional probability of choosing an alternative by a decision maker in the Manski method is the product of the conditional probability of the alternative (given the choice set) and the probability of the choice set. The probability can be presented as follows:

$$P_{ni}(C) = \sum_{C_n \in C} P_n(i/C_n) \times P(C_n), \tag{2-1}$$

where $P_{ni}(C)$ is the unconditional probability of choosing alternative i by individual n, $P_n(i/C_n)$ is the conditional probability of choosing alternative i from the choice set C_n $(C_n \in C)$ and $P(C_n)$ is the probability of the choice set being C_n. Other probabilistic approaches proposed in the literature as alternatives to the Manski method (e.g. Swait and Ben-Akiva, 1987; Swait, 2001; Kaplan, Bekhor and Shiftan 2011; Zolfaghari 2013; Bhat, 2015; etc.) also have computational complexity for a large choice set.

Deterministic constraint-based approaches have also been used in the literature to model choice sets in the context of residential location choice (Zolfaghari, 2013). These methods have assumed that households use non-compensatory decision rules for the screening of alternatives based on some behavioural constraints. Alternatives are removed from the individual choice set when certain attributes of an alternative exceed exogenous thresholds. These exogenous thresholds can be either imposed deterministically based on insights from the data (Farooq and Miller, 2012) or can be computed (Zolfaghari, 2013). Importance sampling techniques have also been used in the context of residential location choice modelling (e.g. Rashidi et. al, 2012; Zolfaghari, 2013; etc.). These techniques are similar to the deterministic constraint-based approaches but allow proportional sampling of alternatives from within and outside the threshold zone. For example, Farooq and Miller (2012) applied importance sampling to construct individual choice sets for residential location choice modelling by taking 75% of alternatives within 15 km of the past location and the remaining 25% from outside the threshold. Since these techniques are based on assumptions made by the analyst, there is a high risk of choice set misspecification and consequently, poor model fit and biased parameter estimation.

Heuristic-based single stage semi-compensatory approaches become popular for estimation simplicity and behaviour persuasiveness. Cascetta and Papola (2001) proposed Availability Perception Random Utility (IAPRU) model having a higher level of efficiency in terms of avoiding the challenges in the probabilistic approach (Manski, 1977) and the risk of elimination by aspect approaches. Membership or availability of alternatives in the individual choice set is simulated implicitly in the utility function based on random constraint on attributes. This method has further been improved by Cascetta and Papola (2009) (called dominant rule-based random utility model), Martínez et al. (2009) (called constrained multinomial logit model), Paleti (2015) (called r^{th} order constrained multinomial logit model). The semi-compensatory techniques are described in detail in section 6.2 of chapter 6.

2.5.3 Sampling of alternatives

If the number of possible alternatives to choose one from them is too many and the decision makers are aware of all the alternatives, ideally, the analyst should include all alternatives in the individual choice set to estimate the model parameters. This approach is behaviourally sound but computationally intractable. Therefore, the analyst can use a reduced set of alternatives that represents the full choice set. The reduced choice set consists of the chosen alternative and a subset from all other available alternatives. This process is commonly known as sampling of alternatives. The purpose of considering reduced set using an appropriate sampling technique is computational whereas considering the reduced set as means of capturing the true choice set (discussed in the previous section 2.5.2) is completely behavioural.

The process of sampling the alternatives for the reduced choice set is non-trivial. Firstly, since the increase in the size of choice set is likely to increase the quality of the estimates but also increases the computational burden, it is difficult for an analyst to decide how many alternatives to include in the choice set. Secondly, the analyst also needs to find a process for determining which alternatives to include in the choice set. A random selection process has been used in literature widely (Bhat and Guo, 2007; Habib and Miller, 2009; Lee and Waddell, 2010; Guevara, 2010). Random sampling of alternatives leads to increase error in parameter estimation. McFadden (1973) proposed a correction term for a consistent estimation of the Multinomial Logit Model (MNL) with a reduced choice set. If the analyst decides to consider a subset of alternatives, D_n,, for an individual decision maker n, the probability of choosing subset D_n given that i is chosen alternative is $\pi_n(D_n/i)$. Using the Bayes theorem, the conditional probability of individual n choosing alternative i from the subset D_n can be expressed as follows (McFadden, 1973)

$$P_n(i|D_n) = \frac{P_n(i)\pi_n(D_n|i)}{\sum_{j\in D_n} P_n(j)\pi_n(D_n|j)} \tag{2-2}$$

$P_n(i)$ is the unconditional probability of choosing alternative i. Replacing the logit probability function of $P_n(i)$, the conditional probability can be simplified as

$$P_n(i|D_n) = \frac{e^{v_{in}+ln\pi_n(D_n|i)}}{\sum_{j\in D_n} e^{v_{jn}+ln\pi_n(D_n|j)}} \quad (2\text{-}3)$$

where, $ln\pi_n(D_n|i)$ is correction term for sampling of alternatives. If $\pi_n(D_n|i)$ satisfies the uniform conditioning property, the correction term will cancel out and the probability of individual *n* choosing alternative *i* will be collapsed to the standard MNL format. Thus, a random sampling of alternatives with a uniform size of the individual choice set allows consistent parameter estimation in the standard multinomial logit model. McFadden (1973) work has been extended for correcting the sampling bias in the nested logit model (Guevara, 2010) and the mixed logit model (Guevara and Ben-Akiva, 2013).

2.5.4 Endogeneity

An econometric model suffers from an endogeneity problem if the error component is correlated with the deterministic component. Although simultaneous determination and specification or measurement errors can result endogeneity bias in the model, omission of attributes that are associated with the explanatory variables in the deterministic part is the most common and significant cause of endogeneity (Guevara, 2010). This problem is nearly unavoidable in many empirical cases. In case of residential location choice, consider two dwelling alternatives that are same in their attributes level except for the year of construction. One is relatively newer than the other and demands a higher price. If the age of the property is not considered in the model, the estimated result might show higher cost sensitivity and the model will suffer from price endogeneity.

In literature, household one behaviour has been used to explain another behaviour such as residential mobility behaviour has been used as an independent variable for explaining car ownership change behaviour (Clark et al., 2016a), and travel mode switching behaviour (Clark et al., 2016b), on contrary, car ownership has been used to explain residential mobility behaviour (Hensher and Taylor, 1983). In those cases, the behaviours which have been considered as independent variables can be endogenous if these variables are correlated with the unobserved utility of the dependent variables.

Several methods have been proposed in the literature to deal with the endogenous variable for consistent estimation of the model parameters. The available methods for endogeneity corrections are the applying Proxys (PR), two steps Control-Function (CF) method, Full Information Maximum-Likelihood (FIML); the Multiple Indicator Solution (MIS), and Latent-Variables (LV) approach and BLP method (Berry, Levinsohn and Pakes method). The appropriateness of these methods depends on the problem that is being analysed. The methods differ considerably in their underlying assumptions, the difficulty of finding appropriate auxiliary variables, estimation tools and computational burden. PR method is the easiest technique but inappropriate in many cases. In the RP method, a proxy variable is used to account for the omitted variable, but the proxy must meet the criteria of independence from error and other explanatory variables (Wooldridge, 2010). CF method is suitable for correcting endogeneity at the level of each alternative. It requires to find an appropriate instrument or auxiliary variable which is correlated with the omitted variable and uncorrelated with other explanatory variables. In many cases, it is very difficult to find a suitable instrument and need to compromise the estimation efficiency since CF is a two-step process. CF is first proposed by Heckman (1978), improved by Rivers and Vuong (1988) for binary Probit, further improved by Petrin and Train (2002) for mixed logit. This method has been applied successfully for addressing endogeneity in many discrete-choice models (Ferreira, 2010; Guevara and Ben-Akiva, 2006, 2012). Multiple Indicator Solution (MIS) method does not require instruments for endogeneity correction. The MIS relies on a couple of indicators that depend on the latent variable that causes endogeneity but is not correlated with other attributes. This method is proposed by Wooldridge (2010) for linear models and improved by Guevara and Polanco (2015) for discrete choice model. This method gains estimation efficiency over the CF method.

Another method of endogeneity correction is the Maximum Likelihood (ML) approach which is suitable when the structural equation of the latent variable is linear, and the endogenous variable is discrete. The estimation using the ML method can be very challenging because the dimensionality of the integral increases as the number of alternatives increases (Cherchi and Guevara, 2012). The Latent-Variables (LV) approach developed by Walker and Ben-Akiva (2002) is similar to the ML but the latent variable in the LV approach could be either discrete or continuous. The methods discussed above are applicable if the endogeneity occurs at a disaggregated level (each observation). If endogeneity occurs at the level of groups of observations, the BLP method is more appropriate for endogeneity correction (Berry et al., 1995).

2.5.5 Behavioural dynamics

Household decisions that aimed to capture in this study (residential change, changes in car ownership level and travel mode switching behaviour) may have dynamic effects. The main source of dynamics in these decision components can be the connection of the current choice with the condition and choice in the previous time period. The following section presents the potential dynamics in household behaviour.

Changes in household circumstances, lifestyle and government intervention in the recent past can influence their new choices. For instance, residential relocation and job switching have been found to influence the choice of commute mode in the following year (e.g. Clark et al., 2016b). Elapsed time since the most recent choice (duration of the current choice) can also play a significant role in the new choice outcomes (e.g. Habib 2009; Clark and Lisowski 2017. For instance, households who have changed their residential location recently may be less likely to change again. It may be noted that several studies in the literature highlighted the influence of variety-seeking behaviour - where decision makers are more likely to accept new options (Rieser-Schüssler and Axhausen, 2012; Song et al, 2018). Finally, previous experience or choice can also affect the new choice (Oakil 2013; Fatmi and Habib, 2017).

The most widely used econometric techniques for modelling dynamic behaviours are 1. Discrete choice modelling and 2. Hazard based duration modelling (Ghasri et al., 2018). The discrete choice technique answers the question of what decision is made in a certain time interval, whereas the hazard model answers the question when the decision is taken place. The relative merits and drawbacks of these two techniques are further discussed in the following subsections.

2.5.5.1. Discrete choice modelling

Discrete choice model has been used in literature for modelling household behaviours that have dynamic effect such as residential mobility behaviour (e.g. Clark et al., 1986; Clark and Huang, 2003; McCulloch, 2010; Coulter and Scott, 2015), vehicle transaction (e.g. Bhat and Pulugurta, 1998; Bhat et al., 2009; Oakil, 2013; Clark et al., 2014) and travel mode switching behaviour (e.g. Oakil, 2013; Clark et al., 2016b; Klinger and Lanzendorf, 2016), etc. In most of the cases, time varying covariates have been used to capture the effects of changes in household conditions or circumstances on the new choices. Lagged dependent variables have also been used for capturing the impact of past experience or choice (especially the most recent one) on the new choice outcome (e.g. Davies and Pickles, 1985; McHugh, Gober and Reid, 1990; Habib, 2009; Clark and Lisowski, 2017). However, the use of lagged variables has been criticized for the risk of endogeneity due to correlation between the lagged variable and other unobserved effects (e.g. Judson and Owen, 1999; Bun and Sarafidis, 2015). Since the lagged variables are likely to be driven by the same underlying factors of the choice outcome, they are more likely to be endogenous. Discrete choice models are found to be straightforward for testing both time varying covariates and lagged dependent variables for capturing dynamic effects in the behaviour.

Although the duration in the current choice can have a significant role on the new choice outcomes, the suitability of discrete choice modelling for capturing the time dependency of behaviour (or duration dynamics) is affected by the discretisation of the data, modelling choices at fixed discrete time period (e.g. a year).

2.5.5.2 Hazard-based duration modelling

The hazard model is a more appropriate technique for capturing the time dependency of the choice behaviour (or duration dynamics). This acknowledges duration dynamics since the duration in the current choice is captured as a stochastic process where the likelihood of changing state depends on the length of time stayed from the start of the event. Although the hazard-based modelling has potentials for capturing the dynamics in the behaviour, the application of this technique is quite slim in transport related literature (Ghasri et al., 2018). One of the reasons can be the challenges of hazard-based modelling when covariates are time dependent. The relationship between the product and variable over time can lead to error unless the interrelationships are well understood (Fisher and Lin, 1999). In some cases, interpretations of the models no longer hold when time dependent covariates are used. Possible ways of incorporating these variables in the hazard-based model are to re-organise the data or apply special treatments in the modelling, but these treatments are very challenging in many cases (Jenkins, 2005). However, hazard based modelling is more straightforward if all explanatory variables are fixed.

2.6 Concluding remarks

The integrated urban model is the most advanced tool to model two interconnected but different areas: land development and transportation. However, the behavioural dynamics and interdependencies in different components of these two areas are poorly addressed in most of the current studies. This chapter reviews the state-of-the-art knowledge in this context and several gaps are identified. Although, existing studies in the literature have captured a long list of behavioural aspects in modelling residential mobility decision such as the connection between the life events and residential mobility, the role of social capital on mobility decision, influence of neighbourhood attributes on residential mobility decision; the geographical scale of residential mobility has not been explored yet. The literature survey in this chapter also confirms strong connections between residential mobility, car ownership and travel decision. However, none of the existing studies have focused on the role of geographical scale of residential mobility on car ownership and travel changes, though, the scales of residential mobility are likely to have varying impacts on the corresponding changes. Moreover, existing studies also have limitations in terms of capturing the full range of transition behaviour (e.g. switching from zero car to one car, single car to two cars in case of car ownership changes) in a single model. The review of the literature on residential location choice also identifies gaps in terms of capturing behavioural aspects and modelling issues. Many behavioural issues in modelling residential location choice have already been captured in the literature as discussed in this chapter. However behavioural differences between two major housing markets and time varying nature of their behaviour in terms of choosing residential location remain as gaps. Choice set formation is also observed as a major challenge in this context. From the detailed review of the literature, no ideal solution is observed for capturing behavioural choice set for disaggregate level residential location choice modelling.

Chapter 3

Modelling residential mobility decision and its geographical scale

3.1 Introduction

Residential decision (a component of the research framework presented in chapter 1, Figure 1-3) consists of two basic layers: residential mobility decision and residential location choice. Although a common set of parameters can influence both of the decisions (Lee and Waddell, 2010), the residential mobility decision is found to be more substantially affected by the households demographic characteristics and life events (Van der Vlist et al., 2002; Clark and Huang, 2003; Coulter and Scott, 2015) while the location choice is mostly associated with land use attributes, dwelling characteristics and transport accessibility (Bhat and Guo, 2007; Haque et al., 2018). These two layers of residential decisions can be interdependent, therefore, the underlying considerations in one choice can potentially influence the other choice. For example, if a household is planning to move home for saving commute distance, the residential location alternatives close to the workplace will have a higher choice probability. For capturing the connection between these two layers of residential decision, joint estimation technique has been considered in a few past studies (e.g. Lee and Waddell, 2010). However, in many studies in literature, these decisions have been modelled sequentially where the residential location choice follows the residential mobility decision (e.g. Brown and Moore, 971; Habib, 2009). In these studies, it is assumed that households made the decisions of residential mobility and choice of location sequentially. This assumption is behaviourally reasonable because it is unlikely that households become active in the housing market to find alternative locations or dwellings unless they are considering to move. Moreover, this simplified approach (sequential) is empirically plausible for piece-wise development of overall empirical relocation choice models capturing the critical issues in the individual components (Cadwallader, 1992; Habib, 2009). This study also aims to model the residential mobility and choice of location sequentially. Moreover, due to the data limitation as mentioned in section 1.3 of Chapter 1, joint estimation of residential mobility behaviour and location choice behaviour was not feasible in this study to capture the reciprocal relation between the decision components (discussed in detail in the conclusion chapter). This chapter focuses on residential mobility behaviour (Figure 3-1 highlights the residential mobility component in the research framework) whereas Chapter 5 captures the residential location choice behaviour. Although this adopted approach is limited in terms of capturing the association between these two layers of residential decision, these independent models may provide several behavioural insights which can be used for policy formulation. The potential policy implications of this study outcomes are discussed at the end of this chapter.

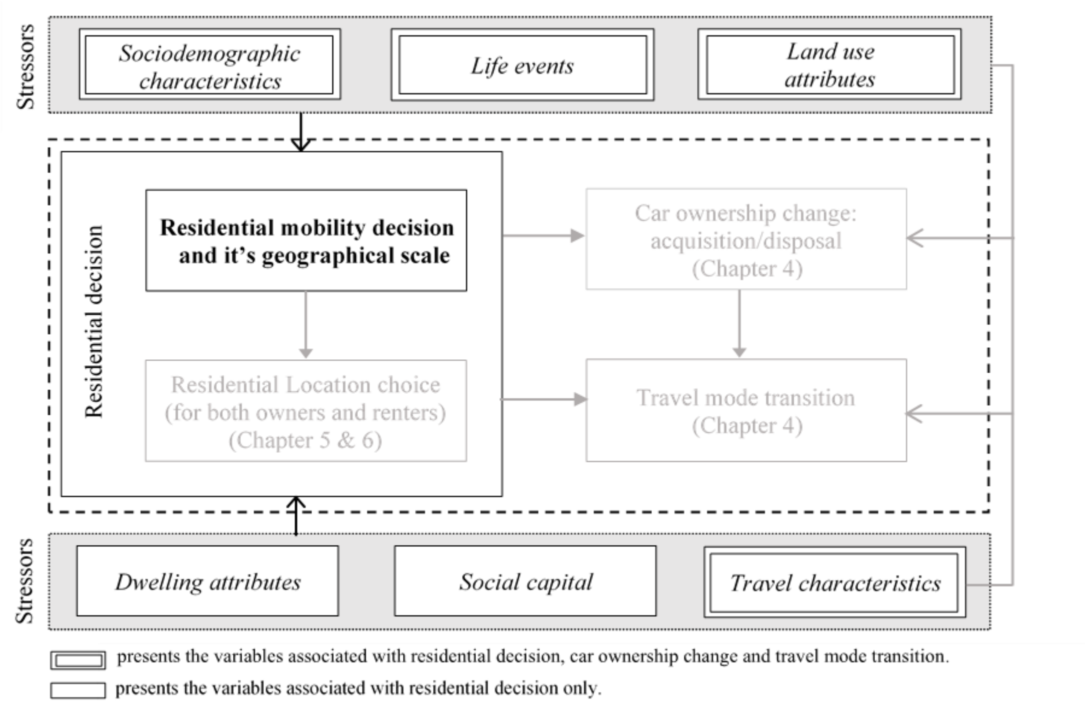


Figure 3-1 Component of the modelling framework that captured in this chapter (highlighted)

In the literature, many studies aim to capture the different aspects of residential mobility decision such as connections between life-course events and residential mobility (Clark and Huang, 2003), role of housing policies on residential mobility (Sánchez and Andrews, 2011), gender role on residential mobility decision (McCulloch, 2010), influence of social network on residential mobility decision (David et al., 2010), residential mobility and travel behaviour (Krizek, 2003), etc. However, there is a research gap in terms of exploring the differences in residential mobility behaviour in different geographical scales. This is important because the process of residential mobility is influenced by a set of circumstances at different geographical scales varying from more general to the most specific context (Dieleman et al., 2000). International mobility is considered as mobility at the highest geographical scale whereas mobility within the same neighbourhood is considered as the most local case although Dieleman et al. (2000) considered metropolitan areas as the most specific scale of residential mobility. Therefore, the residential mobility within a country can be embedded into three geographical scales: local level (within the same ward or zone or neighbourhood), metropolitan/regional level (within a metropolitan area/region) and national level (inter metropolitan areas/regions).

The decisions about moving home and its geographical scale are mostly determined by the social, economic and personal circumstances of households. For example, moving for a more spacious home to accommodate a new member in the household is likely to happen within the same geographical area, which has a little effect on social life, commute and travel behaviour in general. On the other hand, job switching to another metropolitan area has a strong effect on all the above. Housing market characteristics of the geographical areas such as housing cost, tenure composition, and housing policies can also influence household residential mobility decision. Due to the higher cost of housing in the greater London area (GLA) compared to the other UK cities, the rate of migration to the GLA from other UK cities is likely to be low. For example, the number of people moved in and moved out of the GLA during 2004 were 150K and 260K, respectively (Travers et al., 2007). A wide body of literature also demonstrated the connection between housing preferences and the opportunities available in the local housing market (Floor et al., 1996; Molin et al., 1996; Mulder, 1996). If the opportunities are limited, the chance of moving home is likely to be low (Dieleman et al., 2000). In summary, mobility decisions are driven by individual needs, opportunities available and characteristics of the housing markets, therefore, the nature of the mobility decisions is likely to be different in different geographical scales.

Based on the discussion in the preceding section, this chapter aims to capture the following research objectives

- To investigate the factors driving the household residential mobility decision.
- To investigate the potential differences in different geographical scales of residential mobility.

It may be noted that residential mobility is a rare event that may affect the quality of results obtained from cross-sectional data. This has prompted to use a longitudinal dataset (18 waves of the British Household Panel Survey) for model estimation. The long panel helps to examine the choices made by the same households over a span of time. The econometric technique applied in this study allows to quantify the differences between residential mobility in different geographical scales. In addition, the panel nature of the data used in this study facilities to capture the correlation of the choices over time and the impact of the dynamic state of the household on their preferences.

The remainder of this chapter is organised by presenting the details of the data, model structure, choice set construction, results and conclusions. The analyses and modelling works presented in this chapter are directly related to Chapter 4 where the effect of the geographical scale of residential mobility on car-ownership and mode choice has been investigated.

3.2 Data

3.2.1 Data description

The British Household Panel Survey (BHPS) dataset used in this study covers 18 waves from 1991 to 2008. This survey was conducted by the Economic and Social Research Council UK Longitudinal Studies Centre (ULSC), together with the Institute for Social and Economic Research (ISER) at the University of Essex. BHPS is a household-based survey that captured every adult member of the sampled households. The survey was initially designed for understanding social and economic changes at the individual and household level in the United Kingdom. However, BHPS contains information on household residential mobility behaviour, travel characteristics and socio-demographic characteristics. The first wave included 5,511 households and 13,840 individuals from the United Kingdom. A considerable number of households dropped out across the waves and new respondents were added in the subsequent waves. A multi-stage stratified sampling technique was used for sample selection and data was collected through face to face interviews, telephone interviews and self-completed responses. Dataset is released as SPSS, SAS and STATA files and available through the UK Data Service. Although the BHPS data has rich panel information about mobility behaviour, demographics and attitudes, because of the discontinuity of some of the variables across the panels, it was not possible to use all variables. It may be noted that preparing the data for the analysis was a non-trivial task because of the data organization as elaborated in the next section. The effort of data preparation reflects the difficulty of working with such complex data and a potential reason why a small number of past applications have used for developing the choice model.

3.2.2 Data preparation

Data for each wave of the BHPS was recorded in separate files and multiple data files were used to store different types of information within a wave. For example, household-level information, self-reported individual member information and individual-level information of the other household members were recorded in separate files. To build the connections between the information in the files within the waves and across the waves, two types of identifiers were used in the BHPS dataset. (1) Wave specific household identity number (called wHID) and person identity number (called wPNO): these identifiers (that changes across the waves) are provided to find links between the information available in the separate files within a wave. (2) Cross wave personal identifiers (PID): this is provided for linking the information across the waves which are unique for individuals for the full panel. Combining the data from several files considering the different types of identifiers provided was challenging.

Wave specific data files that consist of household level information (wHHRESP) and information of the respondents (wINDRESP) are merged using wave specific identifiers (wHID). Then the merged data files of each wave is combined with the data file containing the information of all individuals in the households (wINDALL) in the corresponding wave using person specific identifiers (wPNO). Finally, the wave specific combined data files that contain all three types of data (wHHRESP, wINDRESP and wINDALL) in each are bound (row bind) together. The final data set thus consists of the information of all households and individuals who attended in all eighteen waves, left the survey, returned into the survey and joined throughout the survey. To investigate who participated in all eighteen waves, cross-wave identifiers (PID) are used. The PIDs that are observed in all waves are considered for the balanced panel. R is used for processing the data. An example code prepared for data processing is given in Appendix A.

3.2.3 Data representativeness

In the first wave of the BHPS, 5511 households were included. Throughout the panel, many households have dropped out from the survey. The rate of dropout is found higher in the first few waves and the dropout rate is found to gradually decrease in the subsequent waves. However, a small number of households who dropped out in the initial waves were observed to re-entered in the survey (Figure 3-2).

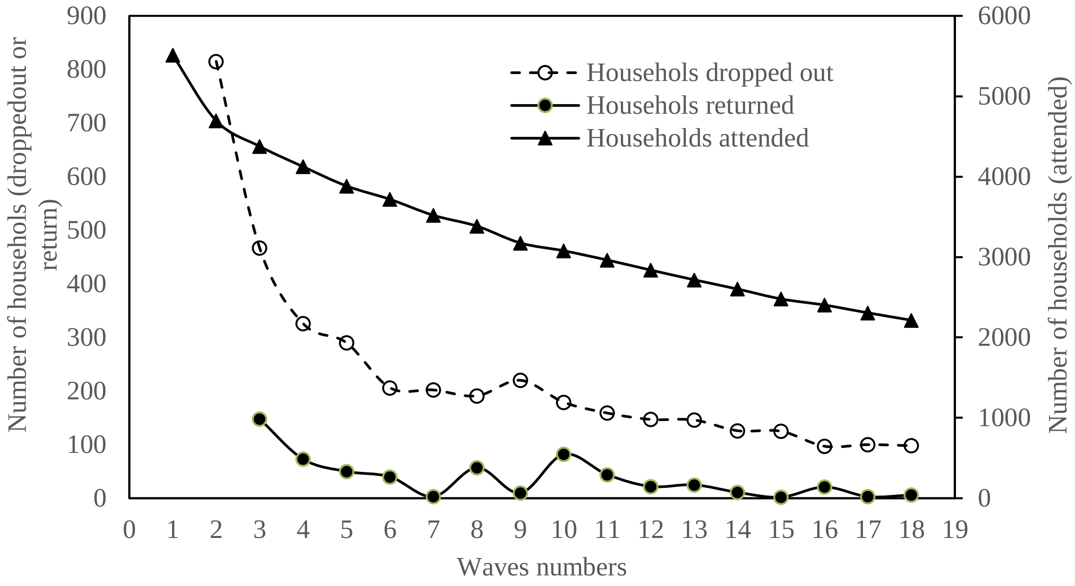


Figure 3-2 Number of retentions, dropout and re-entrance over 18 years

Due to a significant level of dropout, the balanced panel (households who attended the survey continuously) consists of 1617 households and due to missing information (e.g. non-response to the residential location questions), the final dataset of the balance panel further reduced to 1454 households. Since the information at wave t is used to observe behaviour (e.g. residential mobility) at t+1 wave, each individual in the balanced panel consists of 17 observations resulting total of 24718 observations in the final dataset. The unbalanced panel (which includes the respondents for whom data is not available in 18 waves) consists of 50282 observations. Due to the significant number of dropouts from the panel, the representativeness of both balanced and unbalanced panels is likely to be affected. For instance, if the dropout rate is higher among the renters, the panel may have over-representation of the owners and the estimation results will be dominated by their behaviour. Therefore, the representativeness of the subsample (balanced panel) in relation to the full sample (all households included in wave 1) is investigated using the Chi-square test.

Chi-square test of goodness of fit is a widely used technique for assessing the sample representativeness that can be applied at the level of attribute to identify the attributes which may make a sample nonrepresentative (e.g. Griffin et al., 2015; Fasbender, Devos and Lemajic, 2017). The null hypothesis, in this case, is the distribution of household characteristics in the full sample and the sub-sample are similar. The chi-square value for parameters (*k)* are calculated using the formula presented below[3]

$$\chi^2(k) = \sum_{i=1}^{J} \frac{(P_{ik} - Q_{ik})^2}{Q_{ik}} \times \frac{N_s}{100} \qquad (3\text{-}1)$$

P_{ik} *and* Q_{ik} are percentages of observations in the subsample and the full sample respectively corresponding to the category *i* of attribute *k.* N_s is the number of households in the subsample. The degree of freedom (*DF*) is the number of categories under each attribute (*J*) minus 1.

The results of the Chi-square test for the key socio-demographic characteristics of households are presented in Table 3-1. As seen in the table, the Chi-square stat rejects the null hypothesis for eight out of eleven attributes at 95% confidence interval which implies that the dropout in the BHPS is non-random and requires appropriate corrections. A review of the literature reveals that the weighting of the data is a suitable technique to reduce bias due to non-random dropout in the panel survey (Vandecasteele and Debels, 2006).

---

[3] The chi-square value needs to be calculated from the actual frequency. The term $\frac{N_s}{100}$ in the equation *3-1* converts relative frequencies P_{ik} and Q_{ik} into actual frequency.

Table 3-1 Chi-square goodness of fit test for the sub-sample

Variables	Sample distribution (%): Full sample (wave 1)	Sample distribution (%): Sub sample (wave 1)	Chi-square (category)	Chi-square (total)	Chi-square critical value (95% CI)
Household type					
Single member household	26.7	19.5	28.5		
Couple without child	27.8	28.5	0.3	72.1	7.81
Couple with child	33.5	42.6	35.5		
Lone parents	12.0	9.4	7.9		
Household income in GBP					
Less than £20,000	69.7	59.6	21.5		
Between £20,000 to £40,000	25.6	33.7	37.5	71.9	5.99
More than £40,000	4.7	6.7	12.8		
Education attainment of household head					
Below O level	51.5	40.1	36.8		
O and A level degree	34.2	39.4	11.4	89.0	7.81
Graduate degree	12.5	18.2	39.0		
Post-graduate degree	1.8	2.2	1.8		
Age of household head					
Less equal to 30 years	16	13.8	4.3		
Between 31 to 40 years	20.2	24.3	12.0		
Between 41 to 50 years	18.9	25.2	30.3	147.9	9.49
Between 51 to 60 years	14.2	18.9	22.7		
More than 60 years	30.7	17.8	78.7		
Number of employees in the household					
No employee	34.6	19.6	94.5		
One employee	28.8	32.4	6.7	152.3	5.99
More than one employees	36.7	48.0	51.1		
Tenure type					
Owned house	66.5	79.8	38.6		
Rented social housing	20.7	14.2	29.1	120.9	5.99
Rented private housing	12.8	6.0	53.2		
Presence of senior adult (>75years)					
Yes	12.07	2.5	110.9	126.1	3.84
No	87.9	97.5	15.2		
Length of current job of household head					
Less than 5 years	50.0	55.1	7.3		
Between 5 to 10 years	19.8	23.0	7.9	48.2	5.99
More than 10 years	30.2	21.9	33.0		
Having a child in last one year					
Yes	7.1	7.8	0.9	1.0	3.84
No	92.9	92.2	0.1		
Changed job in last one year					
Yes	15.4	16.0	0.4	0.4	3.84
No	84.6	84.0	0.1		
Residential Location before move					
London	9.0	9.6	0.6	0.6	3.84
Other cities	91.0	90.4	0.1		
Sample size	5511	1454	-	-	-

3.2.3.1 Estimation of the sampling weights

Sampling weight is the inverse of the selection probability of a sampling unit. Sampling weights for under or over-represented groups can be calculated as the ratio of their shares in the population and corresponding shares in the sample. The sampling weight estimation is complex when the sample is non-representative for multiple characteristics of the households in the dataset. The easiest way is to calculate the weights for each population's characteristics independently and take the product of them or calculate the weights sequentially. However, neither of these techniques produces accurate weights if the parameters are correlated to each other because the weighting of the sample for one parameter is most likely to change the distribution of other correlated parameters in the sample (Fotini et al., 2013). For example, highly educated people are more likely to be in the high-income group. Correcting the sample representativeness for education qualification will, therefore, change the income distribution in the sample. Therefore, the joint distribution of multiple characteristics of sample nonrepresentation needs to be considered for calculating the weights. Raking or iterative technique is a more accurate and widely used technique for this purpose (Johnson, 2008; Fotini et al., 2013). Raking uses the Iterative Proportional Fitting (IPF) algorithm which uses the known population distribution for adjusting the sampling weights so that the marginal values of a table sum to those known totals. Iteration is used until the weights converged and stop changing (Anderson and Fricker Jr, 2015). Raking forces the sample distribution to match with population distribution by assigning a weight for each respondent.

Weights are calculated using an automatic raking/iterative technique for the household characteristics which are found to be different in the sub-sample and the full sample based on the chi-square test (Table 3-1). The weights ensure that the distribution of the household characteristics in the weighted sub-sample is equivalent to those in the full sample. ‘Rake’ function in the R software package “Survey” is used to calculate the weights. An example of the R code generated for calculating the weights is provided in Appendix B.

Since stratified random sampling technique is adopted in BHPS, initial sampling weights (design weight and weight for non-response) for the households attended in the survey wave *1* are provided with the dataset. These weights are also considered to ensure the representation of the sample to the population. Therefore, the final weight for each household is the product of the initial weight provided with the dataset and the weight calculated to adjust the sub-sample with the full sample. The weights thus correct the over and under-representation of different population groups in the dataset due to non-random dropouts and ensure that the balanced sample (consisting of respondents who have stayed in all 18 waves) is a representative sample in the base year (wave 1). Consequently, the estimated coefficients for the weighted sample are likely to represent the behaviour of the population.

3.2.4 Correlation between panel dropout and household behaviour

The primary cause of panel dropout in the BHPS data is the refusal (no longer interested to join the survey) followed by the non-contact which includes people who died, moved out from the country, staying outside home mostly and a higher likelihood of moving home (Uhrig, 2008). Since the likelihood of moving home is not a leading cause for panel dropout in BHPS, dropout may not affect the mobility behaviour in the balanced panel significantly.

The correlation between the panel dropout and the residential mobility rate in the BHPS is investigated. The households who dropped out from the panel in the first few waves are observed to have a larger number of residential moves compared to the households who remained in the survey. However, the residential mobility behaviour of the households who dropped out after wave seven is very similar to those who attended in all the waves (Figure 3-3). Due to the higher rate of residential mobility of the panel droppers, the rate of residential mobility is higher in the unbalanced panel for the first few waves compared to the balanced panel (residential mobility rate in the unbalanced panel is 5.0% which is 4.2% in the balanced panel).

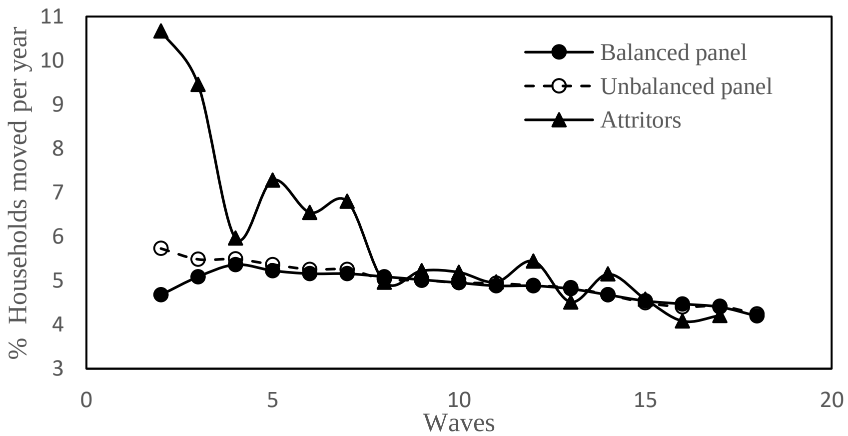


Figure 3-3 Residential mobility rates of the households in the BHPS

3.2.4.1 Adjustment of under-reported residential mobility behaviour in the weighted balanced panel

Panel dropout affects the representativeness of the sample which is most likely to be correlated with the observed behaviour in the dataset. Weighting of the sample can thus help for sample representativeness and associated behavioural adjustment. For example, the share of the young people in a population is 30% who have 20% extra residential moves, if the balanced panel consists of only 10% of young people (due to panel dropout), the number of residential moves is likely to be under-reported. If sampling weights are used to adjust the total number of young people in the balanced panel, it will eventually increase the number of residential moves in the panel. Therefore, the inverse probability weighting technique (or sampling weights) has been used in literature for correcting the bias in the model parameters due to sample non-representativeness and related behaviour issues due to panel dropout (Fitzgerald et al., 1998; Wooldridge, 2010).

The residential mobility rates in the balanced panel of BHPS is found lower than that in the unbalanced panel due to the under-reporting in the balanced panel and over-reporting in the unbalanced panel[4]. Weights corrected the under-reported mobility behaviour in the balanced panel at a very significant level. Therefore, the residential mobility rate in the weighted balanced panel is found very close to that in the unbalanced panel (Table 3-2). The difference in the mobility rates between the weighted balanced and unbalanced panel is most likely to be attributed by the over-representation of the mobility behaviour of the households in the unbalanced panel who dropped out very early. However, the sampling weights may not solve the non-represented behaviour due to drop out if the panel droppers and panel stayers from the same sociodemographic class show different behaviour. Since weighting of the BHPS data has adjusted the overall residential mobility rates in the balanced panel, this issue is likely to be significantly minimized. On the other hand, the use of the unbalanced panel for estimating model parameters is also risky due to the over-representation of the behaviour.

[4] Unbalance panel is most likely to have over representation of the total number of residential moves. Because the households who left the survey very early had larger number of moves. If they continued to the survey until the end, their mobility rate is likely to decrease (a similar trend has been observed for the households who attended in the several waves and then remained or dropped out).

Table 3-2: Residential mobility rates in unbalanced and weighted balanced panels

Residential behaviour	Unbalanced panel	Balanced panel (before weighting)	Balanced panel (after weighting)
% of the households moved	5.0	4.2	4.8

3.2.5 Data analysis

The residential mobility rate of the households in the BHPS dataset is very low. The number of households that moved in a given year varies between 3% to 6% across the waves. Among all the residential moves, more than 60% happened locally (within the same neighbourhood), around 20-25% happened at the regional level and the remaining 15-20% happened at the national level (Figure 3-4).

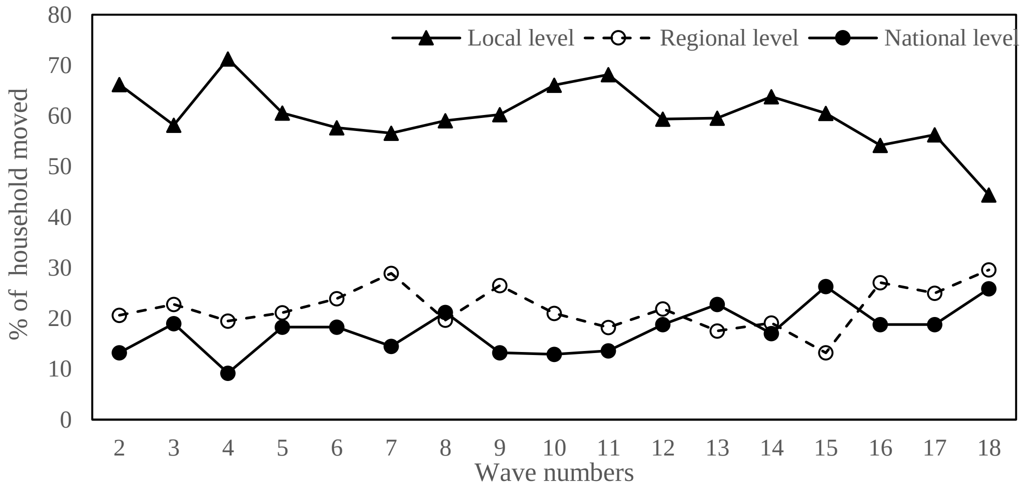


Figure 3-4 Split of relocation at different geographical scales across the waves

Table 3-3 presents the distribution of the characteristics of the households who did not move, who moved at the local level, who moved at the regional level and who moved at the national level. The table values reflect that sociodemographic characteristics and travel behaviour of the households of these four groups (did not move, moved locally, moved regionally and moved nationally) were different from each other before their residential moves. As observed in the table, the group that moved nationally has a considerably higher share of high-income households (annual income above 40,000 GBP) and highly educated people (graduate or postgraduate) compared to the other groups. Similarly, social renters have a higher share at local level relocation compared to regional and national level relocation. On the other hand, private renters have the highest share at regional level relocation compared to the other groups. In case of daily travel behaviour, the average commute distance of the households that have moved at a national level is found to be higher than the average

Table 3-3 Descriptive statistics of the households that have moved in different geographical scales.

Variables	Residential mobility (%) Stayed (SC)	Moved local level (ML)	Moved regional level (MR)	Moved national level (MN)
Sociodemographic characteristics				
Household type				
Single member household	28.2	28.5	29.0	26.7
Couple without child	30.8	23.5	34.4	36.3
Couple with child	31.9	34.1	30.0	31.5
Lone parents	9.1	13.8	6.5	5.6
Household income in GBP				
Less than £20,000	54.3	52.3	51.5	37.3
Between £20,000 to £40,000	30.6	33.0	34.9	38.6
More than £40,000	15.1	14.7	13.7	24.1
Education attainment of household head				
Below O level	47.2	41.6	26.5	29.5
O and A level degree	35.4	37.4	50.6	31.0
Graduate degree	14.7	16.1	18.5	27.4
Post-graduate degree	2.7	4.9	4.3	12.1
Number of employees in the household				
No employee	38.4	31.2	29.5	28.5
One employee	25.5	28.8	34.4	31.7
More than one employees	36.1	40.1	36.1	39.8
Tenure type (%)				
Owned house	75.2	52.3	56.3	72.2
Rented social housing	19.0	24.4	11.5	7.7
Rented private housing	5.8	23.4	32.2	20.1
Presence of senior adult (>75years)				
Yes	16.9	9.1	6.2	15.0
No	83.1	90.9	93.8	85.0
Job length of household head (years)				
Mean	9.8	8.8	6.2	8.8
Standard deviation	14.1	13.0	9.3	13.5
Crowd (household size/number of rooms)				
Mean	0.5	0.8	0.6	0.7
Standard deviation	0.4	0.7	0.7	0.6
Life events				
Having a child in last one year				
Yes	3.8	8.2	8.6	6.6
No	96.2	91.8	91.4	93.4
Changed job in last one year				
Yes	12.2	18.4	22.2	20.5
No	87.8	81.6	77.8	79.5
Travel characteristics				
Travel distance (kilometre)				
Mean	6.3	8.8	7.0	10.3
Standard deviation	18.9	18.9	18.4	24.2
Travel mode				
Car	73.3	56.7	65.1	67.9
Public transport (PT)*	10.5	14.6	3.5	28.5
Active travel mode (AT)*	11.9	9.5	10.4	10.1
Residential location before move				
London	10.6	12.5	0.9	37.1
Other cities	89.4	87.5	99.1	62.9
Number of observations	23675	636	230	177

* Public transport includes underground/tube, train and bus; active travel includes bicycle and walking.

commute distance of two other groups (stayed, moved at the local level and moved at the regional level). Households that have made a national-level move are found to be more avid users of public transport compared to the households that have moved at the local level, regional level or did not move.

3.3 Model development

3.3.1 Model structure

Households who moved in different geographical scales (local, regional and national level) may have different reasons for doing so. In a modelling context, it is proposed to test if there are significant differences among the parameters depending on whether a household moved at the local, regional or national level. The distinct nature of different scales of residential mobility decisions can influence the household car ownership and travel behaviour in a different manner (investigated in Chapter 4).

Residential mobility is a time-dependent household or individual level decision. Therefore, previous choice, length of the current choice and the dynamics in the life course can influence the residential mobility decision. This study aims to capture the dynamic effect in the residential mobility behaviour through time varying covariates (investigating the changes in household behaviour over time in response to the change in their sociodemographic state and life events). Discrete choice model is found very straight forward in this context. However, the limitation of this approach to capture the time dependency of the behaviour (duration dynamics) is acknowledged. It may be noted that hazard-based model has the potential for capturing the time dependency of the residential mobility behaviour since the duration in the current choice is captured here as a stochastic process. However, this technique has less flexibility for testing time varying covariates (Fisher and Lin, 1999) (more details are given in Chapter 2). The mathematical formulation of the random utility based discrete choice modelling technique is discussed below.

In random utility theory, a decision maker maximizes utility for choosing an alternative in a given condition. Therefore, the utility equation to choose alternative *i* by individual *n* at choice condition *t* can be expressed as follows

$$U_{nit} = \beta_i x_{nit} + \alpha_i + \varepsilon_{nit} \quad (3\text{-}2)$$

x_{nit} is a vector of observed variables, β_i is the corresponding coefficient vector and α_i is the alternate specific constant which captures the unobserved utility. ε_{nit} is IID (independent and identically distributed) extreme value type I error term. The multinomial logit (MNL) model formulation for calculating the probability of choosing alternative *i* by individual n at choice situation *t* can be expressed as

$$P_{nit} = \frac{e^{\beta_i x_{nit} + \alpha_i}}{\sum_{j \in C} e^{\beta_j x_{njt} + \alpha_j}} \qquad (3\text{-}3)$$

The logit model formulation has a limitation in terms of capturing the unobserved random heterogeneity across the individuals due to IID restriction. Estimation of logit models in the presence of potential random taste heterogeneities across the individuals, the estimated parameters are more likely to be biased. Mixed multinomial logit (MMNL) formulation of random utility theory has the flexibility to capture the random heterogeneity in the observed and unobserved components of utility across the individual and correlation across the alternatives. To capture the random taste heterogeneity in the unobserved component, the alternate specific constants (α) are decomposed into their mean (μ) and deviation (σ). Then, the equation *3-1* can be revised as follows[5].

$$U_{nit} = \beta_i x_{nit} + (\mu_i + \sigma_i \xi_{ni}) + \varepsilon_{nit} \qquad (3\text{-}4)$$

ξ_{ni} is a vector of random variables that are not observed by other model components. The random variable is normally distributed and the vector dimension is *(NxJ)*. *N* and *J* represent the total number of individuals and the total number of alternatives respectively. Due to the panel nature of the data, an individual might have repeated choice at choice situation *t. t=1,2,3*…………*.T*. The repeated choice probability can be formulated as follows

$$P_{ni}(\sigma) = \prod_{t=1}^{T} \frac{e^{(\beta_i x_{nit} + (\mu_i + \sigma_i \xi_{ni}))}}{\sum_{j \in C} e^{(\beta_j x_{njt} + (\mu_j + \sigma_j \xi_{nj}))}} \qquad (3\text{-}5)$$

To capture the serial correlation, the same random value is needed to be considered for the repeated choices of individual *n*. The unconditional probability can be obtained by integrating the conditional probability in the above equation with respect to the assumed independent normal distributions for the vectors ξ. The mathematical expression of unconditional probability is presented below

[5] To capture the randomness in the observes utility, the equation can be reformulated as $U_{nit} = (\mu_i + \sigma_i \xi_{ni}) \beta_i x_{nit} + \varepsilon_{nit}$

$$P_{ni} = \int_{\xi=-\infty}^{+\infty} \left[\prod_{t=1}^{T} \frac{e^{(\beta_i x_{nit} + (\mu_i + \sigma_i \xi_{ni}))}}{\sum_{j \in C} e^{(\beta_j x_{njt} + (\mu_j + \sigma_j \xi_{nj}))}} \right] f(\xi) d\xi \qquad (3\text{-}6)$$

Since the probability function in the above equation contains a multi-dimensional integral and it does not have a closed-form solution, probabilities are approximated through simulation (Train, 2009). Then, the simulated probabilities are considered into the log-likelihood function to get a simulated log-likelihood. The simulated log-likelihood function is given as bellows:

$$LL = \sum_n \sum_i y_{ni} \, lnP_{ni} \qquad (3\text{-}7)$$

where y_{ni} = 1 if person *n* chose alternative *i* and zero otherwise. Models are evaluated based on the goodness of fit and the t-stats of the estimated parameters. To test the statistical significance of the differences in the estimated parameters in the models, *t-stat* difference test statistic is used in this study[6].

$$t_{diff} = \frac{\beta_{ik} - \beta_{jk}}{\sqrt{Var_{ik} + Var_{jk} - 2Cov_{ijk}}}, \qquad (3\text{-}8)$$

Where $Var_{ik} \left(= (\frac{\beta_{ik}}{t_{ik}})^2\right)$ and $Var_{jk} \left(= (\frac{\beta_{jk}}{t_{jk}})^2\right)$ are the variance of the attributes k in two different contexts and Cov_{ijk} is their covariance. β_{ik} and β_{jk} are the estimates of k^{th} attributes of the model in two different contexts, t_{ik} and t_{jk} are the respective *t* ratio of the estimated parameters. The differences in estimated parameters are significant at the 95% level of confidence if the absolute value of t_{diff} exceeds 1.96.

The likelihood ratio (LR) test value is used for comparing the goodness of fit of competing models (a null model against an alternative model). The LR was calculated using the following equation

$$LR = -2[LL(\beta_a) - LL(\beta_n)] \qquad (3\text{-}9)$$

Where $LL(\beta_a)$ is the log-likelihood for the alternative model and $LL(\beta_n)$ is the log-likelihood of the null model. The LR can be compared to a critical value from a χ^2_n distribution with *n* degrees of freedom, where $n=K_a - K_n$, with K_a and K_n *are* the number of estimated parameters in the alternate model and null model respectively.

[6] For comparing the parameters from the same model where the variances are not independent, the conveniences are needed to be subtracted. $Cov_{ijk}=0$ when the parameters from the two independent modes are comparing.

3.3.2 Design of choice alternatives

Residential mobility decision consists of the decision to move or stay and its geographical scale consists of moved at the local level or regional level or national level. Therefore, the joint decision of residential mobility and its geographical scale consists of four alternatives presented in Table 3-4.

Table 3-4 Alternatives of decision of residential mobility and its geographical scale

Mobility decision	Geographical scales	Joint alternatives
Stayed	Same place	Stayed at same place (SS)
Moved	At the local level	Moved at local level (LL)
	At the regional level	Moved at regional level (RL)
	At the national level	Moved at national level (NL)

3.4 Estimation results

Household residential mobility decision (decision to stay or move) and joint decision of residential mobility and its geographical scales are modelled here. Joint estimation captures the differences in the parameter sensitivity of the households who have moved in different geographical scales.[7] Multinomial logit (MNL) and mixed multinomial logit (MMNL) models are estimated using statistical software R[8]. The panel nature of the data provides an opportunity to capture correlation across the repeated choices over a long period of time and unobserved taste heterogeneities across the individual households. Therefore, in the MMNL model, both observed and unobserved components of the utility are allowed to vary randomly across the alternatives to capture the potential taste heterogeneities. However, heterogeneity in the observed components is found insignificant after capturing the random heterogeneity in the unobserved (constant) terms. Therefore, the final models include the random terms in the constants only. Different model specifications are tested to capture the correlations across the alternatives (using both nesting structure and

[7] The decision of residential mobility and its geographical scale is demographically driven household or individual level decision. Thus, the joint model attempts to capture the differences in the household's circumstances for considering residential mobility in different scales. This joint decision is mostly independent of the characteristics of location or neighbourhood. Choice of location captures household preference for location or neighbourhood characteristics where household evaluates a set of alternative locations to choose the best option.

[8]The R codes from Choice Modelling Center, University of Leeds are used for estimating the MNL and MMNL model parameters. However, these codes are modified according to the model specifications and needs (e.g. adopting the sampling weight).

Cholesky decomposition) but improvements of these models in terms of goodness of fit are not found significant compared to the heteroscedastic model (details are given in Appendix C). To meet the rank and order condition, at least one random term must be normalized. The best normalization is to set the random term of the minimum variance alternative to zero (Walker, 2001). In this model, the random term of the base alternative is found to have the minimum variance and therefore, is normalized it to zero.

The likelihood ratio test value is used to evaluate the goodness of fit of the MMNL estimation over the MNL estimation. The null hypothesis of the MNL model is rejected by the Chi-square statistics for 99.9 % confidence interval revealing a significant level of taste heterogeneity and correlation across the choices over time. The estimation results using the balanced panel are summarized in Tables 3-5 and discussed in the following sections. However, models are also estimated for the unbalanced panel to see the differences in the behaviour of households in the balanced and unbalanced panel. The results of the unbalanced panel are presented in Appendix D.

3.4.1 Estimation of residential mobility decision

Household socio-demographic characteristics, life events and travel characteristics are considered to explain the residential mobility behaviour where did not move is considered as the base alternative[9]. The observed negative significant coefficient for the constant term indicates that the households have baseline preference of staying in their current location which is consistent with the findings in the literature (Kortum et al., 2012). As expected, household-level characteristics are found to influence the residential mobility decision significantly. Single member households are observed to have the highest disposition to move whereas households having children are found to have the least disposition to move. The less propensity of moving of households having kids is more likely to be driven by the connection with the local neighbourhood due to kid's schooling. This finding is consistent with the ongoing literature on residential mobility and family composition (e.g. Van Ham and Clark, 2009; Clark, 2013) although Clark and Drever (2000) argued that couples with children are more

[9] This study acknowledges the multicollinearity issue because some of the independent parameters such as household income, number of employees can be correlated to each other.

Table 3-5 Estimation results for model of decision to move

Parameters	Residential mobility decision				Joint decision of residential mobility and its scale (MMNL)						t difference test		
	MNL		MMNL		Local level (LL)		Regional level (RL)		National level (NL)		LL & RL	LL & NL	RL & NL
	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat			
Alternative specific constants													
(not moved is the base alternative)													
Mean	-4.7874	-29.6	-4.9985	-27.3	-5.6400	-25.5	-6.6415	-30.3	-6.7952	-17.8	3.1	2.3	0.5
Standard deviation	-	-	-0.7719	-14.7	-0.7894	-12.1	1.0096	8.4	0.9223	5.7	-13.2	-9.2	0.4
Household level characteristics													
Household type (base is couple with child)													
Single member household	0.8752	7.0	0.7926	5.6	0.8536	5.0	0.8756	4.1	0.5353	1.8	-0.1	0.9	1.1
Couple without child	0.5536	5.8	0.4900	4.5	0.3953	3.0	0.7186	4.2	0.5448	2.5	-1.6	-0.6	1.2
Lone parents	0.4207	3.6	0.2843	2.1	0.4832	3.1	0.0417	0.1	-0.3728	-1.1	1.4	2.5	0.9
Household income (base is less than £20,000)													
Between £20,00 to £40,000	0.2769	3.2	0.2336	2.5	0.1778	1.6	0.2000	1.0	0.4564	2.2	-0.1	-1.2	-0.8
More than £40,000	0.2797	2.4	0.1568	1.2	0.0572	0.4	0.1570	0.6	0.5921	2.3	-0.3	-1.9	-1.0
Education attainment of household head (base is below O level)													
O and A level degree	0.2949	3.9	0.2980	3.1	0.1791	1.6	0.7410	3.9	0.1474	0.7	-2.7	0.1	2.2
Graduate degree	0.4473	4.5	0.4001	3.2	0.2566	1.7	0.7530	3.1	0.7576	3.3	-1.9	-2.1	0.0
Post-graduate degree	1.0339	6.8	1.1325	5.6	0.9846	4.0	1.1217	2.9	1.4780	4.6	-0.3	-1.4	-0.9
Number of employees in the household (base is no employee)													
One employee	0.0144	0.2	0.0298	0.3	0.0450	0.4	0.0677	0.4	-0.1324	-0.5	-0.1	0.7	0.8
More than one employees	0.0965	0.9	0.1101	0.9	0.3254	2.3	-0.2119	-1.0	-0.3334	-1.2	2.0	2.2	0.4
Length of current job of household head	-0.0149	-3.2	-0.0141	-2.8	-0.0071	-1.2	-0.0321	-2.7	-0.0166	-1.5	2.0	0.8	-1.2
Presence of senior adult (>75 years)	-0.4362	-3.7	-0.4397	-3.4	-0.5982	-3.7	-0.7175	-2.2	0.2877	1.1	0.3	-3.4	-3.3

Table 3-5 Estimation results for model of decision to move (cont.)

Parameters	Residential mobility decision				Joint decision of residential mobility and its scale (MMNL)						t difference test		
	MNL		MMNL		Local level (LL)		Regional level (RL)		National level (NL)		LL & RL	LL & NL	RL & NL
	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat			
Dwelling level characteristics													
Tenure type (base is owned house)													
Rented social housing	0.3571	3.8	0.3967	3.5	0.6570	5.1	0.0537	0.2	-0.6642	-2.2	2.1	4.4	1.8
Rented private housing	1.7242	20.6	1.9368	18.1	1.9209	15.1	2.1278	11.2	1.3763	6.4	-1.0	2.6	3.7
Crowd (household size\number of rooms)	1.2437	10.5	1.2721	9.2	1.3908	8.6	1.0380	4.1	1.0070	3.6	1.5	1.3	0.2
Life course events													
Having child in last one year	0.5433	4.2	0.3916	2.8	0.3506	2.1	0.6260	2.5	0.4402	1.3	-0.9	-0.3	0.4
Changed job in last one year	0.1763	2.0	0.1289	1.4	0.1443	1.2	0.0837	0.4	0.2133	1.1	0.3	-0.3	-0.5
Location characteristics													
Metropolitan area (base is other than London)													
London	0.2192	2.4	0.3038	2.6	0.0011	0.0	-2.6171	-4.2	1.6700	8.7	4.4	-8.3	-8.4
Measures of model fit													
Number of observations	24718		24718		24718								
Initial LL	-17133.2		-17133.21		-34266.40								
Final LL	-4313.72		-4256.64		-5210.83								
Likelihood ratio test	-	-	114.2000										
Chi-square stat (1,0.001)	-	-	10.8280										

likely to move compared to the couple without child. The opposite finding of Clark and Drever (2000) is likely to be driven by the households having new-born child which is captured in this study separately.

The probability of moving is higher for people with a higher level of income. This may be due to inclination for better lifestyle preferences and affordability to change tenure type (i.e. switch from renting to owning) of the middle and higher income people. This finding is consistent with the finding of the studies of Clark (2013) and Van der Vlist et al. (2002) although the later study founds a negative association for renters. Highly educated people are also found to have a higher propensity to move. This phenomenon (also observed by Van der Vlist et al., 2002; Van Ham and Clark, 2009; Kortum et al., 2012), maybe due to their higher access to opportunities (specifically in the job market). The presence of senior adults (more than 75 years of old) in the household is found to reduce the likelihood of moving. (see Van Ham and Clark, 2009; Kortum et al., 2012 for similar finding). This may be due to the fact that most elderly people are more settled in their place and their physical condition constrains to move frequently. Working with the same employer for a long time is found to reduce the propensity of moving home (also observed by Clark and Withers, 1999). The role of dwelling characteristics on residential mobility decision is also found significant. For instance, households living in rented private housing are found to be more likely to move compared to households living in rented social housing or owned houses (also observed by Clark and Drever, 2000; Eluru et al., 2009; Tatsiramos, 2009; Van Ham and Clark, 2009; Clark, 2013). Due to a large investment and high relocation cost owners are less likely to move frequently. On the other hand, social renters do not have free choice to move to another socially rented house and they are also less likely to move in privately rented house or owing a house ultimately leading to less likelihood of relocation. Higher crowding level (denoted as the ratio of number of household members and number of rooms) is also found to increase the likelihood of move (Clark and Drever, 2000; Van Ham and Clark, 2009). Life events such as having a child is found to increase the likelihood of moving home significantly (also reported by Clark, Deurloo and Dieleman, 1984; Clark and Withers, 1999) whereas the impact of job change on residential relocation is also positive but statistically insignificant at 90% confidence interval.

3.4.2 Joint estimation of residential mobility and its geographical scale

A considerable level of differences is observed in household preferences depending on the geographic scale of residential mobility. *t*-difference tests (equation 3-7) are used in order to investigate whether the differences are statistically significant or not. Most of the parameters demonstrate a certain level of sensitivity differences from one to another scale of relocation, however, ten out of the eighteen parameters are found significantly different in at least one pairs (e.g. local level vs regional level relocation, local level vs nation level relocation or regional level vs national level relocation) based on t difference test results. Although single-member households and couples without children(s) are likely to move at the local, regional and national level, the lone parents are not significantly interested to move beyond the local area. Clark (2013) found a similar result that having a child decreases the propensity of moving very far. Households with high income (>£40,000) have a higher likelihood of moving nationally (also observed by Clark and Huang, 2004; Clark, 2013). The likelihood of high-income households to move locally or regionally is not found to be statistically different from 0 as 90% level of confidence. Highly educated (post-graduate) people are found to have the highest propensity to move at the national level than regional and local level (similar observation by Clark and Huang, 2004; Clark, 2013), whereas, less educated people (O and A level degree holder) are more likely to move at the regional level. This difference may be due to limited access and knowledge about the job markets in different regions or metropolitan areas. Households having more than one employed member are found to be more likely to move at the local level. This may be due to the complexity in an adjustment of the commute distances and/or job-relocation issues of multiple working members in the households arising from the regional and national moves. However, Clark (2013) observed that married two-workers households are willing for long distance move to achieve two jobs. The coefficient of the length of current employment of the household head denotes that if the job-tenure is longer, less likely are they inclined to move. The coefficient is however statistically significant in case of regional level move only. Households having senior adults are found to be less likely to move in general with the propensity to move being less for the regional level and statistically insignificant for the national level. Private renters are more likely to move at the local, regional and national levels while social renters are only inclined to move locally. The influence of life events such as childbirth on inter-regional mobility is not found significant. Clark (2013) also observed that the birth of a child reduces the chance of inter-metropolitan relocation. Londoners are found more likely to move out from the greater London area (GLA) but they are unlikely to move within the GLA.

3.5 Validation results

The MMNL models using the full dataset outperform their MNL counterparts in the estimation context. However, there is a risk that the MMNL model overfits the estimation data. To check for potential overfitting issue, the performances of both the MNL and MMNL models are tested using a holdout sample validation (as used by other researchers: de Luca and Cantarella, 2016; Bwambale et al., 2017 for example) where randomly select 60% of the households are considered for estimation (who are consistently available in the panel) and the other 40% of the households for out of sample prediction. Models are re-estimated again using the estimation subsets of the data from the different random draws. Interpretation of the estimation results of the models remains the same as the interpretation of the model estimated using the full dataset (explained in the previous sections). The goodness of fits of the models are presented in Table 3-6. The estimated model parameters are then applied to the validation sample to investigate the predictive performance of each of the models. The same procedure is repeated for three times to check whether the performance is consistent over the different split of the dataset based on different independent random draws.

The predictive power of the models is evaluated in terms of improvement in goodness-of-fit (log-likelihood in prediction sample and predictive rho-square). The results are presented in Table 3-7. It is observed that the MMNL models of residential mobility decision perform better than the corresponding MNL models in the estimation sample and hold a consistent performance in the hold-out sample.

Table 3-6 Goodness of fit of the models estimated using estimation subset of data

Draws	Number of observations	Initial log-likelihood	Final log-likelihood		Adjusted rho-square	
			MNL	MMNL	MNL	MMNL
D1	14824	-20550.4	-3066.9	-3044.8	0.848	0.849
D2	14824	-20550.4	-3090.9	-3058.4	0.846	0.848
D3	14824	-20550.4	-3165.4	-3143.0	0.843	0.844

Table 3-7 Validation results

Draws	Number of observations	Initial log-likelihood	Final log-likelihood		Predictive rho-square	
			MNL	MMNL	MNL	MMNL
D1	9894	-13716.0	-2253.1	-2226.0	0.831	0.833
D2	9894	-13716.0	-2225.3	-2208.3	0.833	0.834
D3	9894	-13716.0	-2159.6	-2131.7	0.838	0.840

Further, to demonstrate the value of the developed models in the context of forecasting, the model performance is assessed in the prediction of future years. To demonstrate the performance of the MNL and MMNL models in the context of forecasting, the data from waves 1-14 is used for estimation and applied the model estimates for predicting the decisions made in the last three years (waves 15-17). The results are presented in Table 3-8. It is observed that the MMNL model performs better than the MNL model in estimation but in the prediction, the improvement of the MMNL model compare to the MNL model is small.

Table 3-8 Model performance in prediction

Sample	Number of observations	Initial log likelihood	Final log likelihood		Predictive rho-square	
			MNL	MMNL	MNL	MMNL
Estimation sample (waves 1 to 14)	20356	-28219.4	-4461.6	-4420.0	0.840	0.841
Validation sample (waves 15 to 17)	4362	-6047.0	-688.7	-682.6	0.876	0.876

3.6 Conclusions and policy recommendations

Two different models are estimated in this chapter for a better understanding of residential mobility decisions. These are:

- A residential mobility choice model to quantify the relative sensitivities of parameters that affecting the decision to move
- Estimation of joint choice of residential mobility decision and its geographical scale to quantify the sensitivities of different factors affecting the decision to move in the local, regional or national level or stay in the current location

BHPS data is used for this study where the same households have been observed over 18 years. MMNL techniques are used to capture the panel effect of the data.

The key findings are as follows:

- Household socio-demographic characteristics (e.g. household income, education level, household type, number of employees, etc.), dwelling characteristics (e.g. household size, tenure type)) and life events (e.g. having child, changing job) are the determinants of residential mobility decision.
- Significant levels of heterogeneities are observed among the different geographical scales of relocation. Ten out of the eighteen parameters are found to be significantly different among different scales of relocation (local vs regional level, local vs nation level, regional vs national level).
- Estimated parameters in the model without considering the geographical scales are significantly different compared to the corresponding scale specific parameters. Therefore, analysing the residential mobility decision without considering the geographical scale may produce biased estimation.

As with most empirical studies, this work has several limitations. The panel nature of the data has the potentiality to capture the dynamics in the household behaviour. Although this study has captured the dynamics by investigating how changes in household circumstances over time influence their behaviour and state-dependence of the choices (included in Appendix E), the modelling technique used in this research did not allow to capture the duration dynamic of decision. The future study is thus recommended for modelling this behaviour using the technique that allows to capture the duration dynamics in the behaviour (e.g. hazard-based model or Markov chain model).

The residential mobility and other decisions are likely to be affected by neighbourhood characteristics such as public transport accessibility, parking availability, land use pattern, etc. These parameters are not available in the dataset and cannot be tested in the current models.

Local or national policies target to benefit residents through improved facilities for a better living standard which indirectly strengthens social connectedness or access to resources. For policy formulation, policymakers assume some degree of residential stability (low residential turnover rate) in the target areas, therefore, a high rate of residential mobility can be a challenge in this context (Kubisch et al., 2010; Silver et al., 2012). Although residential mobility may reflect the improvements in people's circumstances, such as first-time homeownership, moving close to job place, it can also be a sign of job housing imbalance, lack of adequate and quality housing supply, lack of transport and other accessibilities. In addition, residential relocation may create a risk of instability and insecurity in terms of adopting the new environment, losing the social connections and problems with landlords, creditors, or housing conditions. The findings of this chapter have several policy recommendations that may help for minimizing the residential mobility rate. For instance, it is observed from the model output that highly educated peoples have a higher propensity for long-distance relocation. Fewer opportunities in the job market for this group of people (job housing imbalance) or job dissatisfaction may increase the propensity of getting a better job in another area or metropolitan city although it may not be the fact always. Policies targeting the increase in job opportunities and job environment may decrease the long-distance residential mobility rate. A higher level of job facilities is likely to increase the job opportunity for the other members in the household which may further decrease relocation rate (this is also a finding of this study that multiple working member households are less likely to move). For another instance, a higher level of crowd (number of people in the household is high compared to the room available) is also found to push people to move perhaps for the bigger houses. If the dwelling supply is inadequate to meet the demand in the area households are currently living, the dwelling prices are likely to go up (Zhou and Kockelman, 2008; Habib 2009). If the households cannot effort the higher price or rent, they might consider moving further away for affordable housing compromising with the commute and other facilities resulting increase in car dependency and daily VMT. Investigating residential mobility at different scales helps for a better understanding of housing market dynamics in terms of intra and inter housing market mobility. Most importantly, residential relocation has important consequences on household travel behaviour (details are investigated in chapter 4). Therefore, a better understanding of residential mobility will also help in formulating policies for managing the travel demand.

Chapter 4

Modelling the role of residential decision on car ownership change and commute mode transition

4.1 Introduction

Changes in household car ownership and commute mode are medium term household or individual level decisions that can be influenced by household residential change along with other stressors such as sociodemographic change, life events, changes in neighbourhood characteristics, etc. (e.g. Aditjandra et al., 2012; Oakil, 2013; Clark et al., 2016a; Clark et al., 2016b; Fatmi and Habib, 2017). The research framework (Figure 1-3 in Chapter 1) presents these important interconnected household decisions and their dominant causality directions that have been investigated in different chapters. The research framework is also presented here (Figure 4-1) highlighting the components and causality directions captured in this chapter. The rest of this section discusses the gaps identified in the literature in terms of modelling car ownership change and travel mode switching behaviours and their connections with the residential decision.

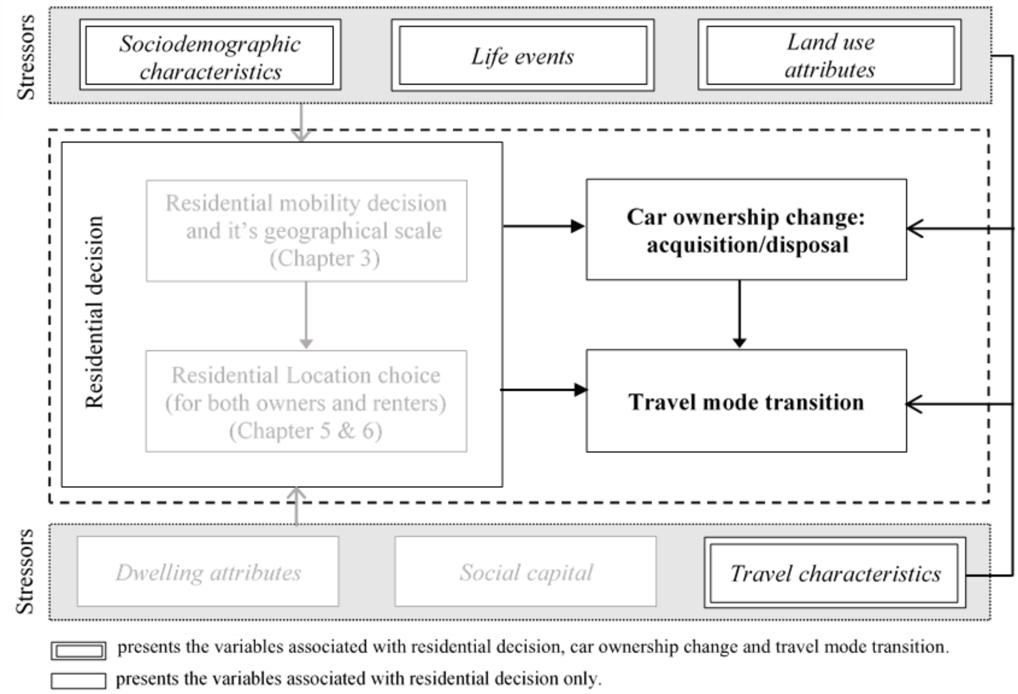


Figure 4-1 Components of the modelling framework that captured in this chapter (highlighted).

As discussed in Chapter 3, the residential mobility can be embedded into three geographical scales (whether moved at the local level or the regional level or the national level) and these scales may have varying impacts on car ownership level. Local level relocation (moving within the same ward) can be regarded as an 'adjustment' move typically prompted by better attributes of the dwelling and is unlikely to have a substantial effect on households' transport and other accessibilities. Therefore, the car ownership level may not be affected significantly due to local level relocation. On the other hand, relocating to another city is more likely to lead to substantial changes in accessibilities and hence car ownership status. For instance, moving to a new city or area with poor access to public transport might push people to buy a car. Previous studies have focused on modelling household car ownership level (Hanly and Dargay, 2000; Dargay and Hanly, 2007; Fox et al., 2017) or changes in car ownership level in two consecutive years (Cao et al., 2007; Aditjandra et al., 2012; Clark et al., 2012; Clark et al., 2016a). However, the effect of the scales of relocations (discussed in detail in section 2.4 of chapter 2) on changes in car ownership level has not been explored yet.

Household commute behaviour can also be associated with the home location. Moving home triggers changes in household commute distance, therefore, may affect the commute mode choice behaviour. Households who relocated locally, commute mode choice is unlikely to be affected if the workplace remains unchanged. However, a regional or national level relocation is more likely to affect household's commute mode choice due to changes in commute distance or changes in transport accessibility. For instance, moving to a location closer to the workplace increases the chance of considering active travel while moving in an area where transport accessibility is very good increases the likelihood of using public transport more (Clark et al., 2016b). So there might be a potential connection between the geographical scale of residential mobility and changes in the commute mode but none of the existing studies have accounted for this issue (e.g. Oakil et al., 2011; Clark et al., 2016b).

The characteristics of location or neighbourhood within the specific geographical scales (local or regional or national) households moved may also influence their car ownership and travel mode changing behaviour. For instance, it has been reported that moving in a deprived area having less access to public transport increases the likelihood of car ownership and car use (Clark et al., 2016a). However, this hypothesis cannot be tested in this study since the dataset does not have information about the characteristics of the neighbourhood households moved.

The existing literature is limited in terms of investigating the specific directions of car ownership change and commute mode switching behaviour. For instance, in case of travel mode switching, Fatmi and Habib (2017) has modelled the behaviour of changing or not changing of commute mode after residential relocation whereas Clark et al. (2016b) has modelled the behaviour of switching to a car or switching from a car. None of these studies have acknowledged the specific directions of switching behaviour such as switching to a car from where: bus or train or active mode of travel. However, each direction of behavioural change may be triggered by a different set of attributes. For example, one household may switch to car from active travel due to a significant level of increase in commute distance whereas another household may switch to car from public transport because of buying a new car or poor accessibility for public transport. Aggregation of alternatives (switching from active travel to car and switching from public transport to car are merged into switching to car only) ignores the potential differences in the conditions specific to switching in different directions. Moreover, the policymakers might also be interested to know about the drivers of switching in each direction. For example, switching to public transport from car may have a different level of impact on the transport system compare to the switching to public transport from active travel. The concluding section presents the potential policy suggestions from this study outcomes. In addition, capturing each switching direction in the separate model as a binary choice (adopted in previous studies such as Oakil, 2013; Clark et al., 2014; Clark et al., 2016b; Fatmi and Habib, 2017) is likely to be cumbersome to estimate and difficult to compare the differences in the behaviour in different directions of switching due to the potential risk of scale differences. In addition, splitting the dataset for capturing each direction of switching separately reduces the benefit of panel data to capture behavioural dynamics and correlation across the choices over time.

Based on the discussion in the preceding section, a few research gaps are identified in the literature. This motivates to address the following research objectives in this chapter:

- To investigate the role of geographical scale of residential mobility on household car ownership and commute mode changes.

- To investigate the relative impact of other factors driving the changes in household car ownership and commute mode and compare them with the contribution of residential relocation.
- To capture all possible directions of switching in a single econometric model.

To investigate these objectives, the British Household Panel Survey (BHPS) dataset (which has been used for modelling residential mobility decision in chapter 3) has been used. The rest of this chapter is organised as follows: the next sections briefly discuss the data used for empirical analysis followed by the model structure. The details of the choice set construction and the model formulations are presented next. This is followed by the model results. The concluding section summarizes the study contributions, limitations and direction for future research.

4.2 Data

4.2.1 Data description

The British Household Panel Survey Dataset (BHPS) is also used in this chapter. As mentioned in chapter 3, the balanced panel that consists of 1,454 households (total 24718 observations in 17 waves) consistently available in all the waves is considered in this chapter. Since the dataset does not have information about the non-work trips of household members, analysis of travel mode switching behaviour in this chapter includes work trips only. Therefore, households having commute members are considered for modelling commute mode transition behaviour and the sample further reduces to around 630 households (10704 observations in total).

4.2.2 Data issues

The representativeness of the data and estimation of sampling weights are discussed in Chapter 3 in detail. Similar to Chapter 3, sampling weights are used in all kinds of analyses in this chapter. It is expected that sampling weights corrected the sample for representativeness and associated behavioural adjustment.

4.2.3 Data analysis

4.2.3.1 Changes in car ownership

Car ownership of the households in the weighted dataset is around 75% which is very close to the national average of 74% (Dargay and Hanly, 2007). The shares of one, two and three car-owning households in the dataset are 44.1%, 24.3% and 6.0% on average respectively. The level of car ownership of each household can change over time. Table 4-1 presents car ownership level changes from one year to the next between 1991 and 2008. It may be noted that the rate of gaining and losing the second car (3.2% and 2.8% respectively) is found higher compared to the rate of gaining and losing the first car (1.3% and 1.4% respectively).

Table 4-1 Household car ownership transaction pathway in two consecutive years

Car ownership transaction pathway			
Number of car (s) at year t	Number of car (s) at year t+1	Number of cases	Percentage
0 car	0 car	5942	24.0
	1 car	316	1.3
	2 cars	16	0.1
	3+ cars	4	0.0
1 car	0 car	345	1.4
	1 car	9817	39.7
	2 cars	797	3.2
	3+ cars	76	0.3
2 cars	0 car	13	0.1
	1 car	694	2.8
	2 cars	4823	19.5
	3+ cars	455	1.8
3+ cars	0 car	6	0.0
	1 car	79	0.3
	2 cars	381	1.5
	3+ cars	955	3.9
Total		24718	100.0

Car ownership level changes are likely to be triggered by changes in sociodemographic status (e.g. income change, change in household size, etc.), life events (e.g. moving house, changing job, getting married, etc.) as well as changes in local and national level policies (e.g. insurance cost, fuel price, etc.). Analysing the data reveals that changes in household size, changes in the number of employees in the households, changes in income level, changes in travel time, moving to a new house, etc. are correlated with gaining and losing of car(s) in two consecutive years (Table 4-2).

Table 4-2 Descriptive statistics of the factors driving the car ownership level changes

Variables	Changes in car ownership level (%)				
	Gained first car	Lost first car	Gained additional car(s)	Lost additional car(s)	No change in car ownership
Sociodemographic characteristics					
Changes in household income					
Income increased	39.7	24.1	53.6	32.0	32.4
Income decreased	18.9	36.2	20.9	43.5	20.0
No change in income	41.4	39.7	25.5	24.5	47.7
Changes in household size					
Household size increased	19.1	6.0	15.2	3.8	3.2
Household size decreased	4.7	22.7	4.7	30.9	3.9
No change in household size	76.3	71.3	80.1	65.3	92.8
Change in number of employment					
Number of employment increased	24.0	10.9	23.5	9.4	7.4
Number of employment decreased	10.1	22.4	9.5	34.3	8.2
No change in employment	65.8	66.7	66.9	56.3	84.4
Presence of senior adults (>75 years)					
Yes	9.4	21.5	3.0	3.8	18.1
No	90.6	78.5	97.0	96.2	81.9
Less educated people (below O level)					
Yes	48.4	54.2	31.1	35.6	47.0
No	51.6	45.8	68.9	64.4	53.0
Tenure type					
Owned house	47.3	54.5	88.5	88.8	73.4
Rented social housing	38.5	32.6	6.6	7.1	19.8
Rented private housing	14.2	12.9	4.9	4.1	6.7
Life events					
Household moved house					
Moved at local level	5.5	7.1	4.2	4.0	2.7
Moved at regional level	1.2	0.9	1.9	0.8	0.9
Moved at national level	3.2	2.1	1.6	1.2	0.7
Stayed	90.1	89.9	92.3	93.9	95.7
Householder changed job					
Yes	15.9	14.0	17.9	16.3	11.9
No	84.1	86.0	82.1	83.7	88.1
Travel characteristics					
Change in travel distance					
Travel distance increased	27.8	32.4	30.0	26.4	25.1
Travel distance decrease	25.1	24.4	23.5	26.7	23.3
No change in travel distance	47.1	43.3	46.4	46.8	51.6
Number of observations	335	364	1328	1154	21537

For example, among the households that have acquired their first car, 19.1% gained members in the households and 24% gained an increase in employed members. On the other hand, among the households that did not acquire or lose car(s), only 3.2% gained new members in the household and only 7.4% gained employment. Elderly peoples are found to have higher proportions of decreasing the number of cars than increasing it with the percentage of moving from one car to no car being the highest (21.5%). The correlation between the geographical scale of residential mobility and car ownership change behaviour is also found statistically significant. For instance, the national level movers are found to have a higher tendency of owning their first car whereas the local level movers are found to have a higher propensity of losing it.

4.2.3.2 Changes in commute mode

Table 4-3 looks at changes in commute mode over time. Only around 6.1% of households in the weighted sample are observed to change their commute mode each year. Many households are found to move to car from both private transport and active travel while the shifting between public transport and active travel is considerably lower. From the statistical analysis of the data presented in Table 4-4, it is observed that moving house, changing job, owning car(s) and a change in commute distance are associated with changes of commute modes in the subsequent year. As seen in Table 4-4, a large shift is observed towards car from public transport and active travel (25.3%) due to the gaining of car(s) by households. Due to an increase in the commute distance, a high rate of switching is observed, particularly to car and/or public transport (74.8% and 63.2% respectively). On the other hand, a decrease in commute distance results in significant levels of shifting towards active travel from both public and private transport (93.3%). Importantly, the correlation between the geographical scale of residential mobility and travel mode switching behaviour is also found significant. The share of switching to car is least among the households who moved at the regional level (1.0%) compared to the households who moved at the local and national levels (6.6% and 5.2% respectively). The national-level movers are found to have lower switches into active travel (0.5%) compared to the other two groups.

Table 4-3 Commute mode switching pathway in two consecutive years

Commute mode switching pathway		Number of cases	Percentage
Commute mode in year t	Commute mode at year t+1		
Public transport (PT)	Public transport (PT)	1164	10.9
	Car travel (CT)	117	1.1
	Active travel (AT)	40	0.4
Car travel (CT)	Public transport (PT)	115	1.1
	Car travel (CT)	7731	72.2
	Active travel (AT)	140	1.3
Active travel (AT)	Public transport (PT)	41	0.4
	Car travel (CT)	192	1.8
	Active travel (AT)	1165	10.9
Total		10704	100

Table 4-4 Descriptive statistics of the factors driving the travel mode changes.

Variables	Changes in commute mode (%)			
	Switched to PT from CT & AT	Switched to CT from PT & AT	Switched to AT from PT & CT	No change
Life events				
Changes in car ownership				
Household acquired car	4.0	25.3	4.8	8.5
Household relinquished car	13.1	6.9	16.0	7.1
No change in car ownership	82.9	67.8	79.2	84.4
Household moved house				
Moved at local level	2.3	6.6	2.8	3.0
Moved at regional level	4.3	1.0	2.5	0.7
Moved at national level	2.9	5.2	0.6	0.5
Stayed	90.4	87.2	94.0	95.8
Householder changed job				
Yes	17.6	22.0	17.0	16.3
No	82.4	78.0	83.0	83.7
Travel characteristics				
Changes in travel distance				
Travel distance increased	74.8	63.2	0.5	23.7
Travel distance decreased	13.2	25.5	93.3	22.1
No change in travel distance	12.0	11.2	6.2	54.1
Number of observations	156	309	180	10059

PT-Public Transport, CT-Car Travel and AT-Active Travel

4.3 Model development

4.3.1 Model structure

Household car ownership and commute more choice behaviour consist of multiple directions of switching such as switching from non-car to car ownership, one car to multiple cars ownership, etc. in case of car ownership and switching from car to public transport, car to active travel, etc. in case of commute mode. This study attempts to capture all possible dimensions of switching in a single model. Details of the designing of alternatives are presents in the next section.

Similar to residential mobility behaviour, changes in car ownership and commute mode may also have dynamic effect. Behavioural dynamics are aimed to capture here by investigating the changes in car ownership and commute mode choice behaviours due to changes in the household conditions or circumstances over time. Discrete choice model is used, however, the limitation of this model to capture the duration dynamics or time dependency of the choice is acknowledged. Although, the hazard-based model is more flexible for capturing duration dynamics, it is less flexible for testing time varying covariates (more detail is in Chapter 2).

Multinomial logit (MNL) and mixed multinomial logit (MMNL) techniques from the family of the discrete choice theory are considered for modelling both car ownership change and commute mode transition behaviours. Details about the MNL and MMNL modelling techniques are presented in section 3.3.1 of chapter 3. Changes in household demographic characteristics, life events and travel characteristics are used as explanatory variables to observe changes in car ownership and commute mode in two consecutive years. To capture the role of the geographical scale of residential mobility on these choices, residential mobility parameters are used as explanatory variables. However, the use of residential mobility decision for explaining car ownership change or travel mode switching behaviours can induce endogeneity bias in the estimated parameters because residential mobility as an independent variable is likely to be correlated with the unobserved factors. For instance, suppose a household has moved to a new location for extra parking spaces and parking availability has opened the scope of buying an extra car. If the information of parking facility is not available in the dataset, the unobserved utility of both residential relocation and car ownership change can be correlated. In this case, residential relocation as an independent variable for explaining household car ownership change behaviour will be endogenous. Similarly, changes in car ownership as an independent variable in the travel mode switching behaviour can also be endogenous. Not only the choice as an independent variable, but any explanatory variables can also be endogenous if they are correlated with the unobserved utility (Guevara, 2010). Therefore, endogeneity in unavoidable in many of the cases (Guevara, 2015). There are several techniques have been proposed in the literature for correcting endogeneity bias (details are presented in chapter 2). However, this study is limited to deal with this issue.[10] Several studies in the literature have considered residential mobility as an independent variable for explaining household car ownership and travel mode switching behaviour where the endogeneity issue has been ignored (e.g. Hensher and Taylor, 1983; Clark et al., 2016a; Clark et al., 2016b).

4.3.2 Design of choice alternatives and individual choice set

4.3.2.1 Car ownership change model

Four levels of car ownership (having no car, one car, two cars and three cars) are observed in the dataset. Therefore, possible dimensions of switching from one level to another level are 16 (4×4). In the data, the number of observations in several directions of switching is very few specifically switching from zero to two or three cars, one to three cars and in the opposite directions (Table 4-1). Moreover, this study aims to capture the sensitivity differences between the first car and additional cars (second or third cars) in terms of gaining and losing. The sensitivity of switching from one to two cars is assumed the same as the sensitivity of switching from one to three cars or two cars to three cars. Therefore, the universal choice set consists of seven alternatives which are presented in Table 4-5 below.

[10] Joint estimation of the residential mobility decision and associated changes in the car ownership and travel mode can help for avoiding this kind of endogeneity. On the other hand, investigating the car ownership and travel mode changing behaviours for the households who moved in different geographical scales using appropriate nesting structure can also be useful. However, none of these approaches were feasible due to small number of observations for each choice direction.

Table 4-5 Universal choice set of car ownership changes

Alternatives in the universal choice set	Switching pairs under each alternative
Gaining car (s)	
Gaining first car (0-1)	0-1, 0-2,0-3
Gaining additional cars (1-2)	1-2,1-3,2-3
Losing car (s)	
Losing first car (1-0)	1-0,2-0,3-0
Losing additional cars (2-1)	2-1,3-1,3-2
Not gaining or losing car	
Zero car to zero car (0-0)	0-0
One car to one car (1-1)	1-1
Two cars to two cars (2-2)	2-2, 3-3

* Each number indicates the number of cars in the household

The choice set of households consist of a subset of alternatives based on the car ownership level at time t. For example, the choice set of a household that owns one car at time t contains the alternatives of gaining additional cars (1-2), losing first car (1-0) and remain in the same status (1-1) at the time (t+1). The choice set of households based on the car ownership level at time t are presented in Table 4-6.

Table 4-6 Choice set of the individual had different car ownership levels at time t

Car ownership level at time t	Choice set at time (t+1)
0 car	0-0, 0-1
1 car	1-1,1-2,1-0
2+ cars	2-2,2-1

4.3.2.2 Commute mode change model

The most commonly used commute modes reported by the respondents in the BHPS data are Rail/train, underground/tube, bus/coach, car, cycle and walk. The possible dimensions of switching explode with the number of alternative modes available. Since the number of switching is very small for some pairs, it is infeasible to capture all possible directions of switching. Therefore, the alternatives are grouped into public transport (PT), car travel (CT) and active travel (AT). Then the total number of alternatives is reduced to 9 (3×3). The individual choice set consists of a subset of them depending on the travel mode in the year t. Household-specific choice sets are presented in Table 4-7.

Table 4-7 Household-specific choice set based on travel mode at time t

Travel mode at time t	Choice set at time (t+1)
CT	CT-CT, CT-PT, CT-AT
PT	PT-PT, PT-CT, PT-AT
AT	AT-AT, AT-CT, AT-PT

4.4 Estimation results

This section presents the estimation results of household car ownership change and commute mode transition behaviours. The models focus on capturing the role of residential mobility decision on changes in car ownership level and commute mode, however, the reverse direction of causality is neglected (discussed in chapter 1 in detail). Moreover, the sequential estimation did not allow to capture the decision simultaneity leading to under/overestimate the correlations among the decisions neglecting the inherent trade-offs (Habib and Kockelman, 2008). The existence of random taste heterogeneities in the observed and unobserved components of the utility are tested using the MMNL technique. Correlation across the alternatives are also investigated using different nesting structures and Cholesky decomposition through the error components. Model parameters are estimated using R[11]. The estimation results are discussed in the following sections.

4.4.1 Modelling car ownership changes

Models are estimated to explore the factors driving the changes of car ownership level in two consecutive years with a special focus on the impact of the geographical scales of residential mobility on car ownership level changes. An MMNL model is estimated to capture the taste heterogeneities and potential correlation structures. The goodness of fit of the MMNL model is then compared with the MNL model using the likelihood ratio (LR) test. The chi-square statistic rejects the null hypothesis of the MNL model at 99.9% confidence interval. The MMNL model also captures a significant level of heterogeneities in the unobserved component. Different nesting structures and lower triangular matrix of Cholesky decomposition have considered to capture the correlations across the alternatives. The model with the Cholesky decomposition has shown a small improvement in terms of goodness of fit compared to the heteroscedastic model (not highly significant) presented in Table 4-8, but the other models to capture the correlation across the alternative have shown poor fit (details are given in Appendix F). Models are estimated without residential mobility parameters to investigate their contribution to the model fit. The chi-square statistic indicates that the model without residential mobility parameters is significantly worse

---

[11] The R codes from the Choice Modelling Center, University of Leeds has been used for estimating the model parameters. However, these codes are modified according to the model specifications and needs.

(LR=55.21, Chi-square stat=32.91 degree of freedom=12, confidence interval = 99.9%). The estimated model without the residential mobility parameters is presented in Appendix G. Estimated results are presented in Table 4-8 and discussed in the next sections based on the MMNL outcomes.

Household socio-demographic characteristics, dwelling characteristics, life events and travel characteristics are considered in the model as independent variables.[12] Changes in household state (sociodemographic, life event and travel behaviour are added) are also added as dummy variables to capture dynamics in the life course (also used in literature, for example, Oakil et al., 2014; Clark et al., 2016a; Clark et al., 2016b; Fatmi and Habib, 2017; Lin et al., 2018). In terms of investigating the changes in the household state on the behaviour, few directions of changes in household state are found to have an insignificant impact, therefore, removed from the final model (i.e. the impact of an increase in household income on the decrease in the car ownership level). More details are given in Appendix G.

From the estimation results (Table 4-8), household income is found to have a strong influence on car ownership level changes. High-income people are more likely to own a second (or third) car while unlikely to relinquish a car. This finding is in agreement with previous studies (e.g. Clark et al., 2016a). An increase in household income also increases the likelihood of acquiring an additional car and a decrease in household income significantly increases the likelihood of disposing of a car. In terms of the number of members in the household, inspired by literature (e.g. Krizek, 2003; Clark, Chatterjee and Melia, 2016a; Fatmi and Habib, 2017), three variables are used to capture this effect: household size, increase in household size and decrease in household size. The size of the household is found to have a positive impact on gaining cars as expected (see Hocherman et al., 1983; Clark et al., 2016a for a similar result). The household size is also found to have a positive effect on losing the first car. Lin et al. (2018) also observed a similar result. This is maybe due to the fact that the households lost the only member having a driving licence or maintaining a car becomes unaffordable due to the large household size.

[12] Few parameters such as household size, number of employees in the household and household income can be correlated. The risk of multicollinearity of the independent variables in the model are acknowledged.

Table 4-8 Estimation results of car ownership change model

Parameters	MNL								MMNL							
	Gained car				Lost car				Gained car				Lost car			
	0 to 1		1 to 2		1 to 0		2 to 1		0 to 1		1 to 2		1 to 0		2 to 1	
	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat
Alternative specific constants (no changes in car ownership is the base alternative)																
Mean	-3.4499	-20.2	-4.2505	-39.4	-4.2336	-25.7	-1.4255	-11.6	-4.2721	-14.2	-5.7862	-26.5	-5.2464	-18.7	-1.3025	-8.6
Standard deviation	-	-	-	-	-	-	-	-	1.8227	11.2	1.9966	20.7	-1.7297	-11.9	-0.7793	-11.2
Household level characteristics																
Household income	-0.0107	-1.7	0.0219	9.8	-0.0143	-2.4	-0.014	-6.8	-0.0006	-0.1	0.0358	9.9	-0.0127	-1.7	-0.0133	-5.7
Change in household income (base is no change)																
Income increased	0.0544	0.4	0.4869	7.3					0.0733	0.5	0.5896	6.9	-	-	-	-
Income decreased					0.3869	2.8	0.3311	4.1	-	-	-	-	0.3401	2.1	0.3134	3.6
Household size	0.3576	7.5	0.2216	8.4	0.1864	4.3	-0.07	-2.1	0.4860	5.8	0.3813	7.3	0.1333	2.0	-0.1112	-2.6
Change in household size (base is no change)																
Household size increased	1.7262	9.1	1.1199	10.3					1.9219	7.1	1.2508	8.5	-	-	-	-
Household size decreased					1.5186	8.8	1.8033	15.8	-	-	-	-	1.6497	7.6	1.9587	15.3
No of employees in the household	0.2488	3.0	0.6902	16.5	-0.1245	-1.5	-0.0585	-1.3	0.4053	3.3	1.0301	14.4	-0.1148	-1.0	-0.1012	-1.9
Change in number of employment (base is no change)																
Number of employment increased	0.9139	5.5	1.027	11.3					1.0696	5.0	1.2502	10.3	-	-	-	-
Number of employment decreased					0.7247	4.2	0.6021	5.9	-	-	-	-	0.6938	3.3	0.6415	5.9
Presence of senior adults	-0.8397	-4.2	-0.6918	-4.0	0.5937	4.1	-0.066	-0.4	-0.8817	-3.2	-1.3171	-4.9	1.0062	4.9	-0.1117	-0.5
Less educated people (below O level)	-0.4126	-3.2	-0.2364	-3.2	-0.0028	0.0	0.342	4.4	-0.6298	-2.7	-0.4953	-3.1	0.3008	1.6	0.4103	3.9

Table 4-8 Estimation results of car ownership change model (cont.)

Parameters	MNL								MMNL							
	Gained car				Lost car				Gained car				Lost car			
	0 to 1		1 to 2		1 to 0		2 to 1		0 to 1		1 to 2		1 to 0		2 to 1	
	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat
Dwelling characteristics																
Tenure type (base is owned house)																
Rented social housing	-0.5488	-4.0	-0.5893	-4.5	1.1362	8.7	0.7345	4.6	-0.7471	-3.1	-0.7720	-3.2	1.6463	7.4	0.9518	4.5
Rented private housing	0.1709	0.9	-0.5665	-3.7	0.9045	5.2	0.249	1.3	0.1908	0.6	-0.8275	-3.4	1.1186	4.3	0.2570	1.1
Life course events																
Moved house																
Moved at local level	0.2119	0.8	0.3047	1.8	0.8322	3.7	0.503	2.6	0.0432	0.1	0.4545	2.1	0.9578	3.4	0.5401	2.5
Moved at regional level	0.1085	0.2	0.8607	3.7	-0.4135	-1.0	-0.2202	-0.6	0.5232	0.8	0.6885	2.0	-0.3130	-0.5	-0.1602	-0.4
Moved at national level	1.6709	4.3	0.9268	3.5	0.9848	2.9	0.1456	0.4	2.1839	4.1	0.9818	2.3	0.8732	1.4	0.1844	0.5
Householder changed employer	0.0298	0.2	0.1016	1.2	0.0712	0.4	0.0191	0.2	-0.1082	-0.5	0.1380	1.2	0.0584	0.3	0.0068	0.1
Travel characteristics																
Travel distance	0.0123	3.2	-0.0116	-5.1	0.0055	1.3	-0.003	-1.4	0.0184	3.3	-0.0060	-1.8	0.0039	0.7	-0.0026	-1.1
Change in travel distance (base is no change)																
Travel distance increased	0.6789	3.1	-0.0543	-0.6					0.3761	1.3	0.0992	0.8	-	-	-	-
Travel distance decreased					-0.1818	-0.8	0.2132	1.9	-	-	-	-	-0.0498	-0.2	0.1763	1.5
Measures of model fit																
Number of observations	24718.000								24718.000							
Initial LL	-21985.800								-21985.800							
Final LL	-8374.2800								-7852.6300							
Likelihood ratio test	-								1043.3000							
Chi-square stat (4,0.001)	-								18.4670							

The variable ‘Change in household size’ captures the likelihood of gaining or losing car(s) due to a recent increase or decrease in the number of members in the household (e.g. childbirth, death, marriage, divorce, etc.)[13]. An increase in the size of the household in the following year is found to increase the propensity of gaining car(s) and a decrease in household size increases the probability of reducing the number of car(s) (also observed by Oakil et al., 2014; Clark et al., 2016a; Lin et al., 2018). Similarly, the effect of the number of employed people in the household is captured by three variables, the latter two capturing the change in the number of employed people in the immediate past. As seen in Table 4-8, the number of employed people in the household significantly influences the gain in household car ownership level. The likelihood of gaining a car increases when a household member gets a job and, similarly, the likelihood of losing a car increases if a household member loses her/his job. Clark et al. (2016a) found a similar result. The presence of senior adult(s) decreases the likelihood of gaining and increases the chance of losing car(s). Less educated people have a lower propensity to gain car and a higher propensity of losing it. Although Clark et al. (2016a) and, Lin et al. (2018) found a similar result, Oakil et al. (2014) observed that the association between education level and car transaction is statistically insignificant. Households living in rented social housing facilities are found to have a lower propensity of acquiring car(s) and the higher propensity of relinquishing car(s) compared to the households living in owned houses.

Importantly, the changes in car ownership levels of households are found to be significantly associated with the residential relocation behaviour and the associated geographical scale. In literature, conflicting outcomes are observed where Oakil et al. (2014) found a strong association between residential relocation and car ownership level change but Clark et al. (2016a) observed week association between them. The propensity of owning the first car is found to be significant for the households who moved at the national level and insignificant for the other two groups. Moving to a different metropolitan area can adversely affect household accessibility to public transport and other facilities which may increase the propensity to own a car. The likelihood of gaining an additional car is found significant for the households that

[13] The variables “Household size” and “Change in household size” provide different insights with the latter capturing the dynamic effect of gaining or losing cars due to adding or losing a new member in the family in the recent year.

moved at the local, regional and national levels. However, the likelihood of losing cars are found significant only among the local level movers. The association between job changing and changes in car ownership level is found insignificant. Householders that reported a longer daily commute are more likely to buy their first car but unlikely to buy additional car(s). A change in commute distance is not found significantly correlated with the household car ownership change.

4.4.2 Modelling commute mode changes

Switching travel mode in two consecutive years is modelled in this section to investigate the factors driving household commute behaviour changes. One of the core aims is to look at the influences of geographical scales of residential mobility on mode choice behaviour along with other drivers such as car ownership change, travel distance, etc[14]. Both the MNL and the MMNL models are estimated here. The MMNL models are allowed for randomness in the unobserved component to capture inter and intra respondent heterogeneity. Potential correlation across the alternatives are also investigated using different nesting structures and Cholesky decomposition. Although the model with Cholesky decomposition gave a small improvement compare to the heteroscedastic model presented in Table 4-9, it requires to estimate a larger number of random parameters resulting substantial increase in the estimation time. To be consistent with the residential mobility and car ownership change model, the heteroscedastic model is included in the main chapter. Details of other models are presented in Appendix H.

The goodness of fit of the MMNL model over the MNL model is investigated using the likelihood ratio test and the chi-square statistic rejects the null hypothesis of the MNL model at 99.9% confidence interval. Similar to the car ownership models, the impact of residential mobility parameters on the goodness of fit of the models are also tested here. The model without residential mobility parameters is found worse compared to the model with residential mobility parameters (LR=30.62, Chi-square stat=27.88, degree of freedom=9, confidence interval=99.9 %) (see Appendix I for more details). Estimated results are presented in Table 4-9 and results are discussed in the following section based on the MMNL outcomes.

---

[14] This study acknowledges the multicollinearity issue because some of the independent parameters can be correlated to each other such as residential mobility scale and travel time.

As observed in Table 4-9, households have significant levels of inertia to switch from one type of mode to another type (see Clark et al., 2016b for a similar result). Across the possible dimensions of switching, all else being equal, moving from public transport to car travel is found to be the least preferred option. Car ownership is found to have a strong association with travel mode change. Households that own cars are more likely to switch from public transport and active mode to car travel and unlikely to switch in other directions (switching from car to PT and AT). The likelihood of moving from PT and AT to car further increases if the household has gained a car in the preceding year. Losing a car in a given year, on the other hand, makes people more likely to switch to public transport or active travel in the following year (see Oakil et al., 2011; Idris et al., 2015 for a similar result).

The commute mode switching behaviour of the head of the households that have relocated is found to be significant in some but not all cases. In particular, estimation results indicate that local level relocation does not result statistically significant change in the likelihood to switch modes. This is probably due to the fact that moving in the same neighbourhood is unlikely to affect household transport circumstances (transport accessibility, commute distance), therefore households are found to use the same commute mode after a local level relocation. Estimation results indicate that regional level movers are more inclined to shift to public transport from car and active travel modes. This may be indirectly related to the fact that while making a regional move, households tend to move to an area with good public transport accessibility and consequently, there is an increase in the likelihood of using public transport. In case of relocations at the national level, there is a significant increase in switches both to car and public transport. This may be due to more significant changes in transport and work accessibilities after national level relocation. To the best of my knowledge, the role of the geographical scale on travel mode change behaviour is ignored in the current literature, however, a strong association between the relocation regardless of its scale and travel mode switching has been observed (e.g. Oakil et al., 2011; Klinger and Lanzendorf, 2016; Oakil, 2013; Clark et al., 2016b; Fatmi and Habib, 2017).

The connection between the change of employer and changing travel mode is found statistically insignificant however Oakil et al. (2011) found a strong connection between them. Travel distance is found to have a strong association with the travel mode changing behaviour for switches to public transport and active travel.

Table 4-9 Estimation results of commute mode switching model

Parameters	MNL						MMNL					
	Switched to CT (from PT & AT)		Switched to PT (from CT & AT)		Switched to AT (from CT & PT)		Switched to CT (from PT & AT)		Switched to PT (from CT & AT)		Switched to AT (from CT & PT)	
	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat
Alternative specific constants												
(no changes in travel mode is the base alternative)												
Mean												
Switched from car travel (CT)	-	-	-4.0482	-10.2	-4.6473	-7.7	-	-	-6.0397	-7.4	-5.8760	-7.1
Switched from public transport (PT)	-5.8208	-18.5	-	-	-3.5541	-9.3	-7.3598	-13.9	-	-	-3.8629	-8.7
Switched from active travel (AT)	-4.6788	-18.3	-4.7723	-19.9	-	-	-6.9939	-7.8	-5.7866	-16.1	-	-
Standard deviation												
Switched from car	-	-	-	-	-	-	-	-	2.2887	7.1	1.6919	6.9
Switched from public transport	-	-	-	-	-	-	2.6795	8.3	-	-	0.1265	0.1
Switched from active travel	-	-	-	-	-	-	2.6628	4.1	1.1433	5.5	-	-
Household owns car	1.3811	8.0	-2.5991*	-7.3	-0.6146*	-1.1	1.9458	7.3	-2.6965*	-4.5	-0.6512*	-0.9
Changes in car ownership												
Household acquired car	1.8772	9.1	-2.2714*	-3.3	-0.7374*	-1.6	2.4394	8.5	-2.8584*	-3.3	-0.9686*	-1.7
Household relinquished car	0.2288	0.8	0.6783*	2.5	0.9615*	4.0	0.4306	1.1	0.8842*	2.6	1.2170*	3.9
Moved house												
Moved at local level	-0.0565	-0.2	-0.3512	-0.7	-0.1925	-0.4	-0.0936	-0.2	-0.2974	-0.5	-0.3640	-0.7
Moved at regional level	0.6508	0.8	1.6592	3.7	1.1208	2.1	0.2378	0.2	2.3575	3.7	1.0994	1.4
Moved at national level	1.3245	3.3	1.6634	3.0	-0.1283	-0.1	1.6951	2.5	1.5369	1.9	-0.1043	-0.1
Householder changed employer	0.3323	1.9	-0.0739	-0.3	-0.0415	-0.2	0.1802	0.8	-0.0931	-0.3	0.0024	0.0
Travel distance	-0.007	-1.5	0.0203	5.9	-0.0696	-8.5	-0.0033	-0.4	0.0291	5.2	-0.0669	-7.7
Changes in travel distance												
Travel distance increased	3.2857	15.7	3.0336	11.7	-1.9716	-1.9	4.0742	14.5	3.3362	10.3	-1.8339	-1.6
Travel distance decreased	3.0335	12.5	1.1371	3.3	3.6893	12.0	3.4740	10.5	1.2165	3.0	3.9887	11.2
Measures of model fit												
Number of observations	10704						10704					
Initial LL	-11759.5						-11759.5					
Final LL	-1901.0						-1784.3					
Likelihood ratio test	-						233.4					
Chi-square stat (6,0.001)	-						22.4580					

* parameters represent switching from car travel only

This may be due to the fact that driving a long distance regularly increases the anxiety level and adversely affects personal stress level and work efficiency; consequently, car is a less preferred option for the long-distance commuters. The effect of increase or decrease in travel distance, however, has a larger and more significant effect. An increase in travel distance makes people more likely to switch to public transport and car while a decrease in commute distance results significant increase in the probability to shift to active travel (Clark et al., 2016b also observed similar relation). Some other sociodemographic characteristics such as income, education level, household size, etc. have also been tested but found to have an insignificant effect and hence dropped out from the final model.

4.5 Validation results

In the above sections, it is observed that the MMNL models outperform over the MNL counterparts. Validations are done here using holdout samples to investigate whether the gains in goodness-of-fit of the MMNL models remain in the prediction context. Therefore, same as the validation technique used in chapter 3, the dataset is randomly divided where 60% of households is considered for estimation (who are consistently available in the panel) and rest 40% of the households is considered for validation (also used by other researchers: de Luca and Cantarella, 2016; Bwambale et al., 2017 for example). The same procedure is repeated for three times to check whether the performance is consistent over the different split of the dataset based on different independent random draws. Both car ownership changes and commute mode switching models are estimated using the estimation subsets of data from three independent random draws. Interpretation of the model parameters estimated using the corresponding subsamples of the data remains the same as the parameters estimated using the full dataset (discussed in the previous section). The goodness of fit of the models estimated using the estimation subset of data are presented in Table 4-10.

The predictive power of the estimated models (both car ownership changes and commute mode switching) are evaluated in terms of improvement in goodness-of-fit (log-likelihood in prediction sample and predictive rho-square) in the validation subset of data. The results are presented in Table 4-11. It is observed that the MMNL models perform better than the corresponding MNL models in the estimation sample and holds consistent performance in the validation sample in both cases.

Table 4-10 Goodness of fit of the car ownership change and travel mode change models estimated using estimation subset of data

Models	Draws	Number of observations	Initial log likelihood	Final log likelihood		Adjusted rho-square	
				MNL	MMNL	MNL	MMNL
Changes in car ownership level	D1	14824	-13123.9	-5334.3	-5004.7	0.588	0.613
	D2	14824	-13091.3	-4845.8	-4538.1	0.625	0.648
	D3	14824	-13104.9	-5119.6	-4789.8	0.604	0.629
Changes in travel mode	D1	6418	-7050.9	-1177.2	-1122.6	0.828	0.835
	D2	6416	-7048.7	-1156.0	-1095.9	0.831	0.839
	D3	6416	-7048.7	-1143.9	-1085.9	0.833	0.840

Table 4-11 Validation results of car ownership change and travel mode change models

Models	Draws	Number of observations	Initial log likelihood	Final log likelihood		Predictive rho-square	
				MNL	MMNL	MNL	MMNL
Changes in car ownership level	D1	9894	-8659.2	-3100.7	-2909.4	0.634	0.656
	D2	9894	-8688.8	-3580.7	-3377.0	0.580	0.603
	D3	9894	-8676.8	-3309.1	-3134.4	0.611	0.630
Changes in travel mode	D1	4286	-4708.7	-754.4	-703.1	0.832	0.842
	D2	4288	-4710.8	-772.2	-744.6	0.828	0.833
	D3	4288	-4710.8	-790.2	-771.3	0.825	0.827

For further demonstration of the performance of the developed models in the forecasting purpose, the model performance is compared in prediction of future years. For this purpose, the data from waves 1-14 is used for estimation and applies the model estimates for predicting the decisions made in the last three years (waves 15-17). The results are presented in Table 4-12.

Table 4-12 Performance of the models in forecasting

Models	Number of observations	Initial log likelihood	Final log likelihood		Predictive rho-square	
			MNL	MMNL	MNL	MMNL
Estimation sample (waves 1 to 14)						
Changes in car ownership level	20356	-18315.2	-7047.0	-6611.8	0.611	0.635
Changes in travel mode	9254	-10166.6	-1662.0	-1492.3	0.833	0.849
Validation sample (waves 15 to 17)						
Changes in car ownership level	4362	-3828.0	-1444.7	-1402.7	0.605	0.615
Changes in travel mode	1450	-1593.0	-307.6	-297.1	0.784	0.787

As seen in Table 4-12, the models of car ownership and commute mode choices perform well in terms of forecasting the decisions of the last three waves. The MMNL models show better fit in prediction than the MNL counterparts. The forecasting results indicate that capturing the panel effect is important for modelling car ownership change and commute mode switching behaviours.

4.6 Conclusions and policy recommendations

Models are estimated in this chapter to investigate the relative impact of residential decision and other factors on the changes in car ownership levels (i.e. increase or decrease in the number of cars) and travel mode transition (i.e. shifts between car, public transport and active travel modes). The key findings are as follows:

- Geographical scales of residential mobility lead to differences in car ownership level changes. Estimation results indicate that household car ownership level changes between two consecutive years are significantly affected by residential mobility decisions (i.e. do not relocate, relocate locally, regionally or nationally) alongside the household sociodemographic characteristics and dwelling characteristics. For example, households who have moved locally are found to be less likely to gain a car whereas households who have moved at the regional or national level are found to be more inclined to acquire a car.
- Household travel mode choice is also found to be significantly affected by the geographical scale of relocation. Households who have moved at the national level are more likely to switch to car while households who have moved at the regional level are more inclined to switch to public transport from car and active travel. Local movers, on the other hand, are observed to have higher inertia and lower probabilities of the mode switch.
- Household travel mode choice is also found to be significantly affected by changes in car ownership levels. For instance, an increase in car ownership is found to increase the propensity of shifting to car and decrease the probability of switching to other modes.

Similar to other empirical studies, this work also has several limitations. The model components such as residential mobility, car ownership change and travel mode switching are estimated sequentially. This sequential estimation may under or overestimate the correlations among the decisions neglecting the inherent trade-offs and simultaneity in choice. However, the nature of the decisions (rare events) does not offer flexibility for simultaneous or joint estimation to capture the reciprocal interaction of decision components because the dataset does not have a representative number of observations for combinations of different choices (households moved in different scales, changed car ownership in different levels and changed travel mode in different directions), a point. For instance, combined choice of household moved at regional level, gained car(s) and switched from public transport to car does not have any observation. Future work should aim for the simultaneous estimation of the interconnected decision components if a suitable dataset is available.

Discrete choice models are estimated in this study. To capture the dynamics in the behaviour, time varying covariates (the effects of changes in contributing factors on the choice) are used and the state-dependency of the decisions are also tested (added in Appendix E). However, the discrete choice technique was less flexible for capturing the duration dynamics in the behaviour. Future work should target for analysing these time dependent household behaviours using the more appropriate technique that allows for capturing the duration dynamics in the behaviour (e.g. hazard-based model or Markov chain model).

This study captured the influence of household residential decision on their car ownership and travel mode switching behaviours. However, these relations may have reverse causalities (e.g. car ownership change can influence residential decision, although, these causality directions are less likely to be critical) that have not been investigated in this research. This issue can also be addressed in the future study.

The number of households consistently observed in the full survey (waves 1 to 18) is limited and a small percentage of households in the dataset changed their car ownership and travel mode in two consecutive years. Therefore, it was not possible to investigate the changes across all possible directions (alternatives are grouped here like public transport, active travel, etc.). Use of larger dataset (for example understanding society) could be useful to achieve this and even for more robust analysis.

Further, neighbourhood characteristics such as public transport accessibility, parking availability, land use patterns are likely to influence household car ownership changes and travel mode transition behaviour. These parameters are not available in the dataset and cannot be tested in the current models.

A key goal of sustainable transport system development is to reduce car dependency and to promote active travel or use of mass transit. From the policy point of view, analysts or policymakers may need to understand the driving forces of car ownership change and travel mode switching behaviours. The estimated models in this chapter can be a guideline in this context. In addition, metropolitan cities in the UK have different rates of intra and inter-regional residential mobility. This study findings will also help for understanding how differently internal mobility and mobility from other cities affect the aggregate level car ownership and commute behaviour of the cities leading to the differences in the policy formulation. However, a few more specific policy guidelines based on the findings from this chapter are further discussed below.

Buying a new car is found to push people for switching from public transport and active travel to car. The driving forces of acquiring a new car is explained by the outcomes of the car ownership change model that can be used as a guideline for policy steps to reduce car ownership and car use. People are more likely to change commute mode when they move home. For instance, national-level movers are more likely to switch to car. This finding is also supported by the findings from the car ownership change model that the national level movers are more likely to own car(s). The main factor for car dependency on the national level moves can be the poor public transport accessibility in the area they moved or they work, insufficient housing supply close to the work station, etc. In addition, new migrants are most likely to be unfamiliar with the housing and other facilities in the new area/city or may not have enough scope to find properties close to their workplaces. These circumstances can influence new migrants to buy a new car for commute and other trips. Policy to encourage employers to provide housing facilities for the new employees who migrated from other cities can minimize this problem. For sustainable transport system development, the policy should encourage people for nonmotorized trips resulting decrease in total VMT and emission. From the model outcome, it is observed that a decrease in the commute distance increases the likelihood of switching to active travel from motorized trips. Therefore, the policy should promote mixed land use development which may increase the chance of living close to the work locations and to consider active travel mode. However, shifting to car due to both increase and decrease in the commute distance may be a result of poor public transport accessibility in the home to work corridor. In this case, transit-oriented development can increase public transport dependency. However further study is required to find the more specific reason behind the switching to car from other transport options. Finally, this study contributes for understanding the changes in car ownership and travel mode in response to change in household state and experience. These behavioural issues can potentially be considered in integrated transport and land use modelling.

Chapter 5

Modelling residential location choice

5.1 Introduction

When households have enough reasons for residential move, they become active in the housing maker for finding a suitable location or neighbourhood (this connection between the decision to move and choice of location to move is shown in Figure 1-3). While the decision to move (or stay) and its geographical scale has been modelled in Chapter 3, this chapter and the next chapter (chapter 6) looks at the choice of location (Figure 5-1). A key behavioural issue in the residential location choice preference is investigated in this chapter (discussed in detail in the rest of the chapter) considering the full choice set for all. However, the next chapter focuses on modelling residential location choice capturing behaviourally persuasive choice set for individual instead of using the universal choice set for all as in this chapter (also discussed in Chapter 1).

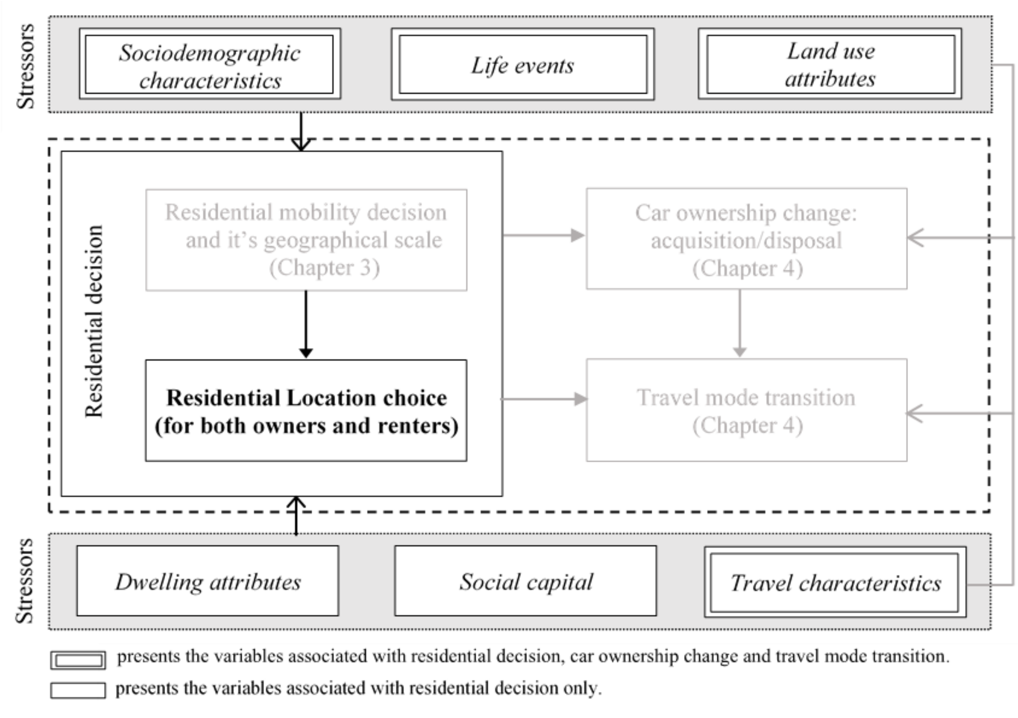


Figure 5-1 Component of the research framework investigated in this chapter and Chapter 6 (highlighted).

Investigating the residential location choice preferences can be useful for housing market analysis and to predict the implications of alternative policy scenarios that can help inform policy decisions. For example, the Willingness-to-pay (WTP)[15] values, which explain the trade-offs people make between cost and location (or dwelling) attributes, can be a valuable tool for the detailed planning of new developments. Investigating the residential location choice behaviour can also be useful for explaining the behaviour of car ownership change and travel mode transition (these connections are shown in the research framework). For instance, households who chose low-density residential areas to live tend to commute long distances for work and other trips and are typically more car-dependent (Naess, 2009; Alexander and Tomalty, 2002). However, people often prefer low-density areas as they offer more green and open spaces, larger homes, greater ease of parking, etc. (Masnavi, 2000). On the other hand, people who live in mixed and compact developments have better access to facilities and are typically less car-dependent (Masnavi, 2000; Brown and Werner, 2008; Ewing and Cervero, 2010; Farber and Li, 2013). Although a potential connection between the characteristics of the location households moved and the changes in their travel behaviour has been observed in the literature, this study cannot explore this relation due to data limitation that may reduce the model predictive power in forecasting. The number of households in the BHPS dataset (used in Chapter 3 and Chapter 4) who moved in a single housing market or city (e.g. London) was very few. Therefore, the BHPS dataset cannot be used for modelling residential location choice behaviour, instead, another dataset (London Household Survey data that is stated in section 1.3 of Chapter 1) is used for investigating the research question in this chapter[16] which is discussed in the next section.

Residential ownership and renting are the two major tenure groups in the housing market. The factors driving the residential location choices of these two groups are unlikely to be the same with possible differences in which attributes matter, and also differences in how much each attribute matters. The potential differences can be explained by the distinct nature of the ownership and renting decisions. Residential ownership is a long-term decision that involves huge investment and high relocation

---

[15] WTP, a widely used tool in marketing and environmental economics, is a measure of how much an individual is willing to pay to acquire desirable attributes and/or to avoid undesirable attributes of the alternatives. It has also been investigated in the context of residential location choice – although only in a relatively limited way (e.g. Jara Diaz et al., 1999; Small et al., 2012; etc.).

[16] LHSD cannot be used for investigating the research questions in Chapter 2 and 3 (residential mobility behaviour, the changes in car ownership and commute mode) because this dataset does not have information prior relocation.

costs while private renting is typically a medium to a short-term decision due to the higher level of flexibility associated with the lower relocation costs, shorter lengths of agreements, and other factors. For instance, the average tenure length in England is 11 years for owners but only around 1 year for private renters and 7 years for social renters (Randall, 2011). The socio-demographic characteristics of the consumers of the ownership and private renting markets are also typically different. For example, high and middle-income households are more likely to be able to afford to buy properties while others may be more likely to rent (Yates and Mackay, 2006). With these clear distinctions, it is important to analyse the ownership and renting choices in detail and identify the similarities and differences in sensitivities of these two groups to different factors. Although there have been some studies focusing on residential ownership (e.g. Bhat and Guo, 2004; Habib and Miller, 2009; Guevara, 2010; Zolfaghari, 2013) or renting decisions (e.g. Hoshino, 2011) in isolation or both (Ho, Hensher, & Ellison, 2017), to the best of our knowledge, none of these studies have quantified the differences in sensitivities towards different factors (or the systematic heterogeneity in elasticity and WTP values) between owners and renters. This study hence aims to address this gap by investigating the similarities and differences between residential location choices of owners and renters. This study objective is addressed by developing Revealed Preference (RP) based residential location choice models for people living in the Greater London Area (GLA)[17]. Several data sources are combined and make use of detailed econometric models to analyse the residential location decisions in those datasets. The findings of this study may have important policy guidelines that are discussed in the concluding section.

5.2 Data

5.2.1 Study area

The Greater London Area (GLA) is considered as a study area. The GLA is divided into 32 boroughs and the City of London. The total number of electoral wards before

[17] It may be noted that in GLA, 23% of the housing market constitutes of social renters who have a constrained choice set and are not able to exercise their residential choices in the same way as private renters. In order to capture the preferences of the social renters, it is critical to know the choice set of each decision maker (i.e. available alternatives during making the decision) and the associated constraints (arising from the allocation policy). However, the London Household Survey dataset did not include these pieces of information. Therefore, the social renters were excluded from our analyses and only focused on private renting (referred as 'renting' in the rest of the paper) and owning.

2002 was 773 where 286 were in Inner London, 462 were in Outer London and the rest were in the city of London. In 2002, the ward boundaries of the GLA were changed significantly and the majority of the wards were physically affected. The total number of wards was reduced to 649 after reshaping, where 221, 403 and 25 were in Inner, Outer and the city of London, respectively. A map view of Inner, Outer and the City of London is presented in Figure 5-1.

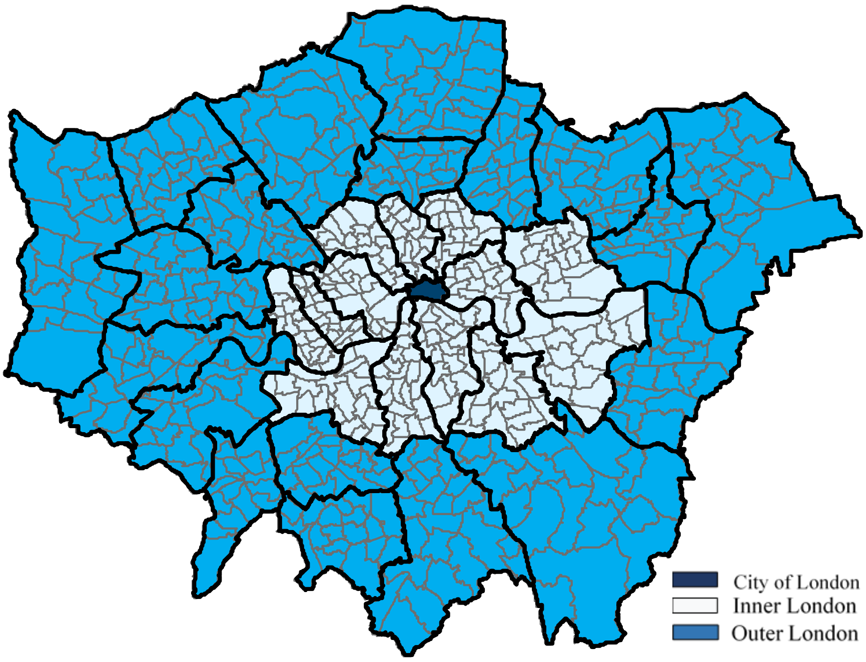


Figure 5-2 Map of Greater London Area. (Source: https://data.london.gov.uk/)

5.2.2 Data description

To estimate the residential location choice model, both household-level data (e.g. residential location, demographic characteristics, dwelling characteristics, travel behaviour) and location characteristics (e.g. land-use, transport accessibility, employment opportunity) are essential. However, in the context of London, no single dataset has all the information. Therefore, several datasets were used as summarised in Figure 5-2 and detailed below. Figure 5-2 shows how the inputs for the model come from three different sources. The dependent variable of the model is the chosen residential location, which comes from the London Household Survey data.

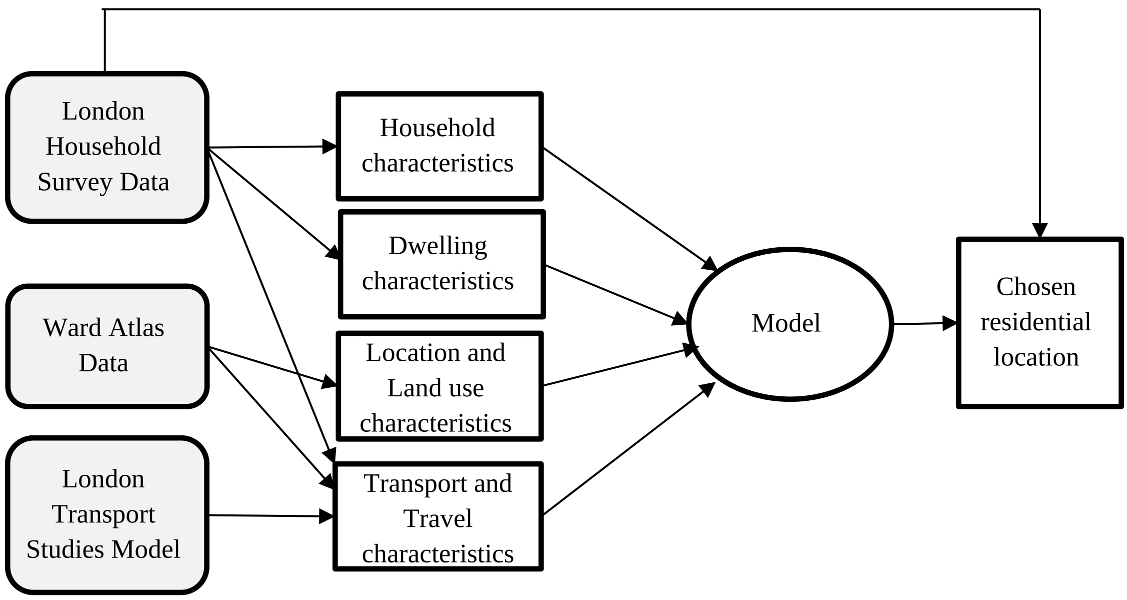


Figure 5-3 Sources of data used for deriving the factors affecting the residential location choices.

5.2.2.1 London household survey data (LHSD)

The LHSD serves as the main source of disaggregate level household and dwelling information used for model estimation. This data set was collected in 2002 and consists of detailed information (e.g. socio-demographic characteristics, dwelling information, employment status, home and work location, car ownership, etc.) of 8,158 households and 20,910 individuals from 498 wards in the GLA. Multistage stratified random sampling has been used in this dataset to ensure representative samples from the selected wards. The dataset contains information of 4,491 households living in their owned houses, 2,489 households living in houses rented from council or housing authorities, 1,087 households living in privately rented houses and 91 households living in shared accommodation. Since only a low number of households have very long tenure length, the households who moved between 1971 to 2002 are retained. This study focuses on households living in owned houses and privately rented houses and having at least one member commuting to work which left us with observations of 2,180 owners and 520 private renters.

5.2.2.2 Ward atlas data (WAD)

WAD includes ward level aggregated information of land use pattern, population density, household composition, ethnic proportion, employment and economic activity, household income, crime rates, land use, public transport accessibility, green space, car use, etc. Data for the year 2002 is used in this study as the source of location attributes used in the model.

5.2.2.3 London Transport Studies Model (LTSM)

Information about the distance of the alternatives from individual workplace and CBD[18] are missing in the LHSD files which are clearly of utmost importance as a determinant of household residential location. The origin-destination (OD) matrix of the GLA from the London Transport Studies Model (LTSM) is used to extract these distances. Since distances did not change for the areas between 2002-2011, distance data from 2011 can be used consistently for the year of 2002.

5.2.3 Data preparation

LHSD is used as the main source of household level information that includes household residential location, travel characteristics and sociodemographic characteristics, etc. Ward boundaries were used as the finest level of residential location identifiers in the LHSD. Ward boundaries in GLA were changed several times. A major change was made in the year 2002 when most ward boundaries were affected in different scales. The ward boundaries before the year 2002 were used in the LHSD (called as old ward boundary in rest of the chapter), ward boundaries after 2002 were used in the WAD (called as new ward boundary in rest of the chapter) and traffic analysis zones (TAZ) were used in the LTSM. Since the ward boundaries used in the datasets are different from each other, combining these three datasets was challenging.

With the help of GIS map matching, WAD and LTSM data are converted under equivalent old ward boundaries. The layer function in ArcGIS is used to investigate the physical changes of old and new ward boundaries and to convert the WAD to equivalent old boundaries. In some cases where old ward boundaries are found to be similar or wards formed part of a new ward area, attributes of the WAD file in new ward boundaries are kept the same for old ward boundaries. In other cases where the

[18] The city of London is considered as the CBD and the ward Cordwainer is considered as centre of the CBD. Although the CBD of London is changing over time and extending, the City of London is the oldest and major part of the London CBD

old ward area was found to be shared across multiple new ward areas, the weighted averages of shared new wards attributes are estimated for the corresponding old ward. Therefore, the attributes are assumed to be constant within each new ward. It is unlikely that this introduces substantial inaccuracies since the ward-level data is already based on an assumption of homogeneity. It may also be noted that a similar approach has been used in previous studies (e.g. Habib and Miller, 2009). TAZ boundaries are also converted to equivalent old ward boundaries with the help of ArcGIS. The centre to centre distances between the converted old ward boundaries are then considered to extract the distance between household work and home locations. However, the potential error introduced due to considering the centre to centre distance instead of the actual disaggregate level distance between work and home locations is acknowledged. After conversion, the three independent datasets are merged based on the old ward boundaries. The final dataset only consists of the information required for analysis and the rest of the information is discarded. More details of data conversion and joining of the datasets are presented in Appendix J.

5.2.4 Limitation of the dataset

Due to the unavailability of a more recent and suitable dataset to address the research question, relatively old data (collected in the year 2002) is used in this study. Although the housing market in the study area might have been changed over the years (which can hinder the direct use of the models for forecasting), the outcome of this study might give a useful indication whether the residential preferences of renters and owners are different and it stresses the need to treat them differently while predicting residential location choices.

This study uses the characteristics of alternatives and individual for the year of data collection (2002) to explain the choice made on and before. Although the ideal case is to explain the residential location choices based on the characteristics at the time point when households had moved, this is very challenging because land use and location attributes may not be available for every year. However, a household has the freedom to move houses at any given point of time (albeit there is likely to be a strong inertia effect). Therefore, it is not unreasonable to assume that if a household is in a certain location in 2002, it is still deemed as their most preferred location. In other words, the households are still getting the highest utilities from their current locations in spite of any potential changes in the attributes of the neighbourhood over time.

Hence, it is reasonable (though not ideal) to use the characteristics at the year of data collection to explain the choices made by the households in earlier years. In fact, because of these facts, previous researchers have also used data from a single year to model the residential location choice behaviour of households who have moved over a spell of time (e.g. Bhat and Guo, 2007; Pinjari et al., 2011; Sener et al., 2011; Zolfaghari, 2013). This approach may invite bias in the estimation result. Because over time some other places might be more attractive compared to the place households are currently living but they do not change for the inertia of living for a long time in the current place and associated cost of relocation.

5.2.5 Data representativeness

A multistage stratified random sampling technique has been used in the London Household Survey. Therefore, sampling weights are provided with the dataset to ensure that the weighted data is representative of the London population according to the 2001 census with regards to demographic variables (such as gender distribution, household type, and ethnic composition), economic variables (such as household income and employment status), housing tenure variable (such as owners, private renter, social renters), etc. (Greater London Authority, 2003). The supplied sampling weights are considered in all analyses in this chapter and the next chapter.

In this study, a subset of the data is considered that consists of the commuter households (households having at least one working member) who lived in either owned or privately rented houses. The data for non-commuter households and households living in socially rented houses is excluded. The characteristics of these two groups (commuter households and non-commuter households) are most likely to be different in the population. For example, the annual income of the households that do not have working member(s) is most likely to be low compare to the households having working member(s). Similarly, the non-commute householders are most likely to be retired elderly people or unemployed people, therefore the car ownership level of this group can be lower. Since the weighted full sample is representative of the London population, the distribution of the characteristics of the commute and non-commute households in the weighted full sample are likely to be different, potentially leading to differences in the characteristics of the households in the subsample and the full sample (see Appendix K for the detailed comparison). As mentioned before, the sampling weights provided with the dataset have been calculated ensuring the representativeness of the different sociodemographic classes (e.g. commuters-noncommuters, owners-renters) in the dataset to the London population. Therefore, the weighted subsample used for analysis in this chapter and the next chapter is representative of the corresponding group in the population.

5.2.6 Data analysis

Descriptive analysis of the weighted data reveals significant differences in location and dwelling attributes, travel behaviour and socio-demographic characteristics between owners and renters which are explained here and presented in Table 5-1. The split of owners in Inner and Outer London is quite different from renters (29.3:70.7 and 56.3:43.7 respectively). The average tenure length of owners is more than three times higher than renters. The percentages of owners and renters belonging to the high-income group (more than £60,000 per year) are 28.7% and 27.3% respectively. This agrees with previous studies in London which also report substantial portions of high-income people preferring to rent in Inner London due to excessive house price (Paccoud and Mace, 2018). For the lower-income group (less than £30000 per year), the corresponding shares are 51.0% and 39.4% for renting and ownership respectively. The average household size (number of members in the household) is found to be higher for owners than renters. The rate of car ownership for households living in their owned properties is more than 50% higher than households living in rented properties with a difference of around 20% more for properties owned in Outer London. Around 50% of households who live in owned properties are married couples whereas only 20.5% of households who live in rented properties belong to this group.

There are substantial differences in the commuting behaviour of owners and renters as well. Owners are more dependent on private car (32.1% in inner London) than renters (11.7% in inner London)– again the percentages varying largely between Inner and Outer London. Households living in Outer London are found to be more car-dependent whereas households living in Inner London are found to be more transit-oriented. Importantly, the average commute distance of owners both in Inner and Outer London are higher on average than that for renters. These differences serve as motivation and provide useful insights for the model specification that are presented in the following section.

Table 5-1 Descriptive statistics of LHSD

Variables	Tenure Group	
	Owners	Renters
Socio-demographic characteristics		
Annual household income		
Less than £30,000	39.4%	51.0%
Between £30,000 to £60,000	31.9%	21.7%
More than £60,000	28.7%	27.3%
Average household size (members in the household)	2.9	2.7
Household composition		
Married couple with and without kids	51.5%	20.5%
Cohabiting couple with and without kids	14.4%	17.4%
Single member household	24.7%	29.4%
Household having more than one member	9.4%	32.7%
Ethnic composition		
White people	79.5%	74.9%
Asian people	12.9%	14.2%
Black people	7.6%	10.9%
Employment status		
Households have at least one working member	98.9%	90.8%
Households do not have any working member	1.1%	9.2%
Location and dwelling features		
Residential location		
Inner London	29.3%	56.3%
Outer London	70.7%	43.7%
Average dwelling size (number of bedrooms)		
Inner London	2.5	2.4
Outer London	2.9	2.6
Average tenure length (in years)		
Inner London	8.8	2.0
Outer London	10.6	2.8
Travel behaviour		
Car ownership		
Inner London	76.0%	40.8%
Outer London	89.5%	66.4%
Travel mode		
Private car		
Inner London	32.1%	11.7%
Outer London	51.2%	34.7%
Public transport (bus, train, tube)		
Inner London	21.8%	31.9%
Outer London	17.6%	28.3%
Others (motorcycle, pedal cycle, walk, etc.)		
Inner London	46.1%	56.4%
Outer London	31.2%	36.9%
Average commute distance (in kilometre)		
Inner London	7.5	7.2
Outer London	11.1	7.9

5.3 Model development

5.3.1 Model structure

Models are estimated to investigate the preference of owners and renters to the residential location choice attributes and existence of attribute-level preference heterogeneities between these two groups.[19] Model parameters are estimated using the discrete choice analysis (DCA) technique. DCA is a widely used technique to analyse consumer choices in which the available options are discrete in nature and mutually exclusive. The estimation was started with the most basic version of a discrete choice model: a Multinomial logit (MNL) model. To capture random taste heterogeneity across individuals as well as differences in error variance between owners and renters, mixed multinomial logit models (MMNL) were estimated later.

A key decision in any study of residential location is the level of disaggregation. This study focused on zone level models where households are considered to choose a zone from a set of available alternative zones. It is assumed that all zones have properties available for renting and buying (a reasonable assumption), therefore, choice sets of both renters and owners can be the same and include all alternatives zones from the study area.[20] However, a zone having many vacant properties is expected to have a higher chance of being selected compare to the area having fewer vacancies. If the market has supply constraints, households might compromise between their preference and investment (housing cost or rent). Indeed, supply constraint (or higher demand) usually increases the housing cost and surplus decreases it (Zhou and Kockelman, 2008; Habib, 2009). Ignoring the supply parameter may increase the unexplained portion of the model and hence makes the price endogenous due to its correlation with the error term (Habib, 2009; Guevara, 2010). To minimize this problem, housing supply information (number of dwelling or number of empty

---

[19] The models developed here reflect the housing demand or housing allocation. However, the characteristics of the housing market are very complex which involves multiple agents in the supply and demand side. In literature, lots of urban simulation models have been proposed where the supply and demand in the housing market have been modelled explicitly (Waddell, 2002; Zhou and Kockelman, 2008; Habib, 2009; Farooq and Miller, 2012). Then the price is determined so that it matches demand and supply at a certain point for market equilibrium. This confirms the allocation of land ensuring that each household chooses their preferred home while developers and landowners maximize profits or rents.

[20] The zones household truly considered in their choice set is difficult to explore because this information is unknow to the analyst. Different techniques are available in the literature for capturing the behavioural choice set. Chapter 6 focused on this issue.

dwelling for selling and renting in each zone) has been used as an independent variable in the literature (Guo and Bhat, 2002). It has been observed that zones where more housing units are available, have a higher probability of being chosen. Due to the unavailability of housing supply information, the current study could not consider this housing supply attribute. Consideration of supply information may have a different level of impact on the preferences of owners and renters due to the differences in the shares of the properties available for selling and buying in each zone. However, chapter 6 focuses on capturing the differences in the choice set consideration of owners and renters based on their preferences. On the other hand, in the dwelling level model, households choose an alternative that offers the highest utility from the pool of available dwellings in the area households are anticipating to move to. This approach is a partial representation of dwelling supply in both ownership and renting markets (Habib, 2009). However, the application of this approach is limited in the literature due to a lack of dwelling supply data for many metropolitan cities like London, thus, a wider application of the zone level approach is observed in the literature (Bhat and Guo, 2004; Zondag and Pieters, 2005; Walker and Li, 2007; Chen et al., 2008; Pinjari et al., 2011; Sener et al., 2011). Moreover, the number of alternative households considered is likely to be constrained due to the dynamic nature of housing supply and household limited capacity for gathering and processing information (Fotheringham et al., 2000).

The modelling work is based on the principle of utility maximisation, assuming that decision makers choose the alternative that provides them with the greatest utility. The modelling work aims to explain the way individual decision makers choose between mutually exclusive alternatives by estimating the importance they place on the characteristics of these alternatives, where this potentially varies across individual decision makers. Of course, the actual process of preference formation is not observed by the analyst, and there thus remains a role for an error term in the models, capturing the various influences on decision making not explained by the analyst.

In the analysis, a number of potential key effects are incorporated, as follows:

- heterogeneity in preferences linked to observed characteristics, such as commute distance;
- differences in preferences between owners and renters;

- random (i.e. unexplained) variations in preferences between individual decision makers; and
- differences in the amount of error variance (i.e. unexplained influences on behaviour) for owners and renters.

The utility for zone j for household n is given by:

$$U_{nj} = \sum_{k=1}^{K^f} (\beta_k^f + \Delta_{kr} r_n) x_{njk}^f + \sum_{l=1}^{L^h} (\beta_{lno}^h o_n + \beta_{lnr}^h r_n) x_{njl}^h + \xi_{rn} r_n + \xi_{on} o_n + \varepsilon_{nj} \qquad (5\text{-}1)$$

Where, r_n is a dummy for renters (*1* if observation *n* corresponds to a renter, *0* otherwise), and o_n is a dummy for owners (*1* if observation *n* corresponds to an owner, *0* otherwise). β_k^f represents parameters which do not vary randomly across the households (i.e. are 'fixed' (*f*)) with a shift $\Delta_{kr} r_n$ for renters with respect to owners; x_{njk}^f are the corresponding attributes. β_{lno}^h and β_{lnr}^h represent parameters that follow a random distribution across the households (i.e. incorporating unobserved heterogeneity (*h*) in preferences), with separate groups for owners and renters ; x_{njl}^h are the corresponding attributes. $\xi_{rn} r_n$ and $\xi_{on} o_n$ are the renter and owner specific error terms, ε_{nj} is the type *I* extreme value error term, distributed randomly across individuals and across zones. The subscript *n* on the attributes relates to the fact that attributes are not just zone-specific but also household specific given the incorporating of the deterministic heterogeneity. For example, for cost, multiple parameters are estimated in the model, with different cost sensitivity for different income groups, and only one of these is used for any given household, with the associated cost attribute set to zero for any income levels that do not apply for that household.

The components of this specification are detailed below.

The first part of the utility specification relates to parameters that do not follow a random distribution across individual households. The model uses K^f such parameters, where these are associated with individual attributes, e.g. x_{njk}^f. In this first part of the utility function, shifts in sensitivity between owners and renters are incorporated; that is, the marginal utility is β_k^f for owners, and $\beta_{kr}^f = \beta_k^f + \Delta_{kr}$ for renters. Statistical significance of Δ_{kr} thus denotes if the sensitivity for renters is significantly different from that for owners for the attribute x_{njk}^f . The standard errors of the renter specific parameters ($\beta_k^f + \Delta_{kr}$) are obtained using the Delta method that produces exact estimates with full maximum likelihood properties (Daly, Hess and de Jong, 2012). The standard errors for β_{kr}^f are calculated using the formula below (Daly, Hess and de Jong, 2012)

$$\sigma_{\beta_{kr}^f} = \sqrt{var(\beta_k^f) + var(\Delta_{kr}) + 2cov(\beta_k^f, \Delta_{kr})} \quad (5\text{-}2)$$

The second part of the utility specification relates to parameters that follow a random distribution across individual households, i.e. incorporating unobserved heterogeneity in preferences. In this case, owner and renter specific coefficients are estimated explicitly, where this is more convenient in the estimation software. The differences between owners and renters in both the mean sensitivities and the level of heterogeneity are allowed.

The third component, $\xi_{rn} r_n + \xi_{on} o_n$, allows for differences between owners and renters in the amount of noise in the utility. In the discrete choice technique, the variance of the unobserved factors for one group can be different from that for the other groups – this can reflect a number of different effects, either more noise in the attributes for one group or a greater role for unobserved attributes. If a model specification does not control for this, then the parameters for the two groups cannot be compared other than in the form of relative sensitivities. (e.g. Carrasco & de Dios Ortúzar, 2002; Train, 2003; Hensher, Rose, & Greene, 2015; Hess and Train, 2017). This study relies on an error components approach (e.g. Brownstone et al., 2000, Hensher, Rose, & Greene, 2015, Hess and Train, 2017) instead of the nested logit "trick" (e.g. Hensher and Bradley, 1993; Ben-Akiva et al., 1994; Bradley and Daly, 1997; Ho & Mulley, 2013) given that other random heterogeneity are also incorporated through mixing. ξ_{rn} and ξ_{on} are Normally distributed disturbances, with a mean fixed to *0* and an estimated standard deviation. They are shared across all zones and vary randomly across individuals within the group (owners or renters). A larger standard deviation for an error component then indicates more noise.

Some normalisation is required for this model, as follows:

- At least one of the attributes needs to be treated as having generic sensitivity between owners and renters in order to be able to also estimate the difference in the error variance (otherwise the estimation would equate to two separate models which would prevent the estimation of the additional error term). After comparison of group specific models, the sensitivity to crime is fixed to be generic between the two groups as the coefficient was most similar for this attribute.

- Only one of the two error components for differences in noise, i.e. ξ_{on} or ξ_{rn}, can be estimated, with the other fixed to zero. After comparing specifications estimating either ξ_{rn} or ξ_{on}, the noise for renters is found higher than for owners, and thus fixed $\xi_{on} = 0$, estimating only ξ_{rn}.

Given the type I extreme value distribution for ε_{nj}, the probabilities in our model are of the Logit form, with the probability of household *n* choosing zone *i* given by:

$$P_{ni}(\beta^f, \beta_n^h, \xi_{rn}, x_n) = \frac{e^{V_{ni}}}{\sum_{j=1}^{J} e^{V_{nj}}} \tag{5-3}$$

Where $V_{nj} = \sum_{k=1}^{K^f} (\beta_k^f + \Delta_k r_n) x_{njk}^f + \sum_{l=1}^{L^h} (\beta_{lno}^h o_n + \beta_{lnr}^h r_n) x_{njl}^h + \xi_{rn} r_n$, i.e. dropping the extreme value error term ε_{nj} and the normalised ξ_{on} term. This probability is conditional on the attributes x_n, estimates for the fixed parameters $\beta^f = < \beta_1^f, \ldots, \beta_K^f >$ and shift parameters $\Delta_r =< \Delta_{1r}, \ldots, \Delta_{Kr} >$, and specific realisations of the heterogeneous parameters $\beta_n^h =< \beta_{1no}^h, \ldots, \beta_{Lno}^h, \beta_{1nr}^h, \ldots, \beta_{Lnr}^h >$ and the error term ξ_{rn}. Given the random distribution of these parameters, the unconditional probability is given by:

$$P_{ni}(\beta^f, \Omega^h, \sigma_r, x_n) = \int_{\beta_n^h} \int_{\xi_{rn}} \frac{e^{V_{ni}}}{\sum_{j=1}^{J} e^{V_{nj}}} h(\beta_n^h \mid \Omega^h) \phi(\xi_{rn}) d\beta_n^h d\xi_{rn} \tag{5-4}$$

where this is now conditional on the estimated parameters only, i.e. the vector β^f of fixed coefficients, the vector of parameters Ω^h for randomly distributed coefficients, and the standard deviation of the error component for renters, i.e. σ_r. In Equation (2), the conditional logit probabilities are then integrated over the distribution of the random terms, with density functions $h(\beta_n^h \mid \Omega^h)$ and $\phi(\xi_{rn})$.

The corresponding log-likelihood function of the model for all the observations is as follows:

$$LL(\beta) = \sum_{n=1}^{N} \sum_{i=1}^{I} y_{ni} \log P_{ni}(\beta^f, \Omega^h, \sigma_r, x_n) \tag{5-5}$$

where, $y_{ni} = 1$ if household *n* chose zone *i* and $y_{ni} = 0$ for all other unchosen alternatives. Maximisation of this *LL* function yields the maximum likelihood estimates for model parameters. This log-likelihood incorporates the integral in Equation (2), which does not have a closed form solution, and the model is thus estimated using numeric simulation.

Models are estimated with different specifications (e.g. generic coefficients for both owners and renters; owner and renter specific coefficients). The likelihood ratio (LR) test value is used for comparing competing models where the *LR* was calculated using equation below

$$LR = -2[LL_r - LL_u] \tag{5-6}$$

Where, LL_r is the log-likelihood for the restricted model, LL_u is the log-likelihood of the unrestricted model. The LR can be compared to a critical value from a χ^2_K distribution with *K* degrees of freedom, where $K = K_u - K_r$, with K_u and K_r are the numbers of the estimated parameters in the unrestricted and the restricted models respectively.

5.3.2 Variables specification

A set of attributes comprising land use, dwelling and transport attributes are considered as explanatory variables for this study. The household characteristics are interacted with the location attributes to capture the systematic taste variation (preference heterogeneity) across different groups of households. A list of potential attributes for residential location choice modelling is identified based on a literature survey (e.g. Bhat and Guo, 2004; Habib and Miller, 2009; Guevara, 2010; Zolfaghari, 2013. Hoshino, 2011). There is a risk of independent variables being strongly correlated to each other which may have serious consequences on the estimated parameters. Therefore, the correlations between the attributes are checked and found a weak correlation in most of the cases (the correlation matrix is attached in Appendix L). For example, the correlation between the commute distance and distance from CBD is found 0.19 which indicates a weak correlation (Rumsey 2016). The model parameters are listed in Table 5-2 and are explained below.

5.3.2.1 Location and land use characteristics

Land use mix

Land use mix is a widely used index of the homogeneity/ heterogeneity of land-use in the wards (zones). Its scale ranges from *0* to *1* where *0* and *1* stand for pure homogeneous and uniform mixed land use patterns respectively. It is computed as (Frank et al., 2004)

$$\text{Land use mix} = \sum_j \frac{[P_j \times \ln(P_j)]}{\ln(J)}, \tag{5-7}$$

where P_j = the proportion of the land area of the j^{th} land-use category. A positive coefficient of this variable will indicate a preference for mixed land use patterns. Six land use categories were considered in this research, namely residential use, commercial use, green space, transport facilities and others.

Land use type

This variable denotes the percentage of residential and commercial areas in the alternative zones. Households who prefer a quiet lifestyle may prefer a residential zone with less commercial activities whereas those who prefer an urban active lifestyle may prefer a residential zone with more commercial activities.

Ethnic composition

This variable denotes the percentage of households belonging to the different ethnic groups (white ethnicity, black ethnicity and Asian ethnicity). To test ethnic preferences in this research, the proportions of the ethnic groups in the zones are interacted with households from the same ethnic background.

Dwelling density

This variable denotes the total number of dwellings per square kilometre in each ward. It is intended that household dislike high dwelling density in their residential area.

School quality

Households having school going children are likely to be interested in the residential areas having good school facilities. This variable is reported by the Greater London Authority based on GCSE (General Certificate of Secondary Education) average point score. This parameter is estimated for the household having at least one school going children.

Crime rate

The crime rate is an indicator of the living quality of an area. An area with a higher crime rate is likely to be less attractive to households. This variable denotes the total number of crimes per year per thousand of the population.

Average household size

The absolute difference between individual household size and average household size of the location alternative is considered in this variable. A lower value of the variable indicates homogeneity in the household size.

Employment opportunity

A household having commute sensitive working members is more likely to be inclined on an area of high employment opportunity. Employment opportunity is the ratio of ward level total job opportunities and total population.

Table 5-2 Variables considered in the models

Variables	Interaction variables	Data sources	Unit	Anticipated impact
Location and land use characteristics				
Land use type				
Residential land use	Inner London	LHSD &WAD	Percentage	+
	Outer London	LHSD &WAD	Percentage	+
Commercial land use	-	LHSD &WAD	Percentage	-
Land use mix	-	LHSD &WAD	Index (0 to 1)	+
Ethnic composition	White people	LHSD &WAD	Percentage	+
	Asian people	LHSD &WAD	Percentage	+
	Black people	LHSD &WAD	Percentage	+
Dwelling density	Inner London	LHSD &WAD	Per square KM	-
	Outer London	LHSD &WAD	Per square KM	-
School quality	School going child	LHSD &WAD	Unitless score	+
Crime rate	-	LHSD &WAD	Per thousand people	-
Average household size	-	LHSD &WAD	Number	+/-
Employment opportunity	-	LHSD &WAD	Per person	+
Distance from CBD	-	LHSD &LTS	Kilometre	+
Dwelling characteristics				
Dwelling cost	Low income	LHSD &WAD	Pound	-
	Middle income	LHSD &WAD	Pound	-
	High income	LHSD &WAD	Pound	-
Dwelling type				
Detached house	Inner London	LHSD &WAD	Percentage	-
	Outer London	LHSD &WAD	Percentage	-
Flat	Inner London	LHSD &WAD	Percentage	+
	Outer London	LHSD &WAD	Percentage	-
Transport and travel characteristics				
Public transport accessibility	Having cars	LHSD &WAD	Score out of 8	+/-
	Don’t have car	LHSD &WAD	Score out of 8	+
Commute distance	-	LHSD &LTS	Kilometre	-

Distance from CBD

Household preference for active urban areas or suburban areas is likely to be heterogeneous. It depends on individual household lifestyle, preference and sociodemographic characteristics. Therefore, the distance between the centre of the City of London and the alternative ward is used as a potential parameter.

5.3.2.2 Dwelling characteristics

Dwelling cost

Dwelling cost (price or rent) sensitivities of households belong to the different income groups are estimated to capture the potential cost heterogeneity. Annual average income less than £30,000 is considered as low income, more than £60,000 is considered as high income and in between is considered as a middle-income group.

Dwelling type

Household sensitivity to the proportion of the different types of houses (percentage of detached houses, the percentage of flats) in inner and outer London are investigated.

5.3.2.3 Transport and travel characteristics

Public transport accessibility

This variable, taken directly from the WAD, is calculated by the Greater London Authority based on walk access time, service availability and network density. The range of this variable is 0 to 8 where 8 represents the highest level of accessibility. This variable is likely to have a significant impact on the utility of transit-dependent households (i.e. who do not own cars). Systematic taste heterogeneity of sensitivity towards this variable is tested among households who own cars and those who do not.

Commute distance

This is the distance in km between individual work location and residential location alternatives. For multiple working member households, the maximum value (distance between work location and potential location alternatives) among the workers in the household is considered (Lee and Waddell, 2010). This is based on the assumption that one commuter's work location and travel choice influences household residential location choice and the other commuter simply selects the work location and other decision conditional on chosen residential location (Pinjari et al., 2011).

5.3.2.4 Zonal constants

Since the total number of alternatives is very large (498), constants are used for broader areas. Five constants are used, dividing the zones into central, north, south, east and west London where the constant for west London is normalized.

5.4 Results

5.4.1 Estimated model parameters

Mixed multinomial logit (MMNL) models are estimated in this study for investigating the residential location choice behaviour of owners and renters using the professional software ALOGIT[21].

A systematic model specification process was used. After incorporating deterministic heterogeneity (e.g. income effects), random heterogeneity was tested. With a large choice set and sample size, this is a computationally burdensome process and was thus carried out prior to the incorporation of differences between owners and renters. Results indicated significant random heterogeneity only for commute distance, where a negative Lognormal distribution was used.

The scale heterogeneity is also captured by means of relative variance where the error variance for the owners is normalized. The estimated standard deviation of the error components of renter specific utilities is found to be very small and not significantly different from zero, indicating no significant scale differences between the owner and renter specific utilities.

The core focus of the analysis then turned to establishing the differences in behaviour between owners and renters. The estimated parameters of the final models are presented in Table 5-3.

Two pooled models are developed first where in the first model generic coefficients are estimated for all variables assuming equal sensitivity for owners and renters, with the only difference between them being the amount of noise in the utility. In the second model, generic and shift parameters are estimated for all variables assuming different sensitivities of owners and renters (called second pooled model in rest of the paper). The null hypothesis is that "the model that assumes different sensitivity for

---

[21]ALOGIT was found to have significantly shorter run times than the other comparable programs like R which prompted the choice.

owners and renters to all variables is not statistically different from the model that assumes equal sensitivity of all variables for both groups". The estimated model with generic coefficients for owners and renters for all variables results in a significant loss of fit compared to the model with specific coefficients for owners and renters. A likelihood ratio test (χ^2=160.2, degree of freedom (*DF*) =30, P=0.001) strongly rejects the null hypothesis. It confirms the existence of preference heterogeneity between owners and renters in their residential location choice, even after accounting for differences in the amount of utility variance (where this was not significant in any case). Opoku and Abdul-Muhmin (2010) also provided evidence about the potential differences in the preferences of ownership and renting. However, among all the parameters, shifts for only five parameters (commute distance, public transport accessibility of the households who owned car(s), percentage of detached houses in Inner London, percentage of detached houses in Outer London and percentage of flats in Outer London) are found to be statistically significant above the 90% confidence interval.

A third pooled model is then estimated, retaining only those shifts that are statistically significant. The null hypothesis is that "the model that assumes owner and renter specific sensitivities for a specific subset of variables is not statistically different from the model that assumes different sensitivity of owner and renters to all variables". A likelihood ratio test (χ^2=11.5, *DF*=21, *P*=0.001) then no longer rejects the third pooled model. The final pooled model helps to reduce the estimation time by minimizing the number of parameters estimated without significantly affecting the goodness-of-fit or the accuracy of the estimates of the models. In the following sections, the similarities and differences in the owner and renter specific parameters are discussed.

5.4.1.1 Similarities between owner and renter specific parameters

As seen in Table 5-3, the parameters for owners and renters have the same direction of sensitivity but the magnitudes of some of the coefficients are found to be significantly different. The influences of dwelling attributes on residential location choices are in general found to be significant for both groups. For example, the housing cost sensitivities of both owners and renters are found negative as expected and different income groups exhibit different levels of cost sensitivities (which is in agreement with the findings of Habib and Miller, 2009; Zolfaghari, 2013). Households from lower-income groups are observed to be more cost-sensitive than higher-income groups both for ownership and renting. All else being equal, the alternative zones having more detached houses are also found to be less preferable options – both for owning and renting. The disutility is found to be higher in Inner London where the zones with a higher percentage of detached houses are fewer in number than in Outer London.

The second group of attributes included land-use and location characteristics. These are also found to have considerable influence on residential location decisions. Households are found to have higher utilities for areas of higher residential activities and less commercial activities for both owning and renting (Habib and Miller, 2009; Zolfaghari et al., 2012; Malaitham et al., 2013 found similar result). Although households are found to have a lower preference for higher levels of dwelling density, results indicate that they prefer mixed land use patterns, with a high accessibility to job, shopping, transport and other facilities. This agrees with the findings of other studies – for instance, Arundel and Ronald (2017) have advocated mixed land-use for ensuring the sustainability of a community while absolute density is mentioned not to be effective. Preferences for ethnic/racial similarity are found to have a positive and statistically significant effect for both groups which suggests that people prefer to live in an area where a higher number of households come from the same ethnic/racial group - this is supported by findings of previous studies (e.g. Bhat and Guo, 2007; Ibraimovic and Hess, 2017). School quality (only considered for households with children) is found to have a positive effect for both owners and renters which is similar to findings in the literature (Zhou and Kockelman, 2008; Malaitham et al., 2013). Crime rates and household size (absolute difference between each household size and the zonal average) are found to affect the utility of owners and renters negatively. This indicates the clustering of households based on the zonal average household size (Zolfaghari et al., 2012 observed a similar finding). Although households are found to be inclined to choose areas having greater employment opportunities (also observed by Malaitham et al., 2013), they are found to be less interested to live in and around the central business district (CBD) (in agreement with Vega and Reynolds-Feighan, 2009 finding).

The third group of attributes consisted of transport and travel attributes. An increase in public transport accessibility is found to increase the utility of 'car-less' households but decreases the utility for 'car-owning' households. Malaitham et al., (2013) also observed that transit dependent households are more inclined to live close to the area having a higher level of transit accessibility than car dependent households. As expected, increased commute distance is found to result in greater disutility (see Bhat and Guo, 2004; Bhat and Guo, 2007; Zhou and Kockelman, 2008; Habib and Miller, 2009; Vega and Reynolds-Feighan, 2009 for similar outcome). With the use of a negative Lognormal distribution, the estimated parameters are the mean and standard deviation of the logarithm of $-\beta$. The standard deviation reveals significant taste heterogeneity across households. Both the mean and standard deviation of $\log(-\beta)$ are significantly different between owners and renters.

Constants are estimated to capture the utility of all factors that are not explained by the included explanatory variables. Since the total number of alternatives is very large (498), constants are estimated at the aggregate level. Therefore, separate constants are estimated for the alternatives in central, north, south, east and west London. The constants were found to be the highest for North London for both owners and renters. It may be noted that the highest value of the constant does not indicate that this is the most preferred zone. Rather, the estimated constants capture the effects of factors that are not included in the model (i.e. are unobserved). Therefore, the result indicates that the share of unobserved factors affecting the choice of North London is higher than that of the other four parts.

Based on the data analyses presented earlier indicating substantial differences between Inner and Outer London, statistical tests are conducted to test if the sensitivities to the variables corresponding to Inner and Outer London are statistically different from each other. The results of these tests indicate that the sensitivity towards four of the variables (% detached houses, % flats, % residential area in the ward and dwelling density) are significantly different between Inner and Outer London (Table 5-3). Parameters which are not significantly different between Inner and Outer London, are estimated as a generic coefficient for the whole of London.

5.4.1.2 Differences between owner and renter specific parameters

This section discussed the differences in the sensitivity of owners and renters in their residential location choice based on the estimated parameters. The interpretation based on comparing the estimated parameters is sound because a) the scale difference between the owner and renter specific parameters are captured and b) the variables are the same for both groups (except cost and commute distance). However, for additional interpretation, the elasticity and WTP (Willingness to Pay) values are calculated which are discussed in the next section.

The shift parameters for four variables: percentage of detached houses in Inner London, percentage of detached houses in Outer London, percentage of flats in Outer London and public transport accessibility of the households who own car(s) are found statistically significant at the 95% level of confidence. Owners are found to be more sensitive (negative) than renters to the areas having a high percentage of detached houses both in Inner and Outer London which is aligned with the previous findings in the literature (e.g. Paccoud and Mace (2017). Owners are found to have a reduced utility for areas with a higher percentage of flats in Outer London whereas renters' preferences are opposite but statistically less significant. House owners who own car(s) are found to have a reduced utility for areas with a high level of public transport accessibility. The preference of renters who own car(s) to the public transport accessibility is positive but insignificant. Many homeowners (who are likely to be from higher income groups and/or have better parking arrangements) may have multiple cars for active travellers in the household resulting in no/reduced demand for public transport use. On the other hand, many renters may have a single car that is used by one member in the household while the other members need to use public transport. These scenarios may result in differences in the sensitivity of owners and renters (who own a car) to the variable public transport accessibility in their residential zone. The shift parameters for housing cost and commute distance are also found to be statistically significant at the 95% confidence interval but the interpretation cannot be made directly based on the estimated parameters since the variables are different for owners and renters. Elasticity analysis in the next section is used for further interpretation of the sensitivity of owners and renters to the residential location choice attributes including housing cost and commute distance.

Table 5- 3 Estimation of long-term and medium-term residential location choices

Parameters	Pooled model with generic coefficients for owners and renters for all variables		Pooled model with generic coefficients and shifts (for renters) for all variables*				Renter-specific (Computed)		Pooled model with generic coefficients and statistically significant shifts (for renters)				Renter-specific (Computed)	
			Generic (Owner-specific also)		Shift (for renters)				Generic (Owner-specific also)		Shift (for renters)			
	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat
Constants														
Central London	0.274	2.7	0.269	2.1	-0.082	-0.4	*0.187*	*1.1*	0.250	2.4			*0.250*	*2.4*
South London	0.359	4.5	0.319	3.5	-0.047	-0.2	*0.271*	*1.3*	0.304	3.8			*0.304*	*3.8*
North London	0.577	5.9	0.489	4.4	0.103	0.4	*0.592*	*2.6*	0.505	5.1			*0.505*	*5.1*
East London	0.555	6.3	0.481	4.7	-0.150	-0.7	*0.331*	*1.7*	0.447	5.0			*0.447*	*5.0*
Dwelling characteristics														
Dwelling cost (price*0.0001, monthly rent*0.01)														
Household income less than £30,000	-0.259	-9.0	-0.557	-8.9	0.353	5.1	*-0.204*	*-7.1*	-0.567	-9.3	0.367	5.6	*-0.199*	*-7.3*
Household income between £30,000 to £60,000	-0.249	-7.9	-0.444	-6.8	0.283	3.9	*-0.161*	*-5.3*	-0.453	-7.0	0.295	4.2	*-0.158*	*-5.4*
Household income more than £60,000	-0.087	-5.5	-0.200	-6.9	0.127	3.7	*-0.073*	*-4.0*	-0.207	-7.3	0.139	4.3	*-0.068*	*-4.1*
Missing values	-0.023	-2.2	-0.028	-1.2	-0.017	-0.6	*-0.045*	*-3.1*	-0.041	-3.6			*-0.041*	*-3.6*
Dwelling type														
Detached house in inner London	-0.107	-7.3	-0.139	-7.4	0.113	3.8	*-0.026*	*-1.1*	-0.131	-7.2	0.085	3.0	*-0.045*	*-2.0*
Detached house in outer London	-0.027	-6.9	-0.029	-6.7	0.032	3.0	*0.003*	*0.3*	-0.028	-6.6	0.029	3.0	*0.001*	*0.1*
Flat in inner London	0.036	11.4	0.034	9.3	0.004	0.6	*0.038*	*6.2*	0.035	11.0			*0.035*	*11.0*
Flat in outer London	-0.010	-4.2	-0.012	-4.3	0.017	2.8	*0.005*	*1.0*	-0.012	-4.9	0.019	4.3	*0.007*	*1.6*
Location and land use characteristics														
Land use type														
Residential land area in inner London	0.167	13.8	0.170	11.6	-0.006	-0.2	*0.164*	*6.8*	0.166	13.9			*0.166*	*13.9*
Residential land area in outer London	0.239	14.7	0.248	13.5	-0.036	-0.9	*0.212*	*6.0*	0.242	14.9			*0.242*	*14.9*
Commercial land area in inner and outer London	-0.059	-7.1	-0.061	-5.5	0.002	0.1	*-0.059*	*-5.1*	-0.060	-7.2			*-0.060*	*-7.2*
Land use mix	1.630	5.5	1.450	4.3	0.770	1.0	*2.220*	*3.2*	1.589	5.3			*1.589*	*5.3*
Ethnic composition													*0.000*	*0.0*
Ratio of white people × white dummy	0.016	9.4	0.017	8.8	0.002	0.5	*0.019*	*4.7*	0.017	10.3			*0.017*	*10.3*
Ratio of asian people × asian dummy	0.041	13.8	0.038	11.6	0.005	0.8	*0.044*	*7.4*	0.040	13.6			*0.040*	*13.6*
Ratio of black people × black dummy	0.058	11.0	0.057	8.9	-0.012	-1.1	*0.045*	*5.3*	0.053	10.3			*0.053*	*10.3*

Table 5- 3 Estimation of long-term and medium-term residential location choices (cont.)

Parameters	Pooled model with generic coefficients for owners and renters for all variables		Pooled model with generic coefficients and shifts for renters for all variables*				Renter-specific (Computed)		Pooled model with generic coefficients and statistically significant shifts (for renters)				Renter-specific (Computed)	
			Generic (Owner-specific also)		Shift (for renters)				Generic (Owner-specific also)		Shift (for renters)			
	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat
Dwelling density														
Inner London	-0.027	-6.5	-0.026	-5.1	0.001	0.1	*-0.025*	*-2.6*	-0.026	-6.3			*-0.026*	*-6.3*
Outer London	-0.125	-17.5	-0.127	-15.2	0.008	0.5	*-0.119*	*-8.7*	-0.126	-17.4			*-0.126*	*-17.4*
School quality	0.007	4.7	0.008	4.7	0.001	0.3	*0.009*	*2.2*	0.008	5.3			*0.008*	*5.3*
Crime rate	-0.117	-3.3	-0.124	-3.3			*-0.124*	*-3.3*	-0.120	-3.3			*-0.120*	*-3.3*
Household size	-0.357	-4.7	-0.404	-5.0	0.183	0.9	*-0.220*	*-1.2*	-0.358	-4.8			*-0.358*	*-4.8*
Employment opportunity	0.176	5.2	0.169	4.6	0.029	0.8	*0.198*	*4.7*	0.185	5.5			*0.185*	*5.5*
Distance from CBD	0.056	8.0	0.050	6.3	-0.017	-1.0	*0.033*	*2.2*	0.046	6.5			*0.046*	*6.5*
Transport and travel characteristics														
Public transport accessibility														
Households own car	-0.156	-3.9	-0.179	-3.8	0.211	2.2	*0.032*	*0.4*	-0.192	-4.5	0.256	3.9	*0.064*	*1.0*
Households do not own car	0.460	9.4	0.359	5.5	-0.013	-0.1	*0.346*	*3.3*	0.354	7.0			*0.354*	*7.0*
Commute distance**														
Mean of *log(-β)*	-1.681	-56.1	-1.711	-63.1	0.165	3.2	*-1.546*	*-34.5*	-1.716	-63.8	0.183	3.6	*-1.533*	*-35.1*
Standard deviation of *log(-β)*	0.185	61.7	0.137	78.9	0.172	5.9	*0.309*	*10.6*	0.134	74.6	0.175	6.0	*0.309*	*10.6*
Standard deviation of the error for renters (inverse function of scale effect)	0.071	0.7			0.057	0.3					0.052	0.4		
Measures of model fit														
Number of observations	2700		2700						2700					
Initial LL	-16768.620		-16768.620						-16768.620					
Final LL	-12944.248		-12864.170						-12869.897					
Adjusted ρ2	0.226		0.229						0.230					

* The parameter crime rate is considered as shared between owners and renters to allow us to capture scale haterogeneity.

** For this random parameter, owners and renters specific coefficients are estimated and the shift parameter is calculated.

5.4.2 Elasticity analysis

The findings of the models are further analysed by looking at elasticities. Elasticity analysis is the more appropriate tool for interpreting the relative impact of model parameters (Washington et al., 2010). It quantifies the percentage change in the choice probability of one alternative due to changes in the value of an attribute of the same alternative (called direct elasticity) or another alternative (called cross elasticity). The well-known formula for the calculating direct elasticity of the MNL model parameters (e.g. Train, 2009) is

$$E_{ix_{ni}} = \frac{\partial V_{ni}}{\partial x_{ni}} x_{ni}(1 - P_{ni}) \qquad (5\text{-}8)$$

where x_{ni} is the attribute of alternative *i* of household *n*, $\frac{\partial V_{ni}}{\partial x_{ni}}$ indicates the changes in the utility of alternative *i* of household *n* due to changes in the attributes of the corresponding alternative, $P_{ni}(\beta_f, \beta_n^h, \xi_r, x_n)$ is the probability of choosing alternative zone *i* by household *n*. The elasticity for the MMNL model is given by the integration of the MNL elasticity. Therefore, the direct elasticity for the MMNL model (see e.g. Hess et al., 2009) is given by

$$E_{ix_{ni}} = \frac{\int_{\beta_n^h} \int_{\xi_{rn}} \frac{\partial V_{ni}}{\partial x_{ni}} x_{ni}(1 - P_{ni}(\beta_f, \beta_n^h, \xi_{rn}, x_n)) P_{ni}(\beta_f, \beta_n^h, \xi_{rn}, x_n) h(\beta_n^h \mid \Omega^h) \phi(\xi_{rn}) d\beta_n^h d\xi_{rn}}{\int_{\beta_n^h} \int_{\xi_{rn}} P_{ni}(\beta_f, \beta_n^h, \xi_{rn}, x_n)\, h(\beta_n^h \mid \Omega^h) \phi(\xi_{rn}) d\beta_n^h d\xi_{rn}} \qquad (5\text{-}9)$$

In this study, direct elasticities are calculated for the MMNL estimation of residential ownership and renting decisions. The elasticities are calculated for all households for changes in attributes of the chosen alternatives (where households currently live)[22]. Then average elasticities across all individuals are computed. The results are presented in Table 5-4.

The direct elasticities calculated in this study reflect the change in the likelihood of choosing a residential zone due to changes in the attributes of the zone where the households are currently living. As observed in Table 5-4, households' residential location choices are found to more elastic (greater or equal to one) to the parameters associated with housing cost for low income owners and low and middle income

[22] Computing the elasticity for the chosen alternative only is appropriate when the number of alternatives in the choice set is very high, making the calculation of cross-elasticities too burdensome (Sener et al., 2011).

renters, some dwelling types (flats in inner London), residential land area in the zone, land use mix, ethnic composition (for white and Asian people), dwelling density, school quality, public transport accessibility for households who do not own car, and commute distance. The household residential choices are found to be less elastic (less than one) for the rest of the parameters.

Table 5-4 Direct elasticities of the owner specific and renter specific parameters

Parameters	Owner-specific	Renter-specific
Dwelling characteristics		
Dwelling cost (price*0.0001, monthly rent*0.01)		
Household income less than £30,000	-1.076	-1.248
Household income between £30,000 to £60,000	-0.886	-1.100
Household income more than £60,000	-0.543	-0.627
Dwelling type		
Detached house in inner London	-0.392	-0.065
Detached house in outer London	-0.225	0.023
Flat in inner London	2.280	2.743
Flat in outer London	-0.419	0.202
Location and land use characteristics		
Land use type		
Residential land area in inner London	2.613	2.589
Residential land area in outer London	2.463	2.264
Commercial land area in inner and outer London	-0.342	-0.476
Land use mix	1.151	1.841
Ethnic composition		
Ratio of White people × White dummy	1.273	1.360
Ratio of Asian people × Asian dummy	1.055	1.162
Ratio of Black people × Black dummy	0.937	0.774
Dwelling density		
Inner London	-1.219	-1.282
Outer London	-2.567	-2.686
School quality	2.298	2.575
Crime rate	-0.152	-0.208
Household size	-0.458	-0.232
Employment opportunity	0.082	0.176
Distance from CBD	0.839	0.430
Transport and travel characteristics		
Public transport accessibility		
Households own car	-0.576	0.120
Households do not own car	1.442	1.578
Commute distance		
Mean	-1.776	-1.666
Standard deviation	0.195	0.422

The interpretation of the differences in the sensitivities of owners and renters based on the estimated coefficients remains the same in the elasticity analysis. As in the estimated parameters, the elasticities for the share of detached houses in Inner and Outer London, the share of flats in Inner London and public transport accessibility (for those who own a car) are considerably higher for owners than renters and in some cases, the signs are opposite (e.g. for detached house in outer London and flats in outer London). The choice of the renters is found to be more elastic to housing cost than that of owners and the opposite applies in the case of commute distance. The elasticities of few other parameters such as commercial land use, land use mix, household size, employment opportunity and distance from CBD are found to vary more than 40% between owners and renters. Therefore, the elasticity analysis also reflects some significant differences in the sensitivities of the owners and renters in their residential location choice attributes.

5.4.3 Willingness-to-Pay (WTP) values

While the analyses in Table 5-3 and Table 5-4 indicate the relative influence of the residential location choice variables, willingness to pay (WTP) analysis can help to translate them into monetary values. As mentioned in the first section, WTP values can be used directly for a cost-benefit analysis to evaluate alternate policies. This makes it a very useful tool for quantifying the monetary value associated with improvement or deterioration in the level of an attribute. For example, WTP for decreasing dwelling density will indicate how much extra rent (or price in the case of ownership) a household is ready to pay for each unit of decrease in dwelling density. WTP can be estimated using the following expression:

$$WTP_k = -\frac{\frac{dV_{ni}}{dk}}{\frac{dV_{ni}}{dcost}} = -\frac{\beta_k}{\beta_{cost}} \tag{5-10}$$

where β_k is the sensitivity to attribute k and β_{cost} is the cost coefficient (monthly rent or dwelling price).

WTP values are calculated for the parameters that influence the residential location decision of owners and renters. The results are presented in Table 5-5 and explained here. As observed in the table, there is a distinct impact of income. The higher-income households are willing to pay significantly more compared to the lower-income group.

Table 5-5 Willingness to Pay (WTP) for owners and renters

Parameters	Unit	GLA average	Owners WTP (price in GBP)			Renters WTP (monthly rent in GBP)		
			Low income	Middle income	High income	Low income	Middle income	High income
Dwelling type								
Detached house in inner London	Percentage	2%	-24973	-31351	-69600	-13	-16	-36
Detached house in outer London	Percentage	9%	-5224	-6559	-14560	1	2	4
Flat in inner London	Percentage	74%	6015	7552	16765	18	23	52
Flat in outer London	Percentage	36%	-2070	-2599	-5770	3	3	7
Location and land use characteristics								
Land use type								
Residential land area in inner London	Percentage	14%	30553	38356	85150	81	102	225
Residential land area in outer London	Percentage	11%	44528	55901	124100	104	132	290
Commercial land area in inner and outer London	Percentage	7%	-10874	-13651	-30305	-29	-37	-81
Land use mix	Index	0.81	260136	326577	725000	1088	1377	3041
Ethnic composition								
Ratio of white people × white dummy	Percentage	72%	3025	3797	8430	9	12	26
Ratio of asian people × asian dummy	Percentage	12%	6880	8637	19175	21	27	60
Ratio of black people × black dummy	Percentage	10.50%	10187	12788	28390	22	28	62
Dwelling density								
Inner London	Dwelling per sq. km.	4956	-47	-59	-132	-0.1	-0.2	-0.3
Outer London	Dwelling per sq. km.	2138	-228	-287	-637	-0.6	-0.7	-1.6
School quality	Score	293	1364	1713	3802	4	6	12
Crime rate	Crime per thousand people	135	-222	-279	-620	-1	-1	-2
Household size	Number	0.4	-72390	-90878	-201750	-108	-137	-302
Employment opportunity	Employment per person	0.6	30319	38063	84500	97	123	272
Distance from CBD	Kilometer	15.1	9055	11367	25235	16	21	46
Transport and travel characteristics								
Public transport accessibility		3.63						
Households own car	Index		-32060	-40248	-89350	16	20	44
Households do not own car	Index		64424	80878	179550	169	214	474
Commute distance		21.4						
Mean	Kilometer		-32723	-41081	-91200	-110	-139	-306
Standard deviation	Kilometer		4521	5676	12600	35	44	97

The willingness to pay for owners is found negative for an increase in the share of detached houses in Inner London, detached houses in Outer London and flats in Outer London and positive for an increase in flats in inner London. However, the willingness to pay for renters is negative for an increase in the share of detached houses in Inner London but positive for an increase in detached houses in Outer London and flats in Inner and Outer London. Households are more interested in residential areas in Outer London than in Inner London, therefore, their willingness to pay for per unit increase of residential area in Outer London is 1.5 times higher than Inner London. Similarly, households are more sensitive to dwelling density in Outer London than Inner London. For instance, the willingness to pay for an increase in dwelling density is negative for both inner and outer London but the rate is five times higher for outer London than in inner London. The WTP for per kilometre saving in commute distance is much higher compared to the per kilometre increase in distance between residential location and the CBD for all income groups (~4 times for owners and 7 times for renters). Both owners and renters are willing to pay more for an increase in the share households from the same ethnic group in their neighbourhood, more balanced land use, better school quality, higher employment opportunity, better public transport accessibility (car-less households only). Both groups are willing to pay less for an increase in the commercial area, crime rate, and household size in their residential zone.

5.5 Conclusions and policy recommendations

In this study, differences between owners and renters in residential location choices are investigated using RP data combined with multiple other data sources. Publicly available real-world data is used to estimate residential location models without requiring sampling of alternatives. The paper contributes to the state-of-the-art by addressing the research gap identified in the introductory section. The results indicate that:

1. Both owners and renters have similar preferences (same signs of parameters) but the sensitivities to many attributes are different. A few parameters are found different significantly between owners and renters such as percentage of detached houses in Inner London, the percentage of detached houses in Outer London, the percentage of flats in Outer London, and public transport accessibility for the households who have car ownership.
2. For investigating the residential location choice behaviour, the potential differences in the sensitivities or preferences of owners and renters towards the attributes should be acknowledged as done in our work.

It may be noted that some of the findings from the study are 'London-specific' – the shift in preferences towards renting vs ownership in inner and outer London for example. However, acknowledging the differences in the elasticity and WTP among renters and owners for different land-use and dwelling attributes shows an important proof-of-concept that incorporating the heterogeneity and the full range of attributes can add value to the detailed cost-benefit analyses.

The use of revealed preference data in this study helped to capture the true preference of households with accurate parameter estimation avoiding the potential bias associated with the hypothetical response in stated preference data. Combining data from a range of sources enabled us to capture a wider range of attributes compared to previous state-of-the-art models (which had mostly dealt with a smaller subset of variables in isolation due to data limitations) and therefore expected to lead to better predictions. There are however several limitations of this study. First, the full choice set is considered for each household which is very large (498 alternatives in this case) and may not be behaviourally representative for all. In reality, the opportunities and constraints do affect the detailed choices. In this case, the location choices refer to the choice of wards as opposed to dwellings and such availability effects are likely to be reduced due to the coarser granularity. However, Chapter 6 is focused to address this issue by constructing restricted choice sets for each respondent based on behavioural rules rather than considering a full choice set. Second, the geographically closer location alternatives are most likely to be more correlated with their unobserved factors than the alternatives that are far from each other. Although, this spatial correlation structure among the residential location choice alternatives has been investigated in the literature (e.g. Bhat and Guo, 2004; Sener et al., 2011), this study is limited in this context. Third, the work location is considered as exogeneous in this study. However, the decisions of residential location and work location can be simultaneous or may have two-way interactions. Ignoring this decision interdependency can under/overestimate the correlations among the decisions neglecting the inherent trade-offs. The decision of residential location can also be interdependent with other decisions as well such as tenure choice, car ownership, travel behaviour amongst others, which was not tested in this study.

Fourth, the data sets used in this study is from 2002. Although the absolute sensitivities are likely to have changed over time due to the market dynamics and continuous gentrification, it is expected that the direction of sensitivity of the estimated parameters still holds in the current context. This is validated by the results of recent literature in the context of London and other similar cities. However, the framework proposed here can be used to investigate the housing market using recent data. Even in its current form, the models provide important behavioural insights on how people trade-off differently when making location choices in different time scales.

Finally, this chapter has captured the residential location choice behaviour of the households who move between 1970 to 2002 but still prefer their current place to live.

Therefore, the attributes in the year 2002 are used for explaining the choices (discussed in section 5.2.4 in detail). However, for many households (specifically those who had longer tenure), the preferences they had during relocation might not be the same as their current preferences, but the factors driving for considering the current place could be the changes in their circumstances, inertia or relocation cost. To capture this longitudinal behavioural change, the dataset used in this study was not suitable because the dataset did not have households and location information when the decision has made. However, some analyses have been conducted to see whether there are any differences in the preferences of the households who moved in the different time periods but still considering their current locations to live (details are presented in Appendix M). Although this analysis may not explain exactly the longitudinal nature of the behavioural change but gives insights about the time-varying nature of the preferences which require to investigate in the future study using the suitable longitudinal data if available.

The behavioural insights from the model outcomes in this chapter can be considered during land use and transport related policy analysis. It can be useful to consider this study findings in the Housing Market analyses which currently use simplifying assumptions and neglect important dimensions of the housing market– for instance, the fact that households have preferences for different house types and neighbourhoods and areas (Jones et al., 2010). Further, the WTP and elasticity analyses can be used for predicting the impacts of alternative policy scenarios due to explicit consideration of the sub-markets. Policymakers might be interested to know the shifting in the market share after implementing a new policy. For example, households are more interested in the higher level of public transport accessibility in their residential zone. If the policy targets to improve the public transport accessibility in an area, it is important to know how people will behave in response to the change. The model findings in this chapter can answer this policy question. Accounting for preference heterogeneity between the owners and renters is also expected to lead to better investment decisions in the housing market. People's preference for balanced (mixed) land use has been found to increase significantly over the years. Therefore, the new policy to shift the urban growth pattern from homogeneous to mix type will help to reduce total travel and car dependency. However, the difference between owners and renters preferences to mixed land use development need to be accounted carefully. Since compact development has claimed advantage but people dislike the higher level of dwelling density, density in mixed development also needs to be handled carefully. Alternatively, better ‘education’ about the benefits of high-density developments or incentives for accepting this kind of development can also be useful. The amount of green/open spaces in compact development is important to the households living there (Boyko and Cooper, 2011). Since both owners and renters are found interested in the flats, multi-storeyed flats instead of conventional detached, semi-detached housing will allow saving land for green space even in compact development. Households who do not own cars are found to be more inclined to the public transport accessibility in their residential zone. Therefore, transit-oriented development or improving public transport accessibility can be a useful policy step to reduce car dependency. For any policy steps based on these findings, it is advised to cross check the outcomes with the recent data.

Chapter 6

Choice set construction in modelling residential location choice

6.1 Introduction

The residential location choice models in chapter 5 are estimated based on the assumption that households evaluated all possible alternatives in the greater London area (498 alternative zones) for selecting one which had the best match with their preferences. Several studies in the literature have used a similar approach where the full or universal choice set which was very large has been considered for each individual for modelling their residential location choice preferences (Bhat and Guo, 2004; Zolfaghari, 2013; Haque et al., 2018). However, this approach may not be behaviourally persuasive resulting bias in the parameter estimation due to choice set misspecification (Swait, 2001; Bell, 2007). In reality, households are unlikely to be aware of the full set of alternatives or to consider all alternatives they are aware of. Households might consider a reduced choice set (a subset of alternatives) based on their preferences, sociodemographic characteristics and their knowledge on available alternatives[23]. For example, household members may not consider an alternative if they do not have enough knowledge about it or if the alternative is very far from the workplace of the commuter in the household. Therefore, it is expected that better ways to model the choice set will make the models behaviourally more representative. This, in turn, will lead to more accurate models for planning and policymaking.

A review of literature on discrete choice models (detailed in section 2.7.2) reveals three types of approaches that have been used to construct the choice set. The two-stage probabilistic approach (Manski, 1977; Swait and Ben-Akiva, 1987) to model individual choice set explicitly is one of them. This approach, however, has combinatorial complexity when the universal choice set is large and is thus typically infeasible in the case of disaggregate residential location choice modelling. The two-stage deterministic approach has been used in more recent studies where a limited set of alternatives are considered for each individual based on some behavioural rules (Farooq and Miller, 2012; Rashidi et al., 2012; Zolfaghari, 2013). For example, a commute sensitive individual is assumed to only consider the residential location

[23] Reduced choice set have also been used in literature for computational purpose, called sampling of alternatives. On the other hand, the use of reduced choice set under choice set construction is behavioural (more detailed discussion is presented in sections 2.5.2 and 2.5.3 in Chapter 2).

alternatives close to his/her workplace. This approach has a high risk of excluding potential alternatives considered by the individual from his/her choice set and including irrelevant alternatives (not considered by the individual in reality). The performance of these elimination based exogenous choice set formation approaches have therefore been criticized in the literature (Zolfaghari, 2013). In both probabilistic and deterministic approaches, the choice sets are generated in the first stage and choice probabilities conditional on the choice sets are calculated in the second stage. The third approach comprises of the single-stage semi-compensatory techniques that model choice set implicitly by reflecting utility penalization of less attractive alternatives instead of direct elimination. The complexity of these methods is linear (as opposed to exponential) with the number of alternatives in the choice set and therefore tractable in the case of residential location choice modelling. Cascetta and Papola (2001) proposed the Implicit Availability/Perception (IAP) technique to model individual choice set implicitly which was then modified by Cascetta and Papola (2009) (Dominant rule-based random utility model), Martínez et al. (2009) (Constrained Multinomial Logit Model, CMNL) and Paleti (2015) (r^{th}-order Constrained Multinomial Logit model, rCMNL). However, these methods also have limitations in their ability to reproduce the true parameters in estimation (Bierlaire et al., 2010).

A review of the literature thus reveals that all existing methods have weaknesses alongside strengths. Zolfaghari (2013) compared the performance of elimination based non-compensatory approaches in the context of residential location choice modelling and observed poor performance of these approaches both in estimation and validation samples. However, a comparison against semi-compensatory approaches was beyond the scope of their research. In particular, it is unclear if a specific approach is better for a particular choice context. Furthermore, previous researchers have either focused only on choice set generation for long-term residential location choices (e.g. ownership) or did not make any distinction between the long and medium term (e.g. renting) decisions. Chapter 5, however, indicates significant differences between the sensitivities to different parameters in the two different residential choice contexts.

Motivated by these points, the specific objectives of the research presented in this chapter are as follows

- To evaluate the performance of semi-compensatory choice set generation techniques in the context of residential location choice modelling;
- To investigate the potential to improve choice set generation techniques without compromising the computational tractability; and
- To investigate the existence of underlying heterogeneity in the choice set of long and medium term residential location choices (residential ownership and renting respectively).

In the next section, the semi-compensatory choice set generation techniques are presented. An improved choice set generation technique is proposed afterword followed by the data for analysis, estimation results, validation results and concluding remarks.

6.2 Semi-compensatory choice set formation models

The two-stage approach of modelling consumer choice involves the modelling or selecting the consideration set in the first stage and modelling the preferences conditional on the consideration set in the second stage. The justification of using consideration sets is that it gives a more realistic presentation of the choice process, a better explanation of the behaviour of the decision makers and improvement in forecasting. If the choice process is not directly observed, the analyst cannot say that two-stage modelling to capture the consideration set in the first stage is behaviourally persuasive, moreover, the two-stage can lead to a misspecified model that makes erroneous forecasts (Horowitz and Louviere, 1995). Therefore, it is better to be open-minded about the existence of consideration set. However, it is argued in the literature that consideration does not provide information outside that contained in the utility function and is simply an indicator of preference (Horowitz and Louviere, 1995). Therefore, the choice could be simulated through a single-stage approach where information about consideration sets can be used in the utility as an endogenous part of the choice process to improve the estimation and prediction efficiency. This process will allow systematic variation in the probability of an alternative being considered by the decision makers as a function of exogenous variables and constraints imposed by the analyst to capture the consideration set. This concept is used in the single-stage semi compensatory approaches. In this approach, the systematic utility is adjusted based on the probability of an alternative being in the individual choice set. A penalty term is introduced in the utility equation for the adjustment. Therefore, the utility function for alternative i and person n can be defined as follows:

$$U_{ni} = V_{ni} + \ln(\phi_{ni}) + \varepsilon_{ni} \quad \text{where } 0 \le \phi_{ni} \le 1 \tag{6-1}$$

where ϕ_{ni} is the probability of alternative i being in the choice set of individual n, V_{ni} is the deterministic utility of the alternative i for individual n and ε_{ni} is the random term assumed to be identically and independently distributed (iid) extreme value across the alternatives. If there is a full probability of an alternative being in the individual choice set, the penalty term becomes zero (i.e. no adjustment is required).

Different functional forms of ϕ_{ni} is used in different semi-compensatory approaches. ϕ_{ni} is expressed as a binary logit function of attributes related to choice set formation in the Implicit Availability Perception Random Utility (IAPRU) model proposed by Cascetta and Papola (2001). The mathematical expression of ϕ_{ni} for attribute k is as follows:

$$\phi_{nik} = \frac{1}{1 + \exp \sum_k (-\mu_k Z_{ik})} \tag{6-2}$$

where z_{ik} is the parameter k correlated with availability/perception and μ_k is the scale parameters. Second order utility penalization is also proposed here based on the Taylor series expansion for further utility cut-off of less attractive alternatives. This method, however, leads to estimation difficulties in complex specifications with multiple constraints. Moreover, second order utility penalization has convergence issues when an extreme penalty is applied to the chosen alternative. For example, if commute distance is considered as an availability/perception attribute in residential location choice modelling and the penalty parameter is 0.8, the utility cut-off of a chosen alternative 10 km away from the individual workplace is 1498 units. Therefore, the choice probability of this alternative is likely to be zero which is behaviourally unrealistic and may also lead to convergence issues.

In a simpler method, Cascetta and Papola (2009) proposed to simulate the choice set (i.e. availability) implicitly based on the rule of dominance among alternatives. The principle is that an alternative i dominates alternative j if j is worse than i with respect to dominant attributes K (where K can be a single or multiple attributes). This framework, however, does not account for how worse one alternative is compared to another alternative. The rule of dominance can be expressed as follows:

$$y_{ij}^{nk} = \begin{cases} 1 \;\; if\; Q_{nik} > Q_{njk} \dfrac{and}{or} C_{nik} < C_{njk},\; \forall k \in K \\ \qquad\quad 0 \qquad otherwise \end{cases} \tag{6-3}$$

y_{ij}^{nk} indicates that alternative j is dominated by alternative i based on attribute k for individual n. Q_{nik} and C_{nik} stand for positive and negative coefficients respectively for attribute k. The penalty term can be expressed as follows:

$$\ln(\phi_{nik}) = \mu_k \left(\sum_{i \in C} y_{ij}^{nk} \right), \quad \mu_k\; is\; the\; scale\; parameter, \mu_k < 0 \tag{6-4}$$

The constrained multinomial logit model (CMNL) proposed by Martínez et al. (2009) has greater flexibility to accommodate multiple constraints with exogenous bounds (both upper and lower) to simulate the individual choice set implicitly. A binary logit functional form is also considered here to estimate the probability of alternatives being in the individual choice set (ϕ_{ni}). The binary logit functional form of ϕ_{ni} with an upper threshold on a constraint attribute k can be presented as follows:

$$\phi_{nik} = \begin{cases} 1 & if\; z_{ik} < t_{nk} \\ \eta_k & if\; z_{ik} = t_{nk} \\ \frac{1}{1+\exp\left(\mu_k (z_{ik} - t_{nk} + \delta_k)\right)} & otherwise \end{cases} \tag{6 -5}$$

$$\delta_k = \frac{1}{\mu_k} \ln\left(\frac{1-\eta_k}{\eta_k}\right) \tag{6-6}$$

where z_{ik} is the value of constrained attribute k, t_{nk} is the threshold of attributes k for individual n, η_k is the cut off tolerance (proportion of decision makers violates the threshold) and μ_k is the scale parameters ($\mu_k > 0$).

If the constraint is applied on multiple attributes, the total penalty term becomes

$$\ln(\phi_{ni}) = ln\left[\prod_{k=1}^{K} \phi_{nik}\right] = \sum_{k=1}^{K} ln\,(\phi_{nik})$$

$$= -\sum_{k=1}^{K} \ln\left(1 + \exp\left(\mu_k (z_{ik} - t_{nk} + \delta_k)\right)\right) \tag{6-7}$$

The exogenous threshold-based heuristic adopted in the CMNL model is relevant for many cases. For instance, it is unlikely that a low-income household considers very expensive houses as options. Therefore, the CMNL model has received considerable attention recently and found wider application in literature in the context of modelling location choice (Martínez and Hurtubia, 2006), parking management (Caicedo et al., 2016), mode choice (Castro et al., 2013), etc. However, Bierlaire et al. (2010) demonstrated the inconsistency of the choice set generated in the CMNL model with the Manski framework using simulated experiments on synthetic data.

Paleti (2015) proposed r^{th}-order CMNL model (called rCMNL) where the complexity is linear with the size of the choice set. A higher order functional form of the CMNL penalty term (ϕ_{in}) is proposed in this regard. The r^{th}-order penalty in the rCMNL model is the natural logarithm of the following r^{th}- order expressions.

$\phi_{ni}^{1} = \phi_{ni}$ *First order*

$\phi_{ni}^{2} = \phi_{ni}[(1 - \bar{P}_{ni}) + \phi_{ni}^{1} \times \bar{P}_{ni}]$ Second order

$\phi_{in}^{3} = \phi_{ni}[(1 - \bar{P}_{ni}) + \phi_{ni}^{2} \times \bar{P}_{ni}]$ Third order

$$\phi_{in}^{r} = \phi_{ni}[(1 - \bar{P}_{ni}) + \phi_{ni}^{r-1} \times \bar{P}_{ni}] \text{ where } \phi_{ni}^{0} = 1 \quad r^{th}\text{-}order \tag{6-8}$$

Where, $\bar{P}_{ni}$ is the probability of choosing alternative *i* from the full choice set without any penalization.

$$\bar{P}_{ni} = \frac{e^{V_{ni}}}{\sum_{j \in C} e^{V_{nj}}} \tag{6-9}$$

Using synthetic data and real-world data, the author demonstrates that higher order penalization performs considerably better than the CMNL model in terms of replicating the Manski model parameters. However, in both examples, the number of alternatives in the universal choice sets was very limited (three and five alternatives respectively). If the number of alternatives in the choice set goes up, the probability of each alternative is likely to go down. For a very large universal choice set (hundreds to thousands of alternatives), $\bar{P}_{ni}$ will be too small and $\phi_{ni}^{r} \simeq \phi_{ni}$ (i.e. the model collapses to the first order CMNL).

6.3 Improved constrained multinomial logit model (ICMNL)

Although the complexity of the CMNL model remains linear with the increase of the number of alternatives in the choice set, it has a weakness to replicate the outcomes of the Manski method (Bierlaire et al., 2010). The penalty term considered in the CMNL model is a first order penalty derived from the attributes that influence individual choice sets. The higher order utility penalization proposed in the rCMNL model can minimize the error in the CMNL model outcomes when the size of the universal choice set is small. In case of a large universal choice set, the higher order penalty in the rCMNL model collapses to the first order CMNL penalty and cannot offer any further improvement. This is due to the fact that the rCMNL model penalty depends on the probability of choosing alternatives from the universal choice set which is likely to be very small for large universal choice sets. Therefore, an alternate formulation of higher order approximation of the availability term (ϕ_{ni}) in the CMNL model is proposed based on the concept of Taylor's series expansion which is independent of the number of alternatives in the universal choice set. This is motivated by the application of Taylor's series expansion in the context of the Implicit Availability Perception (IAP) logit model (Cascetta and Papola, 2001). As mentioned above, the basic utility equation with implicit availability of alternative can be expressed as follows:

$$U_{ni} = V_{ni} + \ln(\phi_{ni}) + \varepsilon_{ni} \quad \text{where } 0 \leq \phi_{ni} \leq 1 \tag{6-10}$$

The availability term ϕ_{ni} can have any value between zero and one. Since the analyst does not know the true value of the ϕ_{ni} for an individual, it can be treated as a random parameter with an expected value of $E(\phi_{ni})$. Therefore the logarithm of the availability term, $ln\ (\phi_{ni})$, can be decomposed into its expected value, $E(\ ln\ (\phi_{ni}))$ and an error, δ_{ni} where the error term captures the difference between the true and expected penalty.

The revised utility equation can be presented as follows

$$U_{ni} = V_{ni} + E(\ln(\phi_{ni})) + \delta_{ni} + \varepsilon_{ni}$$

$$U_{ni} = V_{ni} + E(\ln(\phi_{ni})) + \tau_{ni} \tag{6-11}$$

For simplicity, the total error (τ_{ni}) is assumed to be independently and identically distributed (IID). Based on the 2nd order Taylor series expansion, the expected penalty can be expressed as below

$$E(\ln(\phi_{ni})) = E(\ln\bar{\phi}_{ni}) + E\left(\frac{\phi_{ni}-\bar{\phi}_{ni}}{\bar{\phi}_{ni}}\right) - \mathrm{E}\left(\frac{(\phi_{ni}-\bar{\phi}_{ni})^2}{2(\bar{\phi}_{ni})^2}\right)$$

$$E(\ln(\phi_{ni})) = \ln(\bar{\phi}_{ni}) - \frac{Var\ (\phi_{ni})}{2(\bar{\phi}_{ni})^2} \tag{6-12}$$

Since the distribution of ϕ_{ni} is unknown, the variance of ϕ_{ni} is also unknown. Considering the variance of Bernoulli distribution, $Var\ (\phi_{ni}) = \bar{\phi}_{ni}(1-\bar{\phi}_{ni})$ and the equation 6-12 can be modified as follows

$$E(\ln(\phi_{ni})) = \ln\bar{\phi}_{ni} - \frac{(1-\bar{\phi}_{ni})}{2\bar{\phi}_{ni}} \tag{6-13}$$

The utility equation 6-11 can be presented as below

$$\begin{aligned} U_{ni} &= V_{ni} + \ln\bar{\phi}_{ni} - \frac{(1-\bar{\phi}_{ni})}{2\bar{\phi}_{ni}} + \tau_{ni} \\ &= V_{ni} + \ln\bar{\phi}_{ni}^{T(2)} + \tau_{ni} \end{aligned} \tag{6-14}$$

$\ln\bar{\phi}_{ni}^{T(2)}$ is the second order utility penalty where

$$\bar{\phi}_{ni}^{T(2)} = \bar{\phi}_{ni} * e^{-}\left(\frac{1-\bar{\phi}_{ni}}{2\bar{\phi}_{ni}}\right), \tag{6-15}$$

The average availability $\bar{\phi}_{ni}$ can be estimated implicitly as a binary logit function of attributes related to the choice set membership of alternatives (Cascetta and Papola, 2001). Therefore, the binary logit faction of the availability term proposed in the CMNL model (see equation 6-5) is considered in the ICMNL model for calculating average availability ($\bar{\phi}_{ni}$). Constraints are applied to the attributes related to the alternative to estimate the choice set probability of the alternative using the binary logit function. If the constraint is applied to *K* number of attributes, the penalty can be calculated using the equation (6-16)

$$\begin{aligned} \ln(\bar{\phi}_{ni}^{T(2)}) &= ln\left[\prod_{k=1}^{K}\bar{\phi}_{nik}^{T(2)}\right] = \sum_{k=1}^{K} ln(\bar{\phi}_{nik}^{T(2)}) \\ &= -\sum_{k=1}^{K}\left\{\ln\left(1+\exp(\mu_k(z_{ik}-t_{nk}+\delta_k))\right) + \frac{\exp(\mu_k(z_{ik}-t_{nk}+\delta_k))}{2}\right\} \end{aligned} \tag{6-16}$$

$\bar{\phi}_{nik}^{T(2)}$ represents the availability of alternative *i* to be in the choice set of individual *n* when the constraint is applied on attribute *k*. The probability of choosing alternative *i* by household *n* in ICMNL model is as follow

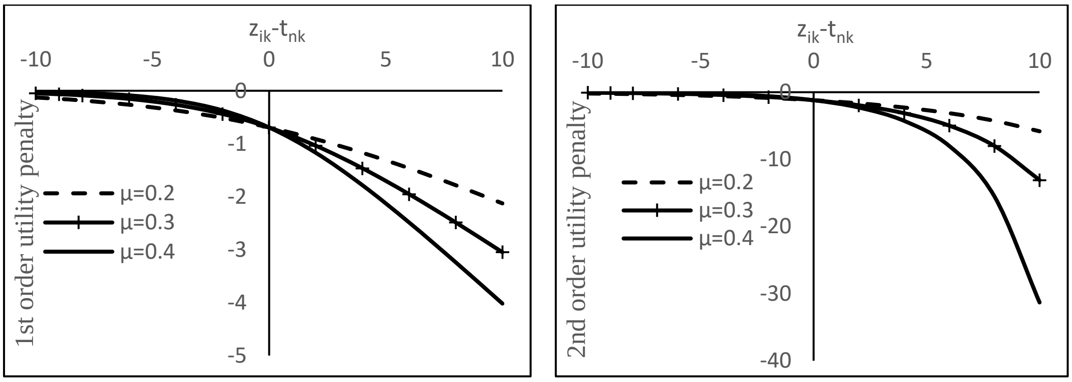
$$P_{in} = \frac{e^{V_{ni}+\ln(\bar{\phi}_{ni}^{T(2)})}}{\sum_{j\in C} e^{V_{nj}+\ln(\bar{\phi}_{nj}^{T(2)})}} \tag{6-17}$$

and the log-likelihood function is

$$LL(\beta) = \sum_{n}\sum_{i} y_{ni}\ln(P_{ni}) \tag{6-18}$$

where y_{ni} = 1 if alternative *i* is chosen by household *n* and y_{ni} = 0 for all nonchosen alternatives. The maximum likelihood estimates of the model parameters are found by maximizing this function. Functional forms used in different methods are explained in the following sections.

If the attributes move away from the bound, the rate of increment of 2nd order penalties in ICMNL becomes considerably stronger than the 1st order penalties. For example, for μ=0.4, if the value of $(z_{ik} - t_{nk})$ moves from 5 to 10, the increment of the first order penalty is 2 units which is 25 units for the second order penalty (Figure 6-1). Therefore, the first order penalty is considered as a soft penalty and the second order penalty is considered as a hard penalty. The scale parameter also determines the size of the penalty.



(a) 1st order penalty (soft) (b) 2nd order penalty (hard)

Figure 6-1 Penalization of the utility function

Due to applying hard penalties on those alternatives that are unlikely to be in the individual choice set, the choice probabilities of these alternatives tend to be zero which is behaviourally reasonable. Therefore, the ICMNL model is expected to be a better approximation of the Manski formulation. The performance of the ICMNL is evaluated here with a simple analysis (Bierlaire et al., 2010). For this analysis, only two alternatives are considered where alternative *1* is always available in the choice set (ϕ_1=1) and alternative *2* has a probability of being in the choice set ($\phi_2 \leq 1$). This hypothesis is similar to the CMNL and ICMNL model concept where alternatives within a threshold are always available in the choice set and choice set membership probabilities are assigned for those alternatives that are outside the threshold zone. In the CMNL, rCMNL, and ICMNL, the probability of choosing alternative *1* is as follows:

$$P_1 = \frac{e^{V_1 + \ln(\phi_1')}}{e^{V_1 + \ln(\phi_1')} + e^{V_1 + \ln(\phi_2')}} \qquad (6\text{-}19)$$

$$P_1 = \frac{e^{V_1}}{e^{V_1} + e^{V_2 + \ln(\phi_2')}},\ since\ \phi_1' = 1 \qquad (6\text{-}20)$$

where, V_1 and V_2 are the systematic utilities of alternatives *1* and *2*, respectively.

The mathematical formulation of penalty terms in the CMNL, rCMNL and ICMNL models can thus be summarized as follows:

$$\phi_2' = \phi_2 \qquad \text{CMNL} \qquad (6\text{-}21)$$

$$\phi_2' = \phi_2[(1 - \bar{P}_2) + \phi_2 \times \bar{P}_2] \qquad 2^{\text{nd}} \text{ order of rCMNL} \qquad (6\text{-}22)$$

$$\phi_2' = \phi_2 * e^{-\left(\frac{1-\phi_2}{2\phi_2}\right)} \qquad \text{ICMNL} \qquad (6\text{-}23)$$

The probability of choosing alternative *1* based on the Manski formulation is as follows:

$$P_1 = P(C[1]) * \frac{e^{V_1}}{e^{V_1}} + P(C[1,2]) * \frac{e^{V_1}}{e^{V_1} + e^{V_2}} \qquad (6\text{-}24)$$

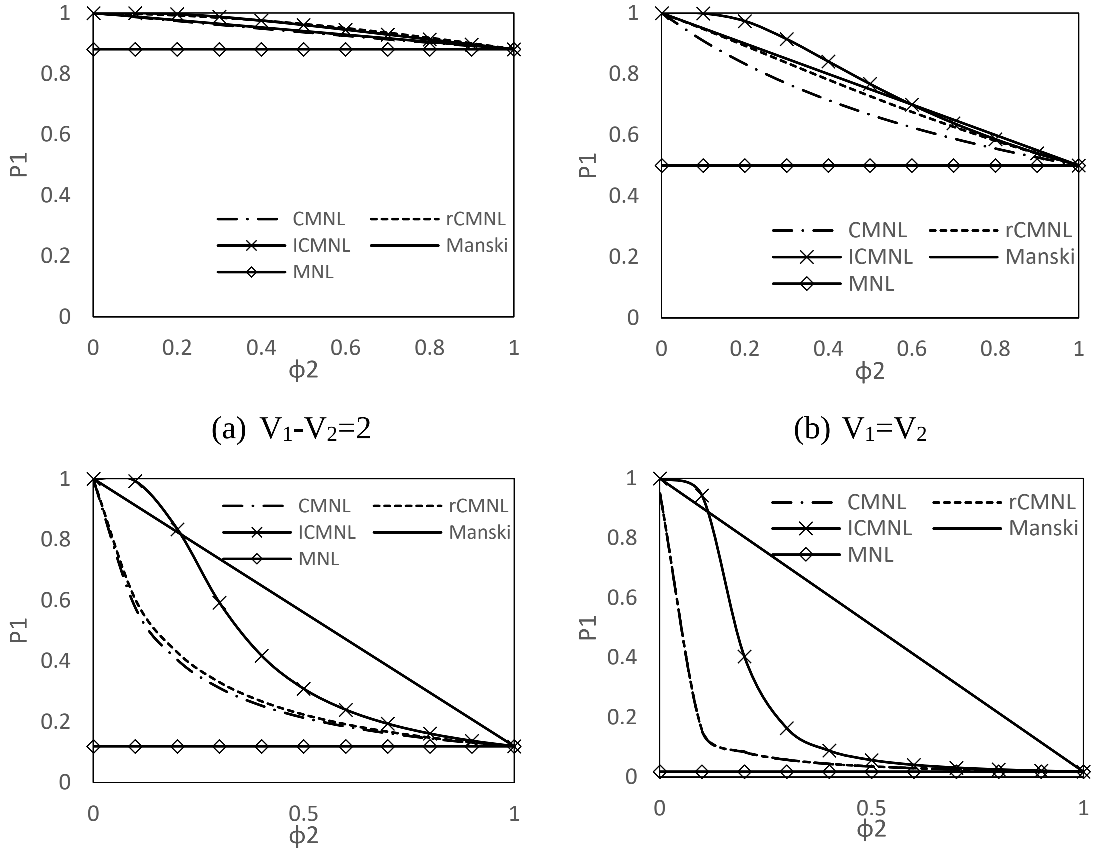
where *P(C[1])* and *P(C[1,2])* are the probabilities of choice sets containing alternative *1* only and both of the alternatives (*1* and *2*), respectively. The probability of a given choice set can be expressed as follows (Bierlaire et al., 2010).

$$P(C[1]) = \frac{\phi_1(1-\phi_2)}{1-(1-\phi_1)(1-\phi_2)} = 1 - \phi_2, since\ \phi_1 = 1 \qquad (6\text{-}25)$$

$$P(C[1,2]) = \frac{\phi_1\phi_2}{1-(1-\phi_1)(1-\phi_2)} = \phi_2, since\ \phi_1 = 1 \qquad (6\text{-}26)$$

Therefore $P_1 = (1 - \phi_2) + \phi_2 * \frac{e^{V_1}}{e^{V_1} + e^{V_2}}$ (6-27)

The choice probability of alternative *1* (P_1) is calculated for different values of ϕ_2 (probability of alternative *2* being in the choice set) under different conditions using the CMNL, rCMNL, ICMNL and Manski formulation presented above. The results are plotted in Figure 6-2. The figures show that the semi-compensatory approaches (CMNL, rCMNL and ICMNL model) can replicate the Manski probability quite well when the utility of alternative *1* (V_1) is larger than the utility of alternative *2* (V_2) (Figure 6-2a). On the other hand (when the utility of alternative *1* becomes smaller than alternative *2*), the semi-compensatory approaches cannot produce the Manski results (Bierlaire et al., 2010). However, the ICMNL model can reduce the gap between the Manski and CMNL model considerably (Figures 6-2c and 6-2d). The rCMNL model is only useful when the utilities of alternatives *1* and *2* are very close to each other (Figure 6-2b).



(a) V_1-V_2=2

(b) V_1=V_2

(c) V_1-V_2=-2

(d) V_1-V_2=-4

Figure 6-2 Choice probability of alternative 1 for different utility differences

6.4 Data for empirical analysis

The data set used in chapter 5 for modelling residential location choice is also used in this chapter. As explained detailed in chapter 5, the final dataset has been obtained by combining three different datasets such as The London Household Survey Data (LHSD), The Ward Atlas Data (WAD) and data from London Transport Studies (LTS) model. Variable specifications remain the same as described in chapter 5. Sampling weights provided with the dataset are also considered for estimation of the models. However, to get an insight into the underlying preferences of household choice set consideration, some additional analysis of the data is performed here.

Statistical analysis of the data shows that households are inclined to choose residential location alternatives close to their current home. However, owners' preferences to relocate near to their current home are found to be stronger than those of renters

(Figure 6-3). For example, around 90% owners chose their new locations within 14 km of their past homes and for the remaining around 10%, they are spread between 14 and 50km whereas around 90% renters chose their new locations within 18 km of their past homes and the remaining 10% around are spread between 18 and 50km. Sharp slope changes of the curves at a certain point in Figure 3 indicate the possible threshold effects of choice set consideration. Since most of the households chose their new locations close to their past home (e.g. 90% owners chose within 14 km of past home), it is unlikely that they considered the alternatives far from their past home (outside a threshold zone).

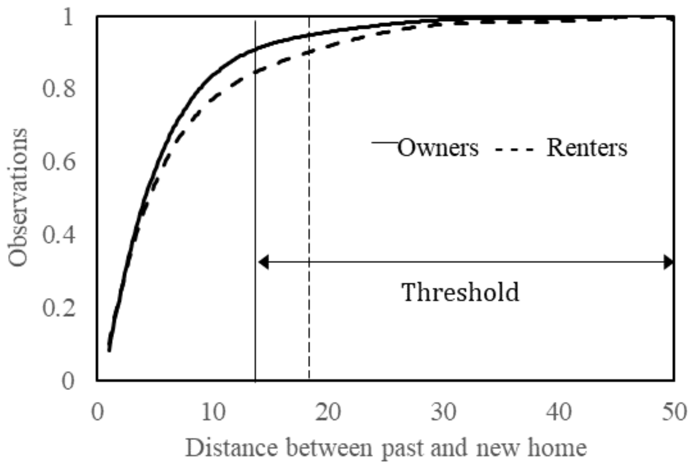


Figure 6-3 Distance between past and new home

6.5 Estimation results

Residential location choices of owners and renters are modelled in this study using the existing (CMNL and rCMNL models) and proposed (ICMNL model) semi-compensatory approaches where choice sets are simulated implicitly based on exogenous constraints on attributes. Models are estimated using R[24]. Although the choice set of an individual is likely to be influenced by a set of parameters (e.g. commute distance, distance of alternatives from the past home, distance of alternatives from the CBD, housing cost, etc.), households are found to have strong preferences to the alternatives close to their past home locations and work locations based on statistical analysis of the data. Therefore, the influence of distance of alternatives from

[24] The author modified the R codes of the Choice Modelling Center, University of Leeds to estimate the CMNL, rCMNL and ICMNL models where modification facilitates to calculate the penalty parameter implicitly. ALOGIT software used in the chapter 5 does not have flexibility for estimating the CMNL, rCMNL and ICMNL models.

the past home location and commute distance on individual choice sets have been explicitly tested. Only distance from the past home is found to have a significant influence on the household choice set consideration. An exogenous threshold is applied to the parameter of past home distance to simulate the choice set. Different thresholds are tested in the models and the threshold value that produces the maximum likelihood estimation is considered for the final model.

As an explanatory variable of individual choice, a large set of parameters including location characteristics, aggregate level dwelling chrematistics, commute characteristics and interaction variables are considered in the models. Different model specifications are tested and the final models contain the parameters statistically significant in any of the models. Several higher order approximations of the rCMNL model were tested in this study and the 3rd order approximation was found to give stable results in terms of improvement in model fit. The performance of the models estimated using the CMNL, rCMNL and ICMNL techniques is analysed based on the improvement of log-likelihood in the estimation sample[25].

6.5.1 Ownership

The estimated parameters of the ownership models are presented in Table 6-1. It is observed that the ICMNL model shows significant improvement in log-likelihood over the CMNL (138.71 units) and the rCMNL models (137.55 units). However, the improvement of the rCMNL model over the CMNL model is insignificant (only 1.2 units). This is due to the fact which is alluded to in the earlier section that the rCMNL model is equivalent to the CMNL model when the size of the universal choice set is large. Estimated parameters are found to be stable across the models estimated using different techniques.

All the parameters considered in the models have the expected sign and most of them are found to be statistically significant. Household cost sensitivity is found to be

---

[25] Goodness of fit of the heuristic based semi compensatory approaches (e.g. IAPRU model, dominance rule based approach, CMNL, etc.) can vary case by case. It depends on the appropriateness of the heuristic for the specific context. For example, dominance rule based approaches might be suitable for one case and exogenous threshold based approaches (CMNL) could perform better in another context. It is unfair to compare the methods based on different heuristics in a single dataset and to draw a general conclusion. Therefore, the performances of only CMNL, rCMNL and ICMNL models are compared in this study where this problem does not arise.

heterogeneous across different income groups. For example, the lower income group is more price sensitive than the higher income group, as expected.

Table 6-1 Estimation results of residential ownership models

Parameters	CMNL		rCMNL		ICMNL	
	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat
Constants						
Central London	0.4181	3.0	0.4249	3.2	0.4935	3.4
South London	0.3572	3.2	0.3614	3.4	0.3874	3.2
North London	0.6912	5.3	0.7000	5.7	0.7462	5.6
East London	0.7722	6.0	0.7827	6.4	0.8950	6.6
Dwelling characteristics						
Dwelling cost (price* 0.0001)						
Income less than £30,000	-0.5145	-6.5	-0.5185	-6.6	-0.5506	-6.7
Income between £30,000 to £60,000	-0.4799	-6.4	-0.4832	-6.4	-0.5145	-6.6
Income more than £60,000	-0.2212	-4.3	-0.2230	-4.4	-0.2323	-4.3
Missing values	-0.0393	-1.2	-0.0397	-1.2	-0.0515	-1.5
Dwelling type						
Detached house in inner London	-0.1247	-4.8	-0.1256	-4.8	-0.1182	-4.5
Detached house in outer London	-0.0275	-5.8	-0.0278	-5.9	-0.0295	-6.1
Flat in inner London	0.0258	5.8	0.0259	5.9	0.0253	5.6
Flat in outer London	-0.0116	-4.1	-0.0117	-4.2	-0.0113	-4.0
Location and land use characteristics						
Land use type						
Residential land area in inner London	0.1417	8.8	0.1426	8.9	0.1321	8.1
Residential land area in outer London	0.2124	9.5	0.2140	9.6	0.2104	9.4
Commercial land area in inner and outer London	-0.0635	-5.4	-0.0637	-5.5	-0.0585	-5.0
Land use mix	1.2774	3.3	1.2854	3.6	1.0715	2.8
Ethnic composition						
Ratio of white people × white dummy	0.0174	7.8	0.0176	7.9	0.0181	7.9
Ratio of asian people × asian dummy	0.0304	7.3	0.0309	7.5	0.0282	6.5
Ratio of Black people × black dummy	0.0492	6.2	0.0497	6.2	0.0462	5.6
Dwelling density						
Inner London	-0.0205	-3.8	-0.0206	-3.8	-0.0194	-3.5
Outer London	-0.1105	-10.5	-0.1114	-10.6	-0.1101	-10.5
School quality	0.0074	4.5	0.0074	4.5	0.0069	4.2
Crime rate	-0.0724	-1.3	-0.0750	-1.3	-0.0691	-1.2
Household size	-0.2817	-2.8	-0.2853	-2.9	-0.2755	-2.7
Employment opportunity	0.1505	0.1505	2.9	0.1538	3.0	0.1547
Distance from CBD	0.0825	9.2	0.0837	9.6	0.0992	10.7
Transport and travel characteristics						
Public transport accessibility						
Households own car	-0.1838	-3.7	-0.1855	-3.7	-0.1979	-3.9
Households do not own car	0.1564	2.1	0.1585	2.2	0.1314	1.8
Commute distance	-0.1464	-28.7	-0.1480	-29.0	-0.1518	-26.7
Penalty parameter (μ)						
Distance from past home	0.1990	38.0	0.2013	38.4	0.0187	34.6
Measures of model fit						
Number of observations	1875		1875		1875	
Initial LL	-11644.8751		-11644.8751		-11644.8751	
Final LL	-7722.9560		-7721.7950		-7584.2520	
Adjusted ρ2	0.334		0.334		0.346	

Preferences for ethnic similarity (where a higher number of households come from the same ethnic group) are found to have a positive and statistically significant effect. Result also shows that households dislike higher levels of dwelling density, commercial activities and crime in their residential areas. Although households prefer to live in areas with higher residential activities, they also prefer areas with more balanced land use patterns. Households do not prefer an area with a higher percentage of detached houses, this may be due to the excess price of detached houses in GLA.

However, households are found to be inclined to flats in inner London areas and seem to dislike flats in outer London areas, all else being equal. Households are also found to prefer areas having greater employment opportunities, good school facilities and those further from the central business district (CBD). The household size (absolute difference between individual household size and zonal average) parameter shows a negative effect on utility. Increases in public transport accessibility increase the utility of 'car-less' households but decrease the utility of 'car-owning' households. It is also observed that increased commute distance adds disutility to the residential location alternatives.

6.5.2 Renting

The goodness of fit of the proposed ICMNL model is found to be better than that of the CMNL model and the rCMNL model also in the renting dataset (Table 6-2). However, a loss of likelihood has been observed in the rCMNL model here compared to the CMNL model. In the rCMNL model, a larger penalty is applied to the alternatives having higher choice probabilities without the penalty term. Therefore, chosen alternatives outside the threshold zone are likely to get a strong penalty resulting in a decrease in the model fit.

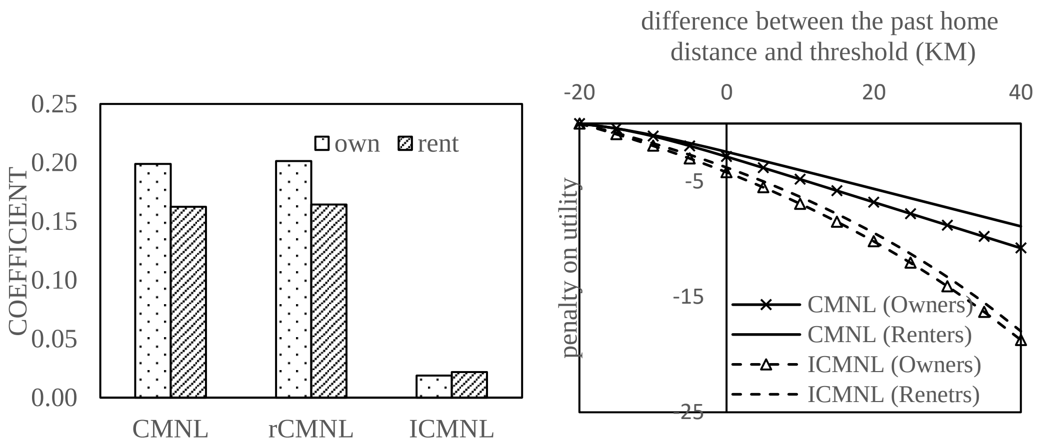
All parameters in the renting models also obtain the expected sign. Some of the estimated parameters are found to be statistically insignificant but are retained in the models to ensure consistent parameter specification in both ownership and renting models. Since the estimated parameters in the renting models give the same sign of the corresponding parameters estimated in ownership models, the interpretations are the same.

Table 6-2 Estimation results of residential renting models

Parameters	CMNL		rCMNL		ICMNL	
	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat
Constants						
Central London	0.5287	2.1	0.5324	2.1	0.5649	2.3
South London	0.3928	1.7	0.3960	1.6	0.4096	1.7
North London	0.4967	1.7	0.5044	1.7	0.5341	1.9
East London	0.8874	3.4	0.8962	3.4	0.9518	3.7
Dwelling characteristics						
Dwelling cost (monthly rent* 0.01)						
Income less than £30,000	-0.2570	-5.2	-0.2596	-5.3	-0.2651	-5.3
Income between £30,000 to £60,000	-0.1336	-2.4	-0.1348	-2.4	-0.1309	-2.4
Income more than £60,000	-0.0949	-2.7	-0.0957	-2.8	-0.0985	-2.8
Missing values	-0.0785	-2.8	-0.0793	-2.8	-0.0809	-2.8
Dwelling type						
Detached house in inner London	-0.0301	-0.7	-0.0301	-0.7	-0.0321	-0.8
Detached house in outer London	-0.0095	-0.7	-0.0097	-0.7	-0.0100	-0.7
Flat in inner London	0.0335	4.0	0.0338	4.0	0.0338	4.0
Flat in outer London	0.0059	0.8	0.0060	0.9	0.0057	0.8
Location and land use characteristics						
Land use type						
Residential land area in inner London	0.1602	5.8	0.1614	5.9	0.1569	5.8
Residential land area in outer London	0.2273	4.4	0.2293	4.5	0.2225	4.3
Commercial land area in inner and outer London	-0.0715	-3.8	-0.0722	-3.8	-0.0696	-3.8
Land use mix	2.5651	3.0	2.5899	3.0	2.4297	3.3
Ethnic composition						
Ratio of white people × white dummy	0.0184	3.4	0.0186	3.5	0.0186	3.5
Ratio of asian people × asian dummy	0.0422	4.6	0.0429	4.7	0.0429	4.5
Ratio of Black people × black dummy	0.0393	2.8	0.0398	2.8	0.0374	2.6
Dwelling density						
Inner London	-0.0196	-2.5	-0.0197	-2.5	-0.0195	-2.5
Outer London	-0.1222	-5.3	-0.1232	-5.3	-0.1207	-5.3
School quality	0.0079	1.5	0.0080	1.6	0.0072	1.4
Crime rate	-0.2545	-2.8	-0.2583	-2.8	-0.2564	-2.8
Household size	-0.0763	-0.4	-0.0775	-0.4	-0.0670	-0.3
Employment opportunity	0.3385	4.0	0.3437	4.0	0.3435	4.1
Distance from CBD	0.0775	4.0	0.0784	4.0	0.0861	4.3
Transport and travel characteristics						
Public transport accessibility						
Households own car	0.1331	1.3	0.1336	1.3	0.1267	1.2
Households do not own car	0.3680	3.6	0.3702	3.7	0.3604	3.5
Commute distance	-0.1849	-16.5	-0.1867	-16.6	-0.1912	-16.3
Penalty parameter (μ)						
Distance from past home	0.1625	14.0	0.1643	14.2	0.0217	14.5
Measures of model fit						
Number of observations	382		382		382	
Initial LL	-2372.4492		-2372.4492		-2372.4492	
Final LL	-1678.9830		-1679.5420		-1667.2620	
Adjusted ρ2	0.280		0.279		0.285	

6.5.3 Contrast between ownership and renting

In terms of the penalty term in the models to simulate the choice set probability implicitly, the result shows consistent preferences with the prior statistical analysis (owners' preference to the alternatives close to the past home location is stronger than renters' preference, Figure 6-1). The differences in the estimated values of the penalty term (μ) in ownership and renting models are found to be statistically significant (Figure 6-4a). Figure 6-4b also confirms that the penalty applied to owners' utility due to the increase of the distance of alternatives from past home is always higher than that of renters. This means that alternatives close to the current home have a higher probability to be included in the choice set of owners than renters. The direction of sensitivity (sign) of the explanatory parameters in the compensatory utility is found to be consistent both in the ownership and renting models but the sensitivity of several parameters are found to be significantly different in both models (e.g. commute distance, distance from CBD, etc.).



(a) Estimated values of penalty parameters

(b) Utility penalization

Figure 6-4 Impact of penalty terms on owners and renters choices

6.6 Validation results

Both the ownership and renting datasets are randomly divided into five rolling subsets. Each subset consists of 80% of the data for estimation and 20% for validation. Models are estimated for five estimation subsets of owners and five estimation subsets of renters. It is observed that the estimated parameters are consistent across the models estimated using the different subsets of the owner and renter data. Details of the parameter estimated using five estimation subsets are presented in Appendix N. Goodness of fit of the owners and renters models are presented in the summary Tables

6-3 and 6-4 respectively. From the tables, it is observed that the ICMNL model shows the higher goodness of fit in all subsets, both for ownership and renting (compare to the CMNL and the rCMNL models).

Table 6-3 Final log-likelihood of models estimated for estimation subsets of owners data

Subset	Number of Observations	Initial LL	Final LL		
			CMNL	rCMNL	ICMNL
Subset1	1500	-9315.90	-6150.51	-6149.05	-6033.88
Subset2	1500	-9315.90	-6175.52	-6175.94	-6072.84
Subset3	1500	-9315.90	-6150.31	-6151.03	-6037.04
Subset4	1500	-9315.90	-6175.38	-6174.81	-6056.92
Subset5	1500	-9315.90	-6129.68	-6129.43	-6018.58

Table 6-4 Final log-likelihood of the models estimated for estimation subsets of renters data

Subset	Number of Observations	Initial LL	Final LL		
			CMNL	rCMNL	ICMNL
Subset1	305	-1894.23	-1345.85	-1345.98	-1336.16
Subset2	305	-1894.23	-1326.31	-1326.15	-1311.82
Subset3	306	-1900.44	-1340.42	-1340.18	-1333.15
Subset4	306	-1900.44	-1334.85	-1334.75	-1325.68
Subset5	306	-1900.44	-1337.80	-1337.91	-1330.65

The five validation subsets (20% of the sample) are then used to validate the estimated model outcomes. The predictive power of each of the models is evaluated using both disaggregate level measures of fit (predictive rho-square and average probability of correct prediction) and aggregate level measures of fit (root mean square error and mean absolute deviation between predicted and actual share). Predictive measures of fit for all the models in different subsets are computed and summarized in Table 6-5 (owners subset) and Table 6-6 (renters subset) where the improvements in percentage over the CMNL model are presented in the parenthesis.

For owners, the ICMNL model shows improved performance over the CMNL and rCMNL models in all subsets in terms of all measures of fit. However, the performance of the rCMNL model is same as the CMNL model performance in most of the subsets and marginally different in some cases in terms of all measures of fit.

For renters, the ICMNL model performs better than the CMNL and the rCMNL models in all validation subsets in terms of the average probability of correct prediction, root means square error and mean absolute deviation between actual and predicted share. The ICMNL model performs worse than the CMNL and the rCMNL models in terms of predicted rho square in one out of five subsets (subset S1). This is also likely due to the fact that this specific subset may contain a high concentration of observations where households have a lower preference for the alternatives close to their current homes.

Table 6-5 Ownership model measures of fit in validation subsets

Validation Tools	Subsets	CMNL	rCMNL	ICMNL
Average probability of correct prediction	S1	0.027 (0)	0.028 (1.5)	0.032 (15.5)
	S2	0.028 (0)	0.028 (0)	0.033 (17.3)
	S3	0.028 (0)	0.028 (0)	0.033 (18.7)
	S4	0.029 (0)	0.030 (1.1)	0.034 (16.6)
	S5	0.03 (0)	0.03 (0)	0.035 (17.8)
Root mean square error between predicted and actual share(RMSE)	S1	1.208 (0)	1.202 (-0.5)	1.142 (-5.5)
	S2	1.385 (0)	1.385 (0)	1.285 (-7.2)
	S3	1.251 (0)	1.251 (0)	1.175 (-6.1)
	S4	0.887 (0)	0.886 (-0.1)	0.863 (-2.7)
	S5	1.54 (0)	1.54 (0)	1.384 (-10.1)
Mean absolute deviation between predicted and actual share (MAD)	S1	0.898 (0)	0.897 (-0.1)	0.83 (-7.6)
	S2	0.972 (0)	0.972 (0)	0.9 (-7.4)
	S3	0.974 (0)	0.974 (0)	0.927 (-4.8)
	S4	0.657 (0)	0.657 (0)	0.638 (-2.9)
	S5	1.17 (0)	1.17 (0)	1.051 (-10.1)
Predicted rho-sq	S1	0.31 (0)	0.311 (0.4)	0.319 (2.9)
	S2	0.311 (0)	0.311 (0)	0.325 (4.7)
	S3	0.311 (0)	0.311 (0)	0.325 (4.6)
	S4	0.322 (0)	0.323 (0.3)	0.337 (4.7)
	S5	0.312 (0)	0.312 (0)	0.322 (3.2)

Table 6-6 Renting model measures of fit in validation subsets

Validation Tools	Subsets	CMNL	rCMNL	ICMNL
Average probability of correct prediction	S1	0.029 (0)	0.029 (0)	0.03 (3.4)
	S2	0.022 (0)	0.022 (0)	0.025 (13.6)
	S3	0.024 (0)	0.024 (0)	0.026 (8.3)
	S4	0.02 (0)	0.02 (0)	0.022 (10)
	S5	0.022 (0)	0.022 (0)	0.025 (13.6)
Root mean square error between predicted and actual share(RMSE)	S1	1.93 (0)	1.929 (-0.1)	1.905 (-1.3)
	S2	1.965 (0)	1.965 (0)	1.956 (-0.5)
	S3	1.719 (0)	1.719 (0)	1.677 (-2.4)
	S4	1.668 (0)	1.668 (0)	1.665 (-0.2)
	S5	1.836 (0)	1.837 (0.1)	1.817 (-1)
Mean absolute deviation between predicted and actual share (MAD)	S1	1.445 (0)	1.444 (-0.1)	1.417 (-1.9)
	S2	1.529 (0)	1.529 (0)	1.487 (-2.7)
	S3	1.398 (0)	1.398 (0)	1.378 (-1.4)
	S4	1.299 (0)	1.299 (0)	1.283 (-1.2)
	S5	1.44 (0)	1.442 (0.1)	1.423 (-1.2)
Predicted rho-sq	S1	0.284 (0)	0.283 (-0.4)	0.282 (-0.7)
	S2	0.263 (0)	0.263 (0)	0.27 (2.7)
	S3	0.261 (0)	0.261 (0)	0.266 (1.9)
	S4	0.242 (0)	0.242 (0)	0.248 (2.5)
	S5	0.272 (0)	0.272 (0)	0.283 (4)

6.7 Conclusions

The CMNL model has attracted considerable interest in recent time due to the inapplicability of the classical Manski approach in the context of large universal choice sets. Since the CMNL model is not always a good approximation of the Manski model, a higher order formulation of the CMNL penalty term has been proposed in the rCMNL model for a better approximation of the Manski model. However, the rCMNL model has limitations in its ability to produce an improvement over the CMNL model when the universal choice set is very large. Therefore, this study proposes an improvement of the existing CMNL model (called ICMNL model) for behavioural choice set consideration to the classical Manski method. The performances of the ICMNL model is evaluated in this study using simulated data and applied to real-world residential location choice data. Models are estimated for residential ownership and renting submarkets. The performance of the methods is evaluated in terms of goodness of fit in the estimation sample and the predictive power of the estimated results in holdout sample validation. The key findings of this study are summarized here:

- In both ownership and renting models, the ICMNL model shows considerably better performance over the CMNL and the rCMNL models in estimation and sample validation.
- The rCMNL model could not produce any significant improvement over the CMNL. This is attributed to the large universal choice set in this study (498 alternatives).
- Modelling of residential location choice with the implicit choice set consideration also produces a behavioural difference in the choice set consideration of owners and renters. The estimated penalty parameters for ownership and renting models indicate that owners are more inclined to the alternatives close to their past home location than renters.

Although the ICMNL model is found to outperform the CMNL and rCMNL models, it still has avenues for further improvements. The threshold effect considered in the model for utility penalization is exogenous and homogeneous across all respondents. The method can be improved by allowing individual specific threshold or threshold specific for a group of respondents who belong to the same characteristics.

Further, the choice sets of all individuals in the proposed ICMNL model are constrained because utilities are penalized if alternatives do not meet the criteria of exogenous constraint. However, some individuals can have unconstrained choice sets. Adopting latent classes in the ICMNL model could be a potential direction for further improvement of the proposed ICMNL model where the choice set for one class can be constrained and choice set for another class can be unconstrained. However, this technique cannot be applied in this study due to data limitation (e.g. data about households social and other connections with the neighbourhood which may make their choice constrained to relocate close to their current location or in the same neighbourhood) and could be a future direction of research. The constrained attribute (distance from past home) is used in the consideration part only, although it may influence the preference as well. Considering the same attribute (e.g. distance from past home) in the consideration part (in the penalty term) and the systematic part of the utility has resulted in an identification issue. The ICMNL model with latent classes for the individuals of constrained and unconstrained choice sets may avoid this issue. However, this technique cannot be applied in this study due to data limitation as mentioned before and could be a future direction of research.

Distance of alternatives from the past home and commute distance are considered in this study to simulate choice set implicitly and finally, distance from the past home is found to be significant. However, other parameters like dwelling cost and dwelling size could also have an influence on the individual choice set. For example, a household having many members may not consider small houses in their choice set. Since aggregate level models have been estimated in this study due to data limitations, the scope was limited to testing a finite number of hypotheses. However, estimation of dwelling level models (which is possible due to the availability of housing supply data in many metropolitan cities) will offer more flexibility to test different hypotheses to capture individual choice set.

Although the potential of the proposed method (ICMNL) that observed in this study to capture individual choice set is promising, more testing is recommended with other data sets as a topic of future research. Testing the validity of the findings in other contexts (e.g., route choice, destination choice, activity choice, etc.) can also be an interesting direction for future research.

However, with better behavioural grounding (supported by the better model fit) as well as computational tractability, the proposed ICMNL model can be an attractive option for modelling with a large universal choice set where the classical probabilistic approach is infeasible.

Chapter 7

Discussion and conclusions

7.1 Summary of the research

Residential decision and car ownership are long term household or individual level decisions that have a close association with travel patterns such as mode choice, trip length, trip frequency, etc. All these interrelated decisions are required to be properly addressed and considered in integrated urban modelling and planning for sustainable policy formulation. Although a large body of literature has attempted to model the household level residential location, car ownership and travel decisions and their interrelationships (Clark and Withers, 1999; Zhou and Kockelman, 2008; Zolfaghari, 2013; Clark et al., 2016a; Clark et al., 2016b), several scopes are identified for modelling these interconnected household level decisions with better behavioural underpinning. Unless the behaviours of the households or individuals are addressed reasonably, the goal of policy implication for sustainable development will be challenging to achieve. Dependable data can be a challenging issue in this context. While RP data is expected to be best suited for capturing the household true behaviour, missing information poses significant modelling challenges for producing dependable model outputs.

Household residential choice is assumed as a two-tier decision process consisting of (a) decision to move or stay and (b) decision of location. The decision to move (or stay) is the upper layer which is mostly driven by life events (such as getting a job, getting married, having baby, etc.), changes in preferences and dissatisfaction in the current place. The stressors for decision to move also determine the geographical scale of the decision such as whether moving in the same area or moving in a different area (another city, metropolitan or country). For example, securing a job in another metropolitan city may require a long-distance move whereas moving for extra space is most likely to happen in the same neighbourhood. This decision is unlikely to depend significantly on the characteristics of neighbourhood households are anticipating to move[26].

Investigating the geographical scales of residential mobility is crucial for housing market analysis and any associated policy formulation (discussed in chapter 3 in detail). In addition, the geographical scales of residential mobility are likely to affect household car ownership and travel behaviours differently. Although investigating the geographical scales in residential mobility decision has significance importance, the existing literature is limited in this context. Therefore, this study has captured geographical scales of residential mobility quantifying the factors affecting the relocation decision at the local, regional or national levels. Mixed multinomial logit (MMNL) models are estimated using 18 years of long panel survey data (British Household Panel Survey) for capturing the dynamics in the life-trajectory decision process. A significant level of behavioural insights in terms of moving home in different geographical scales is observed. The characteristics of the households moved in different geographical scales are observed to be considerably different from one another. For example, the social renters are found very unlikely to move out from the current metropolitan areas whereas the private renters are likely to move within and across the metropolitan areas. Most importantly, parameter sensitivities estimated in the model without considering the geographical scale of relocation are found considerably different from corresponding geographical scale specific sensitivities in many cases. Therefore, capturing household mobility decision without considering its geographical scale is unlikely to produce true parameter sensitivities and most likely to give poor performance in forecasting and policy analysis.

The first layer of the residential choice discussed in the previous section captured decision to move at different geographical scales within the UK. The second layer captured which areas or neighbourhoods within a specific geographical scale household moved. The analysis of the preferences for location or neighbourhood is usually performed for a single housing market such as a town, city, or metropolitan area (e.g. Habib and Miller, 2009; Zolfaghari et al., 2012) which a consumer considers

[26] The characteristics of alternate dwelling and neighbourhood determine the choice of location but can also influence the decision to move in few cases. These are usually short distance relocations, but if households do not find a suitable dwelling or neighbourhood, they may make a long-distance relocation and even change the decision of relocation altogether.

to relocate. The larger is the size of the study area, there are more challenges involved in terms of data acquisition, level of aggregation, assigning the appropriate choice set, correlation structure and computational burden. Since residential mobility is a rare event, the number of households in BHPS (covered the whole UK) who moved in a single metropolitan area is very few, below a hundred in case of London. This small number of observations at the metropolitan level was not enough for testing the research hypothesis. Therefore, another dataset was required to investigate the research questions related to location or neighbourhood preference of the households who moved.

London household survey data (LHSD) is found useful in this regard along with the Ward atlas dataset (WAD) and origin-destination matrix from London Transport Studies Model (LTSM) for location, land use and transport data. Combining these datasets was challenging because the lower level of geographical identifiers was not uniform across the datasets. Combing the different data sets applying GIS map matching technique enhances the opportunity for capturing a wide array of parameters ranges from the dwelling, land use and transport characteristics for more robust analysis. The most important behavioural issue in the choice of residential location that has been investigated in this study is the preference heterogeneity of two major housing markets such as ownership and renting. For modelling, it is assumed that households chose locations (ward) which they perceived to be the best from a set of 498 alternative locations (wards) in GLA. Obviously, this is a poor assumption, therefore an alternative technique is also proposed in this study to capture the behaviourally persuasive choice set which is discussed in the next paragraph. The model with the full choice sets (discussed in chapter 5) presents a significant level of behavioural differences in the residential location choice preferences of owners and renters.

As mentioned in the previous section, consideration of the full choice set for individual households in chapter 5 can be challenged from a behavioural point of view. Rather households are more likely to consider a reduced choice set based on their circumstances. To capture the underlying mechanism of choice set consideration in residential location choice context where the number of alternatives is very high, the classical probabilistic approach is infeasible and elimination based approaches have limitations and criticised in the literature (Zolfaghari, 2013). Therefore, the performances of exiting semi compensatory approaches are revisited in chapter 6 of this study theoretically and with a practical example. The weakness of the existing semi-compensatory methods to model the choice set consistently with the classical probabilistic approach is identified. Therefore, this research proposes a theoretical improvement of exiting constrained multinomial logit (CMNL) model (called improved constrained multinomial logit model, ICMNL) for a better approximation of classical probabilistic approach. The proposed ICMNL model is applied to re-estimate the residential location choice models estimated in chapter 5 considering the full choice set[27]. The performance of the ICMNL model over the other available methods is investigated. The ICMNL model outperforms over other semi-compensatory approaches both in estimation and prediction. Therefore, it can be said that the proposed technique offers an improvement over the state of art semi compensatory approach of modelling reflecting the better power of capturing the underlying choice set consideration mechanism.

Household residential decision discussed in the preceding sections can influence household car ownership change and travel mode switching behaviours. It is also anticipated that the geographical scales of residential decision might have varying impacts on these behaviours. Therefore, this study also investigated the key drivers of car ownership change and commute mode switching behaviours with a spatial focus on the role of the geographical scale of residential decision on these behaviours. Since car ownership change and travel mode switching are time dependent household behaviours, cross-sectional data like LHSD is not appropriate for investigating these behaviours, therefore, the BHPS dataset is used instead.

In the car ownership model, different directions of car transaction such as switching from non-car ownership to car ownership (zero to one or more cars), car ownership to non-car ownership (one or more cars to zero car), acquisition or disposal of additional cars (second or third cars) are captured. All these directions of switching are captured in a single model whereas most of the existing studies have captured each direction of switching (from one level to the next level) as a binary choice in a separate model

[27] The output of the ICMNL models estimated in Chapter 6 cannot be directly compared with the outcomes of the corresponding models estimated in Chapter 5 to see the improvement in terms of capturing the behavioural choice set. There are two reasons a. The parameters used for choice set simulation in chapter 6 has not been considered in chapter 5. b. ICMNL model in chapter 6 has IIA restriction which is relaxed in the models estimated in chapter 5. More details are given in section 7.3.

(Oakil, 2013; Clark et al., 2016a). MMNL estimation allows to capture the correlation of repeated choice and random taste heterogeneity across the individuals. It is observed from the model outcome that the geographical scale of residential mobility has a strong connection with car ownership level changes. For example, households who have moved in other regions or metropolitan areas are found to be more inclined to acquire car(s) compared to the households who have moved within the regions or metropolitan areas. Household socio-demographic characteristics, life events and travel behaviour are also found to influence their car ownership changes.

Similar to the car ownership change model, commute mode switching model also captures different directions of switching (e.g. switching from car to public transport, car to active travel, public transport to car, etc.) in a single model, although the previous studies captured the binary decision of switching behaviour separately (Clark et al., 2016b; Fatmi and Habib, 2017). From the model output, the role of the geographical scale of residential relocation is found significant on travel mode switching behaviour as well. For example, households moved at a national level are found more likely for switching to car, on the other hand, households moved at the regional level are more inclined for switching to public transport. Household car ownership, travel distance and job change are also found to control household travel mode transition behaviour.

7.2 Contributions of this research

The major contribution of this research is the estimation of a richer set of model components of integrated urban modelling system that captures the complexities and interdependencies among the residential mobility, residential location, car ownership change and commute mode switching behaviour. Uses of revealed preference data in this research contributed for capturing the true behaviour avoiding the hypothetical bias in the stated preference data. Uses of long panel data also allowed to observe people for a very long time in terms of capturing temporal and long-term dynamics and correlation across the repeated choices.

The specific contributions of this study are listed below

7.2.1 Methodological contributions

- Proposes a comprehensive approach for capturing multi-directional transition behaviour in car ownership and commute mode choice in a single econometric model.
- Develops advanced econometric models to capture the dynamics in the interconnected household level decisions (residential mobility, car ownership and commute mode change) and the associated unobserved heterogeneities among the households.
- Critically analyses the strength and weakness of the existing semi-compensatory choice set construction approaches.
- Proposes an improvement of an existing semi-compensatory approach which in behaviourally more persuasive and empirically tractable for modelling with the large choice set.

7.2.2 Applied contributions

- Demonstrates the importance of considering the geographical scale of relocation in modelling household mobility decision.
- Explores the role of the geographical scale of residential decision on household car ownership and travel mode switching behaviours.
- Offers important behavioural insights in terms of preferences of different housing submarkets (ownership and renting) in their residential location choices.
- Demonstrates the importance of choice set formation for unbiased parameter estimation.
- Unveils the underlying preferences in choice set consideration for residential ownership and renting decisions.

The empirical findings of this study have the potential for important policy contributions. For integrated transport and land-use modelling, planning and policymaking, a proper understanding of the interconnected household level decisions is crucial, this study can be helpful in this regard. Any policy for society requires a certain level of social stability which can be significantly affected by a high residential mobility rate. This research findings can be used for predicting residential mobility patterns in different geographical scales from household sociodemographic and life events. The insights from this study can also guide for some potential policy steps to minimize the high residential turnover rate but the differences between the geographical scales of relocation should be acknowledged. From this study outcome, it is observed that people are keen to live close to work location to minimize commute distance and are very inclined towards better public transport accessibility. Therefore, areas with low public transport accessibility may have a higher rate of residential turnover. Policies to increase public transport accessibility in the deprived area may decrease the residential turnover rate and can also reduce people's car dependency. Since the mixed type of land development is preferred by the people, policies for promoting this type of development can also reduce car dependency and total vehicle miles travelled. People who have long distance relocation (e.g. national level) are more likely to be car dependent which could be a result of the unfamiliarity with the new area or less time investment for finding a suitable location close to the workplace. Therefore, the policy can force the employer to provide housing facilities for the new employees who moved from another city. More detailed discussions about the policy implications of this study are presented in the corresponding chapters.

7.3 Future research directions

Based on the discussion in the above sections, this study addresses several behavioural and methodological issues in different important components of the research framework. The two rich RP datasets used in this research open the scope of capturing true behaviour. The panel nature of the BHPS dataset allows to capture the dynamics in the life trajectory such as how changes in household state influence their residential, car ownership and travel decisions, correlation among the repeated choices, the role of the previous choices, etc. However, the discrete choice model has a limitation in terms of capturing the duration dynamics of time dependent household behaviour. Therefore, the future study can look for a more appropriate approach (e.g. hazard based model) that allows to capture duration dynamics in the household behaviour.

The causalities among the decision components modelled in this research are very complex. Although this study modelled the dominant directions of dependencies of household decisions sequentially (e.g. impact of residential decision on car ownership), the relation could have reverse causalities (impact of car ownership on residential decision) or people can consider multiple decisions simultaneously. In addition, some other household or individual level decisions such as tenure choice, work location or employer choice, etc. can be associated with the decisions modelled in this study. The future study can find a comprehensive modelling framework and suitable data to capture the interdependencies of the multiple choices and the decision simultaneity if applicable.

In the sequential approach, one decision such as residential mobility decision has been used as an independent variable to explain other decisions such as car ownership change and travel mode switching. In this case, residential mobility decision is more likely to be endogenous. Therefore, it is necessary to find an alternative approach to avoid endogeneity bias or endogeneity correction is needed in the sequential approach used in this study. This study was limited to handle this issue.

Household car ownership and travel behaviours are most likely to be correlated with neighbourhood characteristics (Clark et al., 2016a; Clark et al., 2016b). These characteristics have not been tested in the car ownership and travel mode transition models developed in this study due to data limitation. The developed car ownership change model and travel mode switching model thus can be improved further by incorporating neighbourhood characteristics such as transport accessibility, shopping accessibility, parking facilities, distance from the employment centre, etc. if data is available.

Although this study uses a long panel data (BHPS) for modelling the upper layer of residential decision (decision to move or stay), this rich dataset cannot be used for modelling residential location or neighbourhood (the second layer of residential decision) choice due to the limited number of observations for an area. Therefore, these two layers of residential decision have been modelled separately using two separate datasets. Although, these independent models provide several behavioural insights but limited in terms of capturing the association of these two layers of residential decisions. In many cases, people mobility decision is independent of the characteristics of the new location they are attempting to move. For instance, if someone finds a suitable job in a new city, (s)he is obliged to move where relocation decision may not be conditional to the characteristics of the new area. However, residential mobility decision can depend on the characteristics of the new location in some cases. For example, households may only consider to move if they find suitable properties or neighbourhoods. Two separate models of residential mobility and location choice cannot capture the connection between the two layers of residential decision. Future study can focus on addressing this issue.

In this study, zone level residential location choice models are estimated where each zone (or ward) is considered as location alternative. However, the zone level model has a limitation in terms of capturing the variability at the level of dwelling (Zolfaghari, 2013). For example, within a zone, a dwelling close to the public transport access point is likely to be more attractive compared to the other dwellings far from that point. If dwelling supply data is available, estimation of the dwelling level model can be an interesting direction for future research.

The full choice set is assigned for individual households for modelling residential location choice in chapter 5 ignoring the effect of the consideration set. Chapter 6 has accounted for the consideration effect where the distance from past home is used to simulate the consideration set implicitly. However, this parameter may have the power of explaining the choice as well but could not be used as an explanatory variable due to an identification issue. Since the role of the parameter 'past home distance' for explaining the choice and choice set could not be separated, it may not be a fair comparison between the model finding in chapter 5 and chapter 6 for exploring how much in the gain in terms of capturing the consideration set. Moreover, the models in chapter 6 hold IIA (Independence of Irrelevant Alternatives) restriction, therefore, cannot be directly compared with the models estimated in chapter 5 (mixed logit model where IIA restriction is relaxed). The future direction of research can consider the same specification in the models with consideration effect and without consideration effect to investigate the gain in model fit in terms of capturing the consideration set.

The proposed ICMNL model in chapter 6 has two avenues of further improvement. The threshold effect considered in the model for utility penalization is exogenous and homogeneous across all respondents. The method can be improved for allowing individual specific threshold or threshold specific to the group of respondents belonging to the same behaviour. The proposed ICMNL model assumes the non-compensatory behaviour of all respondents because the choice sets of individuals are constrained by specific criteria and utilities are penalized if alternatives do not meet the criteria of exogenous constraint. However, some individuals can have unconstrained choices set and can play a compensatory role for utility maximization.

The ICMNL model can be improved to accommodate both compensatory and non-compensatory behaviour. Probabilistic partitioning of the sample for households having a constrained and unconstrained choice set could be a potential direction of further improvement of the proposed ICMNL model.

7.4 Concluding remarks

Household level decisions are interconnected and changes dynamically depending on the personal circumstances and the surrounding environment. They are, therefore, difficult to model. This research attempts to model some of these long-term and short-term interconnected household decisions (residential choice, car ownership change and travel mode switching) which will help in a better understanding of household behaviour in response to alternative land use and transport policy context.

References

Aditjandra, P. T., Cao, X. J. and Mulley, C. (2012). Understanding neighbourhood design impact on travel behaviour: An application of structural equations model to a British metropolitan data. *Transportation Research Part A: Policy and Practice,* 46, 22-32.

Adnan, M., Pereira, F. C., Azevedo, C. M. L., et al. (2016). Simmobility: A multi-scale integrated agent-based simulation platform. *95th Annual Meeting of the Transportation Research Board Forthcoming in Transportation Research Record.*

Aero, T. (2006). Residential choice from a lifestyle perspective. *Housing, Theory and Society,* 23, 109-130.

Alexander, D. and Tomalty, R. (2002). Smart growth and sustainable development: Challenges, solutions and policy directions. *Local Environment,* 7, 397-409.

Anderson, L. and Fricker Jr, R. D. (2015). Raking: An important and often overlooked survey analysis tool. *Phalanx,* 48, 36-42.

Arundel, R. and Ronald, R. (2017). The role of urban form in sustainability of community: The case of Amsterdam. *Environment and Planning B: Urban Analytics and City Science,* 44, 33-53.

Auld, J., Hope, M., Ley, H., et al. (2016). POLARIS: Agent-based modeling framework development and implementation for integrated travel demand and network and operations simulations. *Transportation Research Part C: Emerging Technologies,* 64, 101-116.

Baar, K. (1992). The national movement to halt the spread of multifamily housing, 1890-1926. *Journal of the American Planning Association,* 58, 39-48.

Bartel, A. P. (1979). The migration decision: What role does job mobility play? *The American Economic Review,* 69, 775-786.

Beige, S. and Axhausen, K. W. (2012). Interdependencies between turning points in life and long-term mobility decisions. *Transportation,* 39, 857-872.

Bell, M. (2002). Comparing population mobility in Australia and New Zealand. *Journal of Population Research,* (Special ed. 2002),169.

Bell, M. G. (2007). Mixed routing strategies for hazardous materials: Decision-making under complete uncertainty. *International Journal of Sustainable Transportation,* 1, 133-142.

Bell, M. and Rees, P. (2006). Comparing migration in Britain and Australia: harmonisation through use of age–time plans. *Environment and Planning A*, 38(5), 959-988.

Ben-Akiva, M., Bradley, M., Morikawa, T., Benjamin, J., Novak, T., Oppewal, H., & Rao, V. (1994). Combining Revealed and Stated Preferences Data. *Marketing Letters, 5*(4), 335-349.

Berry, S., Levinsohn, J., Pakes, A. (1995). Automobile prices in market The migration decision: What role does job mobility play?equilibrium. Econometrica: J. Econ. Soc., 841–890.

Bhat, C. R. (2015). A comprehensive dwelling unit choice model accommodating psychological constructs within a search strategy for consideration set formation. *Transportation Research Part B: Methodological,* 79, 161-188.

Bhat, C. R. and Guo, J. (2004). A mixed spatially correlated logit model: formulation and application to residential choice modeling. *Transportation Research Part B: Methodological,* 38, 147-168.

Bhat, C. R. and Guo, J. Y. (2007). A comprehensive analysis of built environment characteristics on household residential choice and auto ownership levels. *Transportation Research Part B: Methodological*, 41, 506-526.

Bhat, C. R., Sen, S. and Eluru, N. (2009). The Impact of Demographics, Built Environment Attributes, Vehicle Characteristics, and Gasoline Prices on Household Vehicle Holdings and Use. *Transportation Research Part B: Methodological*, 43 (1), 1–18.

Bhat, C. R., and Pulugurta, V. (1998). A Comparison of Two Alternative Behavioral Choice Mechanisms for Household Auto Ownership Decisions. *Transportation Research Part B: Methodological,* 32 (1), 61–75.

Bierlaire, M., Hurtubia, R. and Flotterod G. (2010). Analysis of implicit choice set generation using a constrained multinomial logit model. *Transportation Research Record: Journal of the Transportation Research Board*, 92-97.

Bill, A., Mitchell, W. and Watts, M. (2006). Examining the relationship between commuting patterns, employment growth and unemployment in the NSW Greater Metropolitan Region. *Australian Journal of Social Issues*, 41, 233-245.

Boyko, C. T. and Cooper, R. (2011). Clarifying and re-conceptualising density. *Progress in Planning,* 76, 1-61.

Bradley, M.A., Daly, A.J., 1997. Estimation of logit choice models using mixed stated preference and revealed preference information. In: Stopher, P.R., Lee-Gosselin, M. (Eds.), Understanding Travel Behaviour in an Era of Change. Pergamon, Oxford, pp. 209–232.

Brown, L. A. and Moore, E. G. (1971). The Intra-urban Migration Process: A Perspective. In L.S. Bourne (ed.), *Internal Structure of the City: Readings on Space and Environment*, New York: Oxford University Press, 200-209.

Brown, B. B. and Werner, C. M. (2008). Before and after a new light rail stop: resident attitudes, travel behavior, and obesity. *Journal of the American Planning Association,* 75, 5-12.

Brownstone, D., Bunch, D.S. and Train, K., 2000. Joint mixed logit models of stated and revealed preferences for alternative-fuel vehicles. *Transportation Research Part B: Methodological*, *34*(5), pp.315-338.

Bun, M. J. and Sarafidis, V. (2015). Dynamic panel data models. *The Oxford handbook of panel data*, 76-110.

Burda, C. (2014). Home Location Preference Survey: Understanding where GTA residents prefer to live and commute. Royal Bank of Canada and the Pembina Institute.

Bwambale, A., Choudhury C. F. and Hess, S., 2017. Modelling trip generation using mobile phone data: a latent demographics approach. *Journal of Transport Geography,* 76, 276-286.

Cadwallader, M. (1992). Migration and Residential Mobility: Macro and Micro Approaches. Madison, Wisconsin: The University of Wisconsin Press.

Caicedo, F., Lopez-Ospina, H. and Pablo-Malagrida, R. (2016). Environmental repercussions of parking demand management strategies using a constrained logit model. *Transportation Research Part D: Transport and Environment,* 48, 125-140.

Cao, X., Mokhtarian, P. L. and Handy, S. L. (2007). Do changes in neighborhood characteristics lead to changes in travel behavior? A structural equations modeling approach. *Transportation,* 34, 535-556.

Capuano, S. (2011). The South–North mobility of Italian college graduates. an empirical analysis. *European Sociological Review,* 28, 538-549.

Carrasco, J. A., & de Dios Ortúzar, J. (2002). Review and assessment of the nested logit model. *Transport Reviews, 22*(2), 197-218.

Cascetta, E. and Papola, A. (2001). Random utility models with implicit availability/perception of choice alternatives for the simulation of travel demand. *Transportation Research Part C: Emerging Technologies*, 9, 249-263.

Cascetta, E. and Papola, A. (2009). Dominance among alternatives in random utility models. *Transportation Research Part A: Policy and Practice*, 43, 170-179.

Castro, M., Martínez, F. and Munizaga, M. A. (2013). Estimation of a constrained multinomial logit model. *Transportation*, 40, 563-581.

Champion, T. (2015) What do the latest official subnational population projections suggest for Great Britain's 63 cities? *Foresight, Government Office for Science*.

Chen, J., Chen, C. and Timmermans, H. (2008). Accessibility trade-offs in household residential location decisions. *Transportation Research Record: Journal of the Transportation Research Board*, 71-79.

Cherchi, E., Guevara, C.A. (2012). A Monte Carlo experiment to analyze the curse of dimensionality in estimating random coefficients models with a full variance–covariance matrix. *Transport. Research Part B: Methodology*, 46 (2), 321–332.

Chevan, A., 1971. Family growth, household density, and moving. *Demography*, *8*(4), 451-458.

Choudhury, C. F. and Ayaz, S. B. (2015). Why live far?—Insights from modeling residential location choice in Bangladesh. *Journal of Transport Geography*, 48, 1-9.

Clark, B., Chatterjee, K. and Melia, S. (2016a). Changes in level of household car ownership: the role of life events and spatial context. *Transportation*, 43, 565-599.

Clark, B., Chatterjee, K. and Melia, S. (2016b). Changes to commute mode: The role of life events, spatial context and environmental attitude. *Transportation Research Part A: Policy and Practice*, 89, 89-105.

Clark, B., Chatterjee, K., Melia, S., et al. (2014). Life events and travel behavior: exploring the interrelationship using UK household longitudinal study data. *Transportation Research Record: Journal of the Transportation Research Board*, 54-64.

Clark, B., Lyons, G. and Chatterjee, K. (2012). Exploring the interactions between life events, neighbourhood choice and car ownership transitions: Insights from a retrospective longitudinal survey. *44th Universities' Transport Studies Group Conference*.

Clark, W. A. (2013). Life course events and residential change: Unpacking age effects on the probability of moving. *Journal of Population Research*, *30*(4), 319-334.

Clark, W. A., Deurloo, M. C. and Dieleman, F. M. (1986). Residential mobility in Dutch housing markets. *Environment and Planning A*, 18, 763-788.

Clark, W. A. and Withers, S. D., (1999). Changing jobs and changing houses: mobility outcomes of employment transitions. *Journal of regional science*, *39*(4), 653-673.

Clark, W. A., Deurloo, M. C. and Dieleman, F. M., (1984). Housing consumption and residential mobility. *Annals of the association of American geographers*, *74*(1), 29-43.

Clark, W.A. and Drever, A. I. (2000). Residential mobility in a constrained housing market: implications for ethnic populations in Germany. *Environment and Planning A*, 32(5), 833-846.

Clark, W. A. and Huang, Y. (2003). The life course and residential mobility in British housing markets. *Environment and Planning A*, 35, 323-339.

Clark, W. and Huang, Y. (2004). Linking migration and mobility: Individual and contextual effects in housing markets in the UK. *Regional studies*, *38*(6), 617-628

Clark, W.A. and Ledwith, V. (2006). Mobility, housing stress, and neighborhood contexts: evidence from Los Angeles. *Environment and Planning A*, *38*(6), 1077-1093.

Coulter, R. and Scott, J. (2015). What motivates residential mobility? Re-examining self-reported reasons for desiring and making residential moves. *Population, Space and Place*, 21, 354-371.

Curran, C., Carlson, L. A. and Ford, D. A. (1982). A theory of residential location decisions of two-worker households. *Journal of Urban Economics*, 12, 102-114.

Daly, A., Hess, S. and de Jong, G., 2012. Calculating errors for measures derived from choice modelling estimates. *Transportation Research Part B: Methodological*, *46*(2), pp.333-341.

Dargay, J. and Hanly, M. (2007). Volatility of car ownership, commuting mode and time in the UK. *Transportation Research Part A: Policy and Practice,* 41, 934-948.

Dargay, J. M. (2001). The effect of income on car ownership: evidence of asymmetry. *Transportation Research Part A: Policy and Practice,* 35, 807-821.

David, Q., Janiak, A. and Wasmer, E. (2010). Local social capital and geographical mobility. *Journal of Urban Economics,* 68, 191-204.

De La Barra, T., Pérez, B. and Vera, N. (1984). TRANUS-J: putting large models into small computers. *Environment and Planning B: Planning and Design,* 11, 87-101.

de Luca, S. and Cantarella, G. E., 2016. Validation and comparison of choice models. In *Travel Demand Management and Road User Pricing*, 57-78.

Dieleman, F. (2017). Households and housing: Choice and outcomes in the housing market. *Routledge*.

Dieleman, F. M., Clark, W. A. and Deurloo, M. C. (2000). The geography of residential turnover in twenty-seven large US metropolitan housing markets, 1985-95. *Urban studies,* 37, 223-245.

Duke-Williams, O. (2009). The geographies of student migration in the UK. *Environment and Planning A,* 41, 1826-1848.

Echenique, M. H., Flowerdew, A. D., Hunt, J. D., et al. (1990). The MEPLAN models of Bilbao, Leeds and Dortmund. *Transport Reviews,* 10, 309-322.

Eluru, N., Sener, I. N., Bhat, C. R., et al. (2009). Understanding residential mobility: joint model of the reason for residential relocation and stay duration. *Transportation Research Record,* 2133, 64-74.

Ettema, D., de Jong, K., Timmermans, H., et al. (2007). PUMA: multi-agent modelling of urban systems. *Modelling land-use change.* Springer, 237-258.

Ewing, R. and Cervero, R. (2010). Travel and the built environment: a meta-analysis. *Journal of the American Planning Association,* 76, 265-294.

Fack, G. and Grenet, J. (2010). When do better schools raise housing prices? Evidence from Paris public and private schools. *Journal of public Economics,* 94, 59-77.

Fagnant, D. J. and Kockelman, K. (2015). Preparing a nation for autonomous vehicles: opportunities, barriers and policy recommendations. *Transportation Research Part A: Policy and Practice,* 77, 167-181.

Farber, S. and Li, X. (2013). Urban sprawl and social interaction potential: an empirical analysis of large metropolitan regions in the United States. *Journal of Transport Geography,* 31, 267-277.

Farooq, B. and Miller, E. J. (2012). Towards integrated land use and transportation: A dynamic disequilibrium based microsimulation framework for built space markets. *Transportation Research Part A: Policy and Practice,* 46, 1030-1053.

Fasbender, D., Devos, W. and Lemajic, S. (2017). Analysis of the sampling representativeness for the Land Parcel Identification System Quality Assurance, EUR 28755 EN, *Publications Office of the European Union, Luxembourg*, ISBN 978-92-79-73181-5, doi:10.2760/458367, JRC107675.

Fatmi, M. R. (2017). Development of Integrated Urban System Models: Population Life-transitions, Household location, and Vehicle Transaction Decisions. *PhD dissertation, University of Dalhousie*.

Fatmi, M. R. and Habib, M. A. (2017). Modelling mode switch associated with the change of residential location. *Travel behaviour and society*, 9, 21-28.

Ferreira, F. (2010). You can take it with you: proposition 13 tax benefits, residential mobility, and willingness to pay for housing amenities. *Journal of Public Economics,* 94 (9), 661–673.

Fisher, L. D. and Lin, D. Y. (1999). Time-dependent covariates in the Cox proportional-hazards regression model. *Annual review of public health, 20*(1), 145-157.

Fitzgerald, J., Gottschalk, P. and Moffitt, R. A. (1998). An analysis of sample attrition in panel data: The Michigan Panel Study of Income Dynamics. National Bureau of Economic Research Cambridge, Mass., USA.

Floor, H., Van Kempen, R. and de Vocht, A. (1996). Leaving Randstad Holland: An analysis of housing preferences with decision plan nets. *Netherlands journal of housing and the built environment,* 11, 275-296.

Florida, R. (2002). The economic geography of talent. *Annals of the Association of American geographers,* 92, 743-755.

Fotheringham, A., Brunsdon, C., and Charlton, M. (2000). Quantitative geography: perspectives on spatial data analysis. London: Sage Publications Ltd.

Fotini, T., Evangelia, V. and Michail, V. (2013). Weighting of responses in the Consumer Survey: Alternative approaches – Effects on variance and tracking performance of the Consumer Confidence Indicator. *Foundation for economic and industrial research*.

Fox, J., Patruni, B., Daly, A., et al. (2017). Estimation of the National Car Ownership Model for Great Britain.

Frank, L. D., Andresen, M. A. and Schmid, T. L. (2004). Obesity relationships with community design, physical activity, and time spent in cars. *American journal of preventive medicine,* 27, 87-96.

Frenkel, A., Bendit, E. and Kaplan, S. (2013). Residential location choice of knowledge-workers: The role of amenities, workplace and lifestyle. *Cities,* 35, 33-41.

Fulton, L. M., Noland. R. B., Meszler, D. J., et al. (2000). A statistical analysis of induced travel effects in the US mid-Atlantic region. *Journal of Transportation and statistics,* 3, 1-14.

Ghasri, M., Rashidi, T. H. and Saberi, M. (2018). Comparing Survival Analysis and Discrete Choice Specifications Simulating Dynamics of Vehicle Ownership. *Transportation Research Record*, *2672*(49), 34-45.

Goodwin, P., Cairns ,S., Dargay, J., et al. (2004). Changing travel behaviour.

Greater London Authority (2003). The 2002 London Household Survey, Data Management and Analysis Group (DMAG).

Griffin, J., et al. (2015). Understanding participant representativeness in deliberative events: A case study comparing probability and non-probability recruitment strategies. *Journal of Public Deliberation,* 11(4),1–26

Gower ,T. L. (2014). 2011 Census Analysis-Distance Travelled to Work.

Guevara, C. A. (2010). Endogeneity and sampling of alternatives in spatial choice models. *Doctoral dissertation, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA.*

Guevara, C. A. (2015). Critical assessment of five methods to correct for endogeneity in discrete-choice models. *Transportation Research Part A: Policy and Practice*, 82, 240-254.

Guevara, C. A., Ben-Akiva, M. (2006). Endogeneity in residential location choice models. *Transportation Research Record*, 1977 (1), 60–66.

Guevara, C., Ben-Akiva, M. (2012). Change of scale and forecasting with the control-function method in logit models. *Transportation Science*, 46 (3), 425–437.

Guevara, C. A., Polanco, D. (2015). Correcting for endogeneity without instruments in discrete choice models: the multiple indicator solution (MIS). Working Paper, Universidad de los Andes.

Guo, J. and Bhat, C. (2002). Residential location choice modeling: Accommodating sociodemographic, school quality and accessibility effects. *Austin: Department of Civil Engineering, Universidade do Texas*.

Habib, M. A. (2009). Microsimulating residential mobility and location choice processes within an integrated land use and transportation modelling system. University of Toronto, Ontario, Canada.

Habib, K. M. N. and Kockelman, K. (2008). Modeling choice of residential location and home type: recent movers in Austin Texas. *87th Annual Meeting of the Transportation Research Board, Washington, DC.*

Habib, M. and Miller, E. (2009). Reference-dependent residential location choice model within a relocation context. *Transportation Research Record: Journal of the Transportation Research Board*, 92-99.

Hanly, M. and Dargay, J. (2000). Car ownership in Great Britain: panel data analysis. *Transportation Research Record: Journal of the Transportation Research Board*, 83-89.

Haque, M. B., Choudhury, C. and Hess, S. (2018). Investigating the Temporal Dynamics of Long-Term and Medium-Term Residential Location Choices: A Case-Study of London.

Heckman, J. (1978). Dummy endogenous variables in a simultaneous equation system. *Econometrica*, 46, 931–959.

Hensher, D.A., Bradley, M., 1993. Using stated response data to enrich revealed preference discrete choice models. Marketing Letters 4 (2), 139–152.

Hensher, D. A., Rose, J. M., & Greene, W. H. (2015). *Applied Choice Analysis* (2 ed.). Cambridge: Cambridge University Press.

Hensher, D. A. and Taylor, A. K. (1983). Intraurban residential relocation choices for students: an empirical inquiry. *Environment and Planning A*, 15(6), 815-830.

Hensher, D. A. and Ton, T. (2002). TRESIS: A transportation, land use and environmental strategy impact simulator for urban areas. *Transportation,* 29, 439-457.

Hess, S., Ben-Akiva, M., Gopinath, D. (2009). Taste heterogeneity, correlation and elasticicities in latent class choice models. Paper presented at the International Choice Modelling Conference 2009, Harrogate.

Hess, S. and Train, K., 2017. Correlation and scale in mixed logit models. *Journal of Choice Modelling*, *23*, pp.1-8.

Ho, C. Q., Hensher, D. A., & Ellison, R. (2017). Endogenous treatment of residential location choices in transport and land use models: Introducing the MetroScan framework. *Journal of Transport Geography, 64*, 120-131.

Ho, C. Q., & Mulley, C. (2013). Multiple purposes at single destination: A key to a better understanding of the relationship between tour complexity and mode choice. *Transportation Research Part A: Policy and Practice, 49*(0), 206-219.

Hocherman, I., Prashker, J. N. and Ben-Akiva, M., 1983. Estimation and use of dynamic transaction models of automobile ownership (No. 944).

Horowitz, J. L. and Louviere, J. J. (1995). What is the role of consideration sets in choice modeling? *International Journal of Research in Marketing*, 12, 39–54.

Hoshino, T. (2011). Estimation and analysis of preference heterogeneity in residential choice behaviour. *Urban studies*, 48, 363-382.

Hunt, J. D. (2003). Design and application of the pecas land use modelling system. 8th International Conference on Computers in Urban Planning and Urban Management (CUPUM).

Ibraimovic, T. and Hess, S. (2017). Changes in the ethnic composition of neighbourhoods: Analysis of household's response and asymmetric preference structures. *Papers in Regional Science*, 96(4), 759-784

Idris, A. O., Habib, K. M. N. and Shalaby, A. (2015). An investigation on the performances of mode shift models in transit ridership forecasting. *Transportation Research Part A: Policy and Practice*, 78, 551-565.

Jenkins, S. P. (2005). Survival Analysis, Institute for Social and Economic Research, University of Essex, Colchester, UK.

Jiao, P. and Harata, N. (2007). Residential location choice behavior for different households: methodology and case study. *Proceedings of the Eastern Asia*

Society for Transportation Studies The 7th International Conference of Eastern Asia Society for Transportation Studies, 2007. Eastern Asia Society for Transportation Studies, 171-171.

Johnson, D. R. (2008). Using Weights in the Analysis of Survey Data. *Population Research Institute, The Pennsylvania State University*.

Jones, C., Coombes, M. and Wong, C. (2010). Geography of housing market areas: Final report. *London, CLG*.

Judson, R. A. and Owen, A. L. (1999). Estimating dynamic panel data models: a guide for macroeconomists. *Economics letters*, 65(1), 9-15.

Keirstead, J., Samsatli, N. and Shah, N. (2010). SynCity: an integrated tool kit for urban energy systems modelling. *Energy efficient cities: Assessment tools and benchmarking practices*, 21-42.

Kim, J. H., Pagliara, F. and Preston, J. (2005). The intention to move and residential location choice behaviour. *Urban studies*, 42(9),1621-1636.

Klinger, T. and Lanzendorf, M. (2016). Moving between mobility cultures: what affects the travel behavior of new residents? *Transportation*, 43, 243-271.

Kockelman, K. (1997). Travel behavior as function of accessibility, land use mixing, and land use balance: evidence from San Francisco Bay Area. *Transportation Research Record: Journal of the Transportation Research Board*, 116-125.

Kortum, K., Paleti ,R., Bhat, C. R., et al. (2012). Joint model of residential relocation choice and underlying causal factors. *Transportation Research Record*, 2303, 28-37.

Krizek, K. J. (2000). Pretest-posttest strategy for researching neighborhood-scale urban form and travel behavior. *Transportation Research Record: Journal of the Transportation Research Board*, 48-55.

Krizek, K. J. (2003). Residential relocation and changes in urban travel: Does neighborhood-scale urban form matter? *Journal of the American Planning Association*, 69, 265-281.

Lee, B. H. and Waddell, P. (2010). Residential mobility and location choice: a nested logit model with sampling of alternatives. *Transportation*, 37, 587-601.

Lee, B. H., Waddell, P., Wang, L., et al. (2010). Reexamining the influence of work and nonwork accessibility on residential location choices with a microanalytic framework. *Environment and Planning A*, 42, 913-930.

Lin, T., Wang, D. and Zhou, M. (2018). Residential relocation and changes in travel behavior: what is the role of social context change? *Transportation Research Part A: Policy and Practice*, 111, 360-374.

Litman, T. (2016). Evaluating transportation land use impacts: considering the impacts, benefits and costs of different land use development patterns.

Malaitham, S., Nakagawa, D., Matsunaka, R., et al. (2013). An Analysis of Residential Location Choice Behavior in Bangkok Metropolitan Region: An Application of Discrete Choice Models for the Ranking of Alternatives. *Journal of the Eastern Asia Society for Transportation Studies*, 10, 989-1008.

Manski, C. F. (1977). The structure of random utility models. *Theory and decision*, 8, 229-254.

Martinez, F., Aguila, F. and Hurtubia, R. (2009). The constrained multinomial logit: A semi-compensatory choice model. *Transportation Research Part B: Methodological*, 43, 365-377.

Martinez, F. and Donoso, P. (2004). MUSSA: a behavioural land use equilibrium model with location externalities, planning regulations and pricing policies. *U Chile-SECTRA*.

Martinez, F. and Hurtubia, R. (2006). Dynamic model for the simulation of equilibrium status in the land use market. *Networks and spatial economics*, 6, 55-73.

Masnavi, M. R. (2000). The new millennium and the new urban paradigm: the compact city in practice. *Achieving sustainable urban form*, 64-73.

McCulloch, A. (2010). Disparities in residential mobility among women with children. Paper.

McFadden, D. (1978) Modeling the choice of residential location. *Transportation Research Record*, 673, 72-77.

McHugh, K. E., Gober , P. and Reid, N. (1990). Determinants of short-and long-term mobility expectations for home owners and renters. *Demography,* **27**(1), 81-95.

Milakis, D., Efthymiou, D. and Antoniou, C. (2015). Quasi-longitudinal analysis of links between built environment, travel attitudes and travel behavior: A case of Greeks relocating from US to Greece. In *Proceedings of the 94th Annual Meeting of the Transportation Research Board*, 1-21.

Miller, E. J., Kriger, D. S. and Hunt, J. D. (1998). Integrated urban models for simulation of transit and land-use policies.

Moeckel, R. (2017). Constraints in household relocation: Modeling land-use/transport interactions that respect time and monetary budgets. *Journal of Transport and Land Use,* 10, 211-228.

Molin, E., Oppewal, H. and Timmermans, H. (1996). Predicting consumer response to new housing: A stated choice experiment. *Netherlands journal of housing and the built environment,* 11, 297-311.

Morimoto, A. (2015). Transportation and land use. *K. Doi, Traffic and Safety Sciences: Interdisciplinary Wisdom of IATSS* 22-30.

Mulder, C. H. (1996). Housing choice: Assumptions and approaches. *Netherlands journal of housing and the built environment,* 11, 209-232.

Naess, P. (2009). Residential self-selection and appropriate control variables in land use: Travel studies. *Transport Reviews,* 29, 293-324.

Nelson, B. (1997). Suburbs under Siege: Race, Space, and Audacious Judges. *The Journal of American History,* 84, 316.

Noland, R. B. (2001), Relationships between highway capacity and induced vehicle travel. *Transportation Research Part A: Policy and Practice,* 35, 47-72.

Oakil, A. (2013). Temporal dependence in life trajectories and mobility decisions. Utrecht University.

Oakil, A., Ettema, D., Arentze, T., et al. (2011). A longitudinal analysis of the dependence of the commute mode switching decision on mobility decisions and life cycle events. *Transportdynamics: Proceedings of 16th International Conference of Hong Kong Society for Transportation Studies (HKSTS). Hong Kong Society for Transportation Studies, Hong Kong.*

Oakil, A., Ettema, D., Arentze, T., et al. (2014). Changing household car ownership level and life cycle events: an action in anticipation or an action on occurrence. *Transportation,* 41, 889-904.

Opoku, R. A. and Abdul-Muhmin, A. G. (2010). Housing preferences and attribute importance among low-income consumers in Saudi Arabia. *Habitat international*, 34(2), 219-227.

Paccoud, A. and Mace, A. (2018). Tenure change in London’s suburbs: Spreading gentrification or suburban upscaling? *Urban studies*, 55(6), 1313–1328.

Paleti, R. (2015). Implicit choice set generation in discrete choice models: Application to household auto ownership decisions. *Transportation Research Part B: Methodological,* 80, 132-149.

Pawson, H. and Bramley, G. (2000). Understanding recent trends in residential mobility in council housing in England. *Urban studies*, 37, 1231-1259.

Pendyala, R., Konduri, K., Chiu, Y. C., et al. (2012). Integrated land use-transport model system with dynamic time-dependent activity-travel microsimulation. *Transportation Research Record: Journal of the Transportation Research Board*, 19-27.

Petrin, A., Train, K. (2010). A control function approach to endogeneity in consumer choice models. *Journal of Marketing Research*, 47 (1), 3–13.

Pickvance, C. G. (1974). Life cycle, housing tenure and residential mobility: a path analytic approach. *Urban Studies*, 11(2), 171-188.

Pinjari, A. R., Pendyala, R. M., Bhat, C. R., et al. (2011). Modeling the choice continuum: an integrated model of residential location, auto ownership, bicycle ownership, and commute tour mode choice decisions. *Transportation*, 38, 933.

Pisman, A., Allaert, G. and Lombaerde, P. (2011). Urban and suburban lifestyles and residential preferences in a highly urbanized society. Experiences from a case study in Ghent (Flanders, Belgium). *Belgeo*, 89-104.

Prillwitz, J., Harms, S. and Lanzendorf, M. (2006). Impact of life-course events on car ownership. *Transportation Research Record*, 1985, 71-77.

Prillwitz, J., Harms, S. and Lanzendorf, M. (2007). Interactions between residential relocations, life course events, and daily commute distances. *Transportation Research Record: Journal of the Transportation Research Board*, 64-69.

Rabe, B. and Taylor, M. (2010). Residential mobility, quality of neighbourhood and life course events. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 173(3), 531-555.

Rajamani, J., Bhat, C., Handy, S., et al. (2003). Assessing impact of urban form measures on nonwork trip mode choice after controlling for demographic and level-of-service effects. *Transportation Research Record: Journal of the Transportation Research Board*, 158-165.

Randall, C. (2011). Housing-Social Trends 41. *Office for National Statistics, UK.*

Rashidi, T. H., Auld, J. and Mohammadian, A. K. (2012). A behavioral housing search model: Two-stage hazard-based and multinomial logit approach to choice-set formation and location selection. *Transportation Research Part A: Policy and Practice*, 46, 1097-1107.

Rashidi, T. H., Mohammadian, A. and Koppelman, F. S. (2011). Modeling interdependencies between vehicle transaction, residential relocation and job change. *Transportation,* 38, 909.

Rhoads, T. A. and Shogren, J. F. (2006) .Why do cities use supply side strategies to mitigate traffic congestion externalities? *Economics Letters,* 92, 214-219.

Rieser-Schussler, N., Axhausen, K. (2012). Investigating the influence of environmentalism and variety seeking on mode choice. *Transportation Research. Record Journal of Transportation Research Board.* 2322, 31–41.

Ritchey, P. N. (1976). Explanations of migration. *Annual review of sociology,* 2, 363-404.

Rivera, M. A. I. and Tiglao, N. C. C. (2005). Modeling residential location choice, workplace location choice and mode choice of two-worker households in Metro Manila. *Proceedings of the Eastern Asia Society for Transportation Studies,* 1167-1178.

Rivers, D. and Vuong, Q. (1988). Limited information estimators and exogeneity tests for simultaneous probit models. Journal of Econometrics, 39, 347–366.

Rumsey, D J. 2016, Statistics for Dummies, 2nd Edition, ISBN: 978-1-119-29352-1

Salvini, P. and Miller, E. J. (2005). ILUTE: An operational prototype of a comprehensive microsimulation model of urban systems. *Networks and spatial economics,* 5, 217-234.

Sanchez, A. C. and Andrews, D. (2011). Residential Mobility and Public Policy in OECD Countries. *OECD Journal: Economic Studies*.

Satterthwaite, D. (2005). The scale of urban change worldwide 1950-2000 and its underpinnings: IIED.

Scheiner, J. (2006). Housing mobility and travel behaviour: A process-oriented approach to spatial mobility: Evidence from a new research field in Germany. *Journal of Transport Geography,* 14, 287-298.

Schirmer, P., Van Eggermond, M. and Axhausen, K. (2013). Measuring location in residential location choice-an empirical study in the C anton of Zurich. *13th International Conference on Computers in Urban Planning and Urban Management, Utrecht.*

Sener, I. N., Pendyala, R.M. and Bhat, C. R. (2011). Accommodating spatial correlation across choice alternatives in discrete choice models: an application to modeling residential location choice behavior. *Journal of Transport Geography,* 19, 294-303.

Song, F., Hess, S. and Dekker, T. (2018). Accounting for the impact of variety-seeking: Theory and application to HSR-air intermodality in China. *Journal of Air Transport Management*, *69,* 99-111.

Speare, A. (1970). Home ownership, life cycle stage, and residential mobility. *Demography*, *7*(4), 449-458.

Stillwell, J. and Thomas, M. (2016). How far do internal migrants really move? Demonstrating a new method for the estimation of intra-zonal distance. *Regional Studies, Regional Science,* 3, 28-47.

Swait, J. (2001). A non-compensatory choice model incorporating attribute cutoffs. *Transportation Research Part B: Methodological*, 35, 903-928.

Swait, J. and Ben-Akiva, M. (1987). Incorporating random constraints in discrete models of choice set generation. *Transportation Research Part B: Methodological,* 21, 91-102.

Tatsiramos, K. (2009). Geographic labour mobility and unemployment insurance in Europe. *Journal of population Economics*, 22, 267-283.

Tayyaran, M. R., Khan, A. M. and Anderson, D. A. (2003). Impact of telecommuting and intelligent transportation systems on residential location choice. *Transportation Planning and Technology*, 26, 171-193.

Train, K. E. (2009). Discrete choice methods with simulation: Cambridge university press.

Travers, T., Tunstall, R., Whitehead, C. M., et al. (2007). Population mobility and service provision: a report for London Councils.

Uhrig, S. (2008). The nature and causes of attrition in the British Household Panel Study. ISER Working Paper Series.

Van Acker, V. and Witlox, F. (2010). Car ownership as a mediating variable in car travel behaviour research using a structural equation modelling approach to identify its dual relationship. *Journal of Transport Geography,* 18, 65-74.

Van der Vlist, A. J., Gorter, C., Nijkamp, P., et al. (2002). Residential mobility and local housing-market differences. *Environment and Planning A,* 34, 1147-1164.

Van Ham, M. and Clark, W. A. (2009). Neighbourhood mobility in context: household moves and changing neighbourhoods in the Netherlands. *Environment and planning A*, 41(6), 1442-1459.

Van Ommeren, J. N., Rietveld, P. and Nijkamp, P. (1998). Spatial moving behavior of two-earner households. *Journal of regional science,* 38, 23-41.

Vandecasteele, L. and Debels, A. (2006). Attrition in panel data: the effectiveness of weighting. *European Sociological Review,* 23, 81-97.

Vega, A. and Reynolds-Feighan, A. (2009). A methodological framework for the study of residential location and travel-to-work mode choice under central and suburban employment destination patterns. *Transportation Research Part A: Policy and Practice*, 43(4), 401-419.

Waddell, P. (2002). UrbanSim: Modeling urban development for land use, transportation, and environmental planning. *Journal of the American Planning Association,* 68, 297-314.

Waddell, P. (2006). Reconciling household residential location choices and neighborhood dynamics. *Under revision, Sociological Methods and Research*.

Wagner, P. and Wegener, M. (2007). Urban land use, transport and environment models: Experiences with an integrated microscopic approach. *DisP-the planning review*, 43, 45-56.

Walker, J. L. (2001). Extended discrete choice models: integrated framework, flexible error structures, and latent variables. Massachusetts Institute of Technology.

Walker, J. L. and Ben-Akiva, M. (2002). Generalized random utility model. Math. Soc. Sci. 43 (3), 303–343.

Walker, J. L and Li J. (2007). Latent lifestyle preferences and household location decisions. *Journal of Geographical Systems,* 9, 77-101.

Washington, S. P., Karlaftis, M. G. and Mannering, F. (2010). *Statistical and econometric methods for transportation data analysis: Chapman and Hall/CRC.*

Watts, M. J. (2009). The impact of spatial imbalance and socioeconomic characteristics on average distance commuted in the Sydney metropolitan area. *Urban studies,* 46, 317-339.

Wegener, M. (2011). From macro to micro—how much micro is too much? *Transport Reviews,* 31, 161-177.

Weisbrod, G., Lerman, S. R. and Ben-Akiva, M. (1980). Tradeoffs in residential location decisions: Transportation versus other factors. *Transport Policy and Decision Making,* 1, 13-26.

White, M. J. (1977). A model of residential location choice and commuting by men and women workers. *Journal of regional science,* 17, 41-52.

Wooldridge, J. M. (2010). Econometric analysis of cross section and panel data: MIT press.

Yamamoto, T. (2008). The impact of life-course events on vehicle ownership dynamics: The cases of france and Japan. *IATSS research,* 32, 34-43.

Yates, J. and Mackay, D. F. (2006). Discrete choice modelling of urban housing markets: A critical review and an application. *Urban studies,* 43, 559-581.

Zhang, J., Yu, B. and Chikaraishi, M. (2014). Interdependences between household residential and car ownership behavior: a life history analysis. *Journal of Transport Geography,* 34, 165-174.

Zhou, B. and Kockelman, K. (2008). Microsimulation of residential land development and household location choices: bidding for land in Austin, Texas. *Transportation Research Record: Journal of the Transportation Research Board*, 106-112.

Zhuge, C., Shao, C., Gao, J., et al. (2016). Agent-based joint model of residential location choice and real estate price for land use and transport model. *Computers, Environment and Urban Systems*, 57, 93-105.

Zolfaghari, A., Sivakumar, A. and Polak, J. W. (2012). Choice set pruning in residential location choice modelling: a comparison of sampling and choice set generation approaches in greater London. *Transportation Planning and Technology*, 35(1), 87-106.

Zolfaghari, A. (2013). Methodological and empirical challenges in modelling residential location choices. Doctoral dissertation, Imperial College, London.

Zondag, B. and Pieters, M. (2005). Influence of accessibility on residential location choice. *Transportation Research Record: Journal of the Transportation Research Board*, 63-70.

Appendices

Appendix A Example code of BHPS data processing

```
library(data.table)
install.packages("haven")
install.packages("tidyr")
install.packages("gtools")
install.packages("dplyr")

HHHRESP=read_sav("Z:/PhD/Model Estimation/w3_residential _mobility/BHPSD/ahhresp.sav")
HINDRESP=read_sav("Z:/PhD/Model Estimation/w3_residential _mobility/BHPSD/aindresp.sav")
HINDALL=read_sav("Z:/PhD/Model Estimation/w3_residential _mobility/BHPSD/aindall.sav")
HHHRESP=as.data.table(HHHRESP)
HINDRESP=as.data.table(HINDRESP)
HINDALL=as.data.table(HINDALL)

# To ensure column PID does not change
colnames(HINDRESP)[which(names(HINDRESP)=="PID")]="ZZZ"
colnames(HINDALL)[which(names(HINDALL)=="PID")]="ZZZ"

#To get the unique column name for all waves.
names(HHHRESP)=sub("A","",names(HHHRESP))
names(HINDRESP)=sub("A","",names(HINDRESP))
names(HINDALL)=sub("A","",names(HINDALL))

# To get the actual name of PID back
colnames(HINDRESP)[which(names(HINDRESP)=="ZZZ")]="PID"
colnames(HINDALL)[which(names(HINDALL)=="ZZZ")]="PID"

# Select the data required. Can add any new column inside list
data_HHHRESP=HHHRESP[,list(HID,AGECHY,NCH02,NCH1215,NCH1618,NCH34,NCH511,N
KIDS,NEMP,HHMOVE,HSOWND,RENTF,HSCOST,HSJB,HSVAL,RENT,RENTG,HSROOM,H
STYPE,FIHHYR,NA75PL,NCARS)]

data_HINDRESP=HINDRESP[,list(HID,PNO,PID,QFACHI,QFEDHI,JBSAT,JUSPEC,JBSOC,SPS
OC,JBRGSC,J2HAS,JBFT,JBSTAT,CJSTEN,JBTTWM,JBTTWT,FIYRL,JBPL,RACE,FISIT,REGI
ON,MOVJB,MOVJBA,MOVJBB,MOVJBC,MOVJBD,MOVJBE,MOVJBF,MOVJBG,MOVJBH,M
OVJBI,MOVY1,MOVY2,PLNOWY4,F139,TENURE,FIYR,FIYRB,LKMOVE,LKMOVY,LKNBR
D,HHSIZE,HHTYPE)]

data_HINDALL=HINDALL[,list(HID,PNO,PID,DEPCHL,DISTMOV,MOVEST,MASTAT,HOH)]

# Merge individual data sets (indresp and indall)
data_HIND_HHH=merge(data_HINDRESP,data_HHHRESP, by="HID", all.x=TRUE)

# Add household level data into the individual data file
data_HIND_HHH_HALL=merge(data_HIND_HHH,data_HINDALL,by="PNO",all.x=TRUE)

#Select the data for head of the household
data_HIND_HHH_HALL_HHOH=data_HIND_HHH_HALL[HOH==1,]

# Steps to get the data where HOH is not interviewed
```

```
data_HIND_HALL=merge(data_HINDRESP,data_HINDALL, by=c("PID","HID","PNO"),
all.x=TRUE)
data_HIND_HALL_HHOH=data_HIND_HALL[HOH==1,]
data_HIND_HALL_HHH=merge(data_HHHRESP,data_HIND_HALL_HHOH, by="HID",
all.x=TRUE)
data_HIND_HALL_HHH1=replace_na(data_HIND_HALL_HHH,list(PID=9999,PNO=9999))
data_NHHOH=data_HIND_HALL_HHH1[PID==9999]
data_NHHOH_HHID=data_NHHOH[,"HID"]
data_NHHOH_FULL=merge(data_NHHOH_HHID,data_HHHRESP, by="HID", all.x=TRUE)
data_NHHOH_FULL1=merge(data_NHHOH_FULL,data_HIND_HALL, by="HID", all.x=TRUE)
data_NHHOH_FULL2=data_NHHOH_FULL1[!duplicated(data_NHHOH_FULL1$HID), ]

# Bind the data of HOH interviewed and non interviewed
data_full=rbind(data_HIND_HHH_HALL_HHOH,data_NHHOH_FULL2)

#Omit any missing observations
data_full_w1=na.omit(data_full)
data_full_w1[,WAVE:=1]
data_full_w1=as.data.table(data_full_w18)
write.csv(data_full_w1,file="data_full_w11.csv")
#Repeat the same steps for all eighteen waves

#####################################################################
# Bind data for all 18 waves
data_full_w1_w18=rbind(data_full_w1,data_full_w2,data_full_w3,data_full_w4,data_full_w5,data_f
ull_w6,data_full_w7,data_full_w8,data_full_w9,data_full_w10,data_full_w11,data_full_w12,data_fu
ll_w13,data_full_w14,data_full_w15,data_full_w16,data_full_w17,data_full_w18)

data_full_w1_w18=as.data.table(data_full_w1_w18)
write.csv(data_full_w1_w18,file="data_full_W1_w18.csv")
```

Appendix B Example code of sampling weight calculation using raking technique

```
library(data.table)
library(survey)
datafilename="data_BHPS_bal.csv"
data_BHPS=read.csv(datafilename,header=T)
data_BHPS=data.table(data_BHPS)
N=nrow(data_BHPS)
data_BHPS_unweighted <- svydesign(ids=~1, data=data_BHPS)
HHTYPE_dist <- data.frame(HHTYPE_GR = c(1, 2, 3, 4),
                Freq = N * c(0.267, 0.278, 0.335, 0.120))
HHINCOME_dist <- data.frame(FIHHYR_GR = c(1, 2, 3),
                Freq = N * c(0.697, 0.256, 0.047))
EDU_dist <- data.frame(QFACHI_GR = c(1, 2, 3, 4),
                Freq = N * c(0.019, 0.128, 0.342, 0.511))
NEMP_dist <- data.frame(NEMP_GR = c(0, 1, 2),
                Freq = N * c(0.345, 0.288, 0.367))
TENURE_dist <- data.frame(TENURE_GR = c(1, 2, 3),
                Freq = N * c(0.666, 0.206, 0.128))
ELDERLY_dist <- data.frame(NA75PL_GR = c(0, 1),
                  Freq = N * c(0.879, 0.121))
JOBLENGTH_dist <- data.frame(CJSTEN_GR = c(1, 2, 3),
                   Freq = N * c(0.497, 0.197, 0.306))
NEWCHILD_dist <- data.frame(NCH_GR = c(0, 1),
                    Freq = N * c(0.929, 0.071))
data_BHPS_rake <- rake(design = data_BHPS_unweighted,
                sample.margins = list(~HHTYPE_GR, ~FIHHYR_GR, ~QFACHI_GR, ~NEMP_GR,
                ~TENURE_GR, ~NA75PL_GR, ~CJSTEN_GR, ~NCH_GR),
                population.margins = list(HHTYPE_dist, HHINCOME_dist, EDU_dist, NEMP_dist,
                TENURE_dist, ELDERLY_dist, JOBLENGTH_dist, NEWCHILD_dist))
wt=as.data.table(weights(data_BHPS_rake))
names(wt)[1]<-paste("weight_rake_final")
data_wt <- cbind(data_BHPS,wt)
summary(weights(data_BHPS_rake))
write.csv(data_wt,file="data_BHPS_weight.csv")
```

Appendix C Testing correlations across the random terms in joint model of residential mobility and its geographical scale

Several nested specifications are tested in the residential mobility model to capture the correlations across the random parameters of the alternatives such as correlation between moved at the local level and moved at the regional level, correlation between moved at the regional level and moved at the national level, correlation between moved at the local, regional and national level, etc. Although goodness fit of all the models that captured different nesting structures are found better than the MNL estimation (Table B1 to Table B4), a poor fit is observed compared to the heteroscedastic model presented in chapter 3 (Table 3-4). To capture the full scale of correlations among the random parameters, the lower trigonal matrix of Cholesky decomposition has been considered. The estimated model with Cholesky decomposition offers larger fit compare to the models that captured other forms of nesting structures, however, the improvement in the fit of Cholesky model compare to the heteroscedastic model is negligible. The sign and magnitude of the estimated parameters are found consistent in all the models.

Table C1 MNL estimation results for model of decision to move in different geographical scales

Parameters	Joint decision of residential mobility and its scale					
	Local level (LL)		Regional level (RL)		National level (NL)	
	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat
Alternative specific constants (not moved is the base alternative)	-5.3203	-26.6	-6.3613	-19	-6.431	-19.2
Household level characteristics						
Household type (base is couple with child)						
Single member household	0.8986	5.75	1.012	4.03	0.6016	2.17
Couple without child	0.3886	3.16	0.9432	4.72	0.5894	2.85
Lone parents	0.5936	4.35	0.124	0.4	-0.303	-0.85
Household income (base is less than £20,000)						
Between £20,00 to £40,000	0.2184	2	0.3055	1.72	0.4588	2.25
More than £40,000	0.1925	1.31	0.2497	1.03	0.6054	2.39
Education attainment of household head (base is below O level)						
O and A level degree	0.1733	1.85	0.7343	4.35	0.1709	0.89
Graduate degree	0.2861	2.29	0.7059	3.19	0.7481	3.48
Post-graduate degree	0.8303	4.13	1.1129	2.98	1.395	4.96
Number of employees in the household (base is no employee)						
One employee	0.0395	0.34	0.0715	0.38	-0.134	-0.6
More than one employees	0.2776	2.03	-0.0508	-0.2	-0.278	-1.07
Length of current job of household head	-0.0081	-1.48	-0.0389	-3.3	-0.018	-1.63
Presence of senior adult (>75 years)	-0.5784	-3.83	-0.7409	-2.5	0.2695	1.08
Dwelling level characteristics						
Tenure type (base is owned house)						
Rented social housing	0.5916	5.32	0.095	0.4	-0.63	-2.17
Rented private housing	1.737	16.49	2.0447	12.6	1.3162	6.63
Crowd (household size\number of rooms)	1.3153	9.49	1.1495	5	1.0501	3.7
Life course events						
Having child in last one year	0.4702	2.91	0.8162	2.94	0.4553	1.43
Changed job in last one year	0.1592	1.43	0.1241	0.68	0.2458	1.22
Location characteristics						
Metropolitan area (base is other than London)						
London	0.0424	0.35	-2.661	-3.1	1.4558	9.33
Measures of model fit						
Number of observations	24718					
Initial LL	-34266.40					
Final LL	-5263.96					

Table C2 MMNL estimation results for model of decision to move in different geographical scales which captures the correlation between local and regional level moves

Parameters	Joint decision of residential mobility and its scale					
	Local level (LL)		Regional level (RL)		National level (NL)	
	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat
Alternative specific constants (not moved is the base alternative)	-5.5228	-28.7	-6.4672	-24.5	-6.5684	-20.9
Household level characteristics						
Household type (base is couple with child)						
Single member household	0.8326	5.4	0.8825	3.8	0.5844	2.6
Couple without child	0.3345	2.7	0.8629	4.7	0.5700	3.2
Lone parents	0.4870	3.3	-0.0151	-0.1	-0.3216	-1.2
Household income (base is less than £20,000)						
Between £20,00 to £40,000	0.1783	1.6	0.2221	1.2	0.4634	2.3
More than £40,000	0.1121	0.7	0.1117	0.4	0.6231	2.5
Education attainment of household head (base is below O level)						
O and A level degree	0.1878	1.8	0.7499	4.3	0.1771	0.9
Graduate degree	0.2528	1.8	0.6760	3.0	0.7695	3.5
Post-graduate degree	0.8970	3.8	1.1573	3.1	1.3879	4.6
Number of employees in the household (base is no employee)						
One employee	0.0556	0.5	0.0144	0.1	-0.1435	-0.6
More than one employees	0.2957	2.1	-0.1233	-0.6	-0.3101	-1.2
Length of current job of household head	-0.0066	-1.1	-0.0354	-3.0	-0.0176	-1.6
Presence of senior adult (>75 years)	-0.5761	-3.7	-0.7445	-2.4	0.2765	1.1
Dwelling level characteristics						
Tenure type (base is owned house)						
Rented social housing	0.6231	5.0	0.1311	0.5	-0.6402	-2.2
Rented private housing	1.9255	15.6	2.2481	12.7	1.3521	6.8
Crowd (household size\number of rooms)	1.3544	9.0	1.1242	4.6	1.0200	3.6
Life course events						
Having child in last one year	0.3555	2.2	0.6730	2.5	0.4730	1.9
Changed job in last one year	0.1174	1.0	0.0767	0.4	0.2382	1.2
Location characteristics						
Metropolitan area (base is other than London)						
London	0.0395	0.3	-2.6942	-3.8	1.5465	8.4
Correlated random parameters						
$\sigma_{LL\text{-}RL}$	0.7170	10.9				
σ_{NL}	0.5679	2.0				
Measures of model fit						
Number of observations	24718					
Initial LL	-34266.40					
Final LL	-5231.68					

Table C3 Estimation results for model of decision to move in different geographical scales which captures the correlation between regional and national level moves.

Parameters	Joint decision of residential mobility and its scale					
	Local level (LL)		Regional level (RL)		National level (NL)	
	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat
Alternative specific constants (not moved is the base alternative)	-5.5997	-25.6	-6.6817	-19.3	-6.7901	-12.6
Household level characteristics						
Household type (base is couple with child)						
Single member household	0.8427	4.9	0.9987	3.7	0.5681	1.6
Couple without child	0.3766	2.9	0.8773	4.2	0.5418	2.0
Lone parents	0.5053	3.3	0.0730	0.2	-0.3798	-1.0
Household income (base is less than £20,000)						
Between £20,00 to £40,000	0.1863	1.6	0.2822	1.5	0.4546	2.2
More than £40,000	0.1152	0.7	0.1972	0.8	0.6132	2.3
Education attainment of household head (base is below O level)						
O and A level degree	0.1931	1.7	0.7854	4.1	0.2186	1.0
Graduate degree	0.2308	1.6	0.6316	2.5	0.8005	3.2
Post-graduate degree	0.9299	3.8	0.9026	2.2	1.3775	3.8
Number of employees in the household (base is no employee)						
One employee	0.0455	0.4	0.0597	0.3	-0.1364	-0.5
More than one employees	0.3059	2.1	-0.1369	-0.6	-0.3288	-1.1
Length of current job of household head	-0.0065	-1.1	-0.0369	-3.1	-0.0169	-1.5
Presence of senior adult (>75 years)	-0.5909	-3.7	-0.6877	-2.2	0.3139	1.1
Dwelling level characteristics						
Tenure type (base is owned house)						
Rented social housing	0.6404	5.0	0.0703	0.3	-0.6699	-2.1
Rented private housing	1.9189	15.2	2.1625	12.0	1.4722	6.8
Crowd (household size\number of rooms)	1.3725	8.6	1.1589	4.8	1.0578	3.1
Life course events						
Having child in last one year	0.3813	2.2	0.6978	2.5	0.4174	1.2
Changed job in last one year	0.1268	1.1	0.0895	0.5	0.2049	1.0
Location characteristics						
Metropolitan area (base is other than London)						
London	0.0428	0.3	-2.5041	-3.7	1.6325	9.2
Correlated random parameters						
σ_{LL}	0.7698	10.1				
$\sigma_{RL\text{-}NL}$	0.8950	8.0				
Measures of model fit						
Number of observations	24718					
Initial LL	-34266.40					
Final LL	-5223.65					

Table C4 Estimation results for model of decision to move in different geographical scales which captures the correlation between local, regional and national level moves.

Parameters	Joint decision of residential mobility and its scale					
	Local level (LL)		Regional level (RL)		National level (NL)	
	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat
Alternative specific constants (not moved is the base alternative)	-5.5021	-35.5	-6.4447	-19.4	-6.6344	-9.6
Household level characteristics						
Household type (base is couple with child)						
Single member household	0.7965	5.4	0.8421	3.3	0.5140	2.1
Couple without child	0.3154	2.7	0.8339	4.2	0.5514	3.4
Lone parents	0.4534	3.0	-0.0407	-0.1	-0.4407	-1.3
Household income (base is less than £20,000)						
Between £20,00 to £40,000	0.1781	1.7	0.2232	1.3	0.4596	2.2
More than £40,000	0.0874	0.6	0.0792	0.3	0.5620	2.2
Education attainment of household head (base is below O level)						
O and A level degree	0.1742	1.6	0.7473	4.2	0.1788	1.4
Graduate degree	0.2011	1.4	0.6115	2.6	0.7051	4.0
Post-graduate degree	0.8868	3.7	1.1109	2.8	1.4931	8.4
Number of employees in the household (base is no employee)						
One employee	0.0560	0.5	0.0131	0.1	-0.1753	-0.8
More than one employees	0.2869	2.0	-0.1404	-0.6	-0.3178	-1.3
Length of current job of household head	-0.0068	-1.2	-0.0351	-3.0	-0.0151	-1.4
Presence of senior adult (>75 years)	-0.5745	-3.5	-0.7230	-2.4	0.2965	1.5
Dwelling level characteristics						
Tenure type (base is owned house)						
Rented social housing	0.6181	5.3	0.1270	0.5	-0.5975	-3.5
Rented private housing	1.9376	17.0	2.2473	14.6	1.5180	7.4
Crowd (household size\number of rooms)	1.3401	9.1	1.1123	4.5	1.0462	3.6
Life course events						
Having child in last one year	0.3480	2.1	0.6539	2.7	0.3189	1.2
Changed job in last one year	0.1067	0.9	0.0706	0.4	0.2082	1.0
Location characteristics						
Metropolitan area (base is other than London)						
London	0.1202	0.9	-2.5895	-3.8	1.5373	9.2
Correlated random parameters						
$\sigma_{LL\text{-}RL\text{-}NL}$	0.7643	13.5				
Measures of model fit						
Number of observations	24718					
Initial LL	-34266.40					
Final LL	-5213.11					

Table C5 Estimation results for model of decision to move in different geographical scales which captures the full range of correlation by means of Cholesky decomposition

Parameters	Joint decision of residential mobility and its scale					
	Local level (LL)		Regional level (RL)		National level (NL)	
	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat
Alternative specific constants (not moved is the base alternative)	-5.6378	-25.6	-6.4286	-16	-6.643	-15.6
Household level characteristics						
Household type (base is couple with child)						
Single member household	0.8537	4.91	0.7573	2.54	0.4215	1.32
Couple without child	0.3597	2.63	0.6663	2.88	0.4589	2
Lone parents	0.4854	3.02	-0.0343	-0.1	-0.448	-1.19
Household income (base is less than £20,000)						
Between £20,00 to £40,000	0.1781	1.55	0.1904	0.95	0.4379	1.98
More than £40,000	0.0908	0.59	0.0836	0.29	0.5479	1.97
Education attainment of household head (base is below O level)						
O and A level degree	0.19	1.69	0.7238	3.97	0.1493	0.67
Graduate degree	0.1671	1.07	0.6671	2.79	0.7787	3.09
Post-graduate degree	0.9136	3.7	1.01	2.55	1.464	4.37
Number of employees in the household (base is no employee)						
One employee	0.068	0.54	0.0325	0.15	-0.157	-0.59
More than one employees	0.3305	2.24	-0.2625	-1	-0.369	-1.25
Length of current job of household head	-0.0075	-1.26	-0.0331	-2.8	-0.015	-1.37
Presence of senior adult (>75 years)	-0.5929	-3.58	-0.7221	-2.2	0.2919	1.04
Dwelling level characteristics						
Tenure type (base is owned house)						
Rented social housing	0.6422	4.97	0.0869	0.3	-0.64	-1.97
Rented private housing	1.936	15.23	2.1422	11.4	1.4906	6.87
Crowd (household size\number of rooms)	1.3735	8.48	1.013	3.98	0.984	3.01
Life course events						
Having child in last one year	0.3664	2.16	0.63	2.14	0.3463	1.05
Changed job in last one year	0.105	0.9	0.0465	0.24	0.1794	0.88
Location characteristics						
Metropolitan area (base is other than London)						
London	0.0977	0.68	-2.4786	-2.9	1.6295	8.96
Cholesky parameters						
$\sigma_{LL\text{-}LL}$	0.8292	12.17				
$\sigma_{RL\text{-}LL}$	0.3721	2.57				
$\sigma_{NL\text{-}LL}$	0.5315	3.26				
$\sigma_{RL\text{-}RL}$	0.8734	5.86				
$\sigma_{NL\text{-}RL}$	0.7009	4.14				
$\sigma_{NL\text{-}NL}$	Fixed	-				
Measures of model fit						
Number of observations	24718					
Initial LL	-34266.40					
Final LL	-5207.96					

Appendix D Differences in the residential mobility behaviour in the balanced and unbalanced panel

It is mentioned in chapter 3 that the weighting of the balanced panel has made an adjustment of the sample for representativeness and behaviour at a significant level. Moreover, models are estimated for the unbalanced panel to see the differences in the parameters of the models for unbalanced and the weighted balanced panel. It is observed that the direction of sensitivity (sign) of all the significant parameters in the unbalanced panel model remains the same as the corresponding parameters estimated for the balanced panel (Table C-1). The magnitude of most of the estimated parameters is also found very similar in both cases although small differences are observed for few parameters. Since the balanced and unbalanced panels are not independent, t-stat cannot be used to check whether the differences are statistically significant or not. The small differences of few parameters in the model for balanced and unbalanced panel are most likely to be driven by the many moves of the household who dropped out very early in the unbalanced panel resulting over-representation of the mobility behaviour. Since the behaviour of the households who dropped out is not observed, there is no straightforward way to handle this issue.

Table D1 Estimation results of the models for weighted balanced and unbalanced panel

Parameters	Decision to move			
	Balanced panel		Unbalanced panel	
	Coeff.	t-stat	Coeff	t-stat
Not moved is the base alternative				
Moved at local level				
Alternative specific constants				
Mean	-5.6400	-25.5	-4.9642	-28.9
Standard deviation	-0.7894	-12.1	-0.7124	-10.89
Household level characteristics				
Household type (base is couple with child)				
Single member household	0.8536	5.0	0.7106	5.24
Couple without child	0.3953	3.0	0.2972	1.85
Lone parents	0.4832	3.1	0.3944	4.1
Household income (base is less than £20,000)				
Between £20,00 to £40,000	0.1778	1.6	0.1581	1.86
More than £40,000	0.0572	0.4	0.0807	0.67
Education attainment of household head (base is below O level)				
O and A level degree	0.1791	1.6	0.1939	2.23
Graduate degree	0.2566	1.7	0.2714	2.21
Post-graduate degree	0.9846	4.0	0.744	1.47
Number of employees in the household (base is no employee)				
One employee	0.0450	0.4	-0.0102	-0.12
More than one employees	0.3254	2.3	0.1484	1.32
Length of current job of household head	-0.0071	-1.2	-0.014	-2.89
Presence of senior adult (>75 years)	-0.5982	-3.7	-0.4687	-2.92
Dwelling level characteristics				
Tenure type (base is owned house)				
Rented social housing	0.6570	5.1	0.6353	6.83
Rented private housing	1.9209	15.1	1.7144	15.66
Crowd (household size\number of rooms)	1.3908	8.6	1.2061	7.83
Life course events				
Having child in last one year	0.3506	2.1	0.4731	3.78
Changed job in last one year	0.1443	1.2	0.2892	3.37
Location characteristics				
Metropolitan area (base is other than London)				
London	0.0011	0.0	-0.2835	-2.42

Table D1 Estimation results weighted balanced and weighted unbalanced panel (cont.)

Parameters	Decision to move			
	Balanced panel		Unbalanced panel	
	Coeff.	t-stat	Coeff	t-stat
Moved at regional level				
Alternative specific constants				
Mean	-6.6415	-30.3	-6.4571	-18.02
Standard deviation	1.0096	8.4	1.0583	9.38
Household level characteristics				
Household type (base is couple with child)				
Single member household	0.8756	4.1	0.7711	3.46
Couple without child	0.7186	4.2	0.7069	2.53
Lone parents	0.0417	0.1	0.4475	2.19
Household income (base is less than £20,000)				
Between £20,00 to £40,000	0.2000	1.0	0.3355	2.31
More than £40,000	0.1570	0.6	0.3169	1.47
Education attainment of household head (base is below				
O and A level degree	0.7410	3.9	0.8119	5.41
Graduate degree	0.7530	3.1	0.7611	4.45
Post-graduate degree	1.1217	2.9	0.9665	2.32
Number of employees in the household (base is no				
One employee	0.0677	0.4	0.1013	0.63
More than one employees	-0.2119	-1.0	0.0564	0.28
Length of current job of household head	-0.0321	-2.7	-0.0357	-3.93
Presence of senior adult (>75 years)	-0.7175	-2.2	-0.706	-1.89
Dwelling level characteristics				
Tenure type (base is owned house)				
Rented social housing	0.0537	0.2	0.056	0.53
Rented private housing	2.1278	11.2	1.7127	9.95
Crowd (household size\number of rooms)	1.0380	4.1	0.9268	3.02
Life course events				
Having child in last one year	0.6260	2.5	0.5427	2.2
Changed job in last one year	0.0837	0.4	0.173	1.17
Location characteristics				
Metropolitan area (base is other than London)				
London	-2.6171	-4.2	-1.9051	-3.16

Table D1 Estimation results weighted balanced and weighted unbalanced panel (cont.)

Parameters	Decision to move			
	Balanced panel		Unbalanced panel	
	Coeff.	t-stat	Coeff	t-stat
Moved at national level				
Mean	-6.7952	-17.8	-6.3223	-22.78
Standard deviation	0.9223	5.7	0.998	6.39
Household level characteristics				
Household type (base is couple with child)				
Single member household	0.5353	1.8	0.5417	2.14
Couple without child	0.5448	2.5	0.4501	2.44
Lone parents	-0.3728	-1.1	0.23	0.95
Household income (base is less than £20,000)				
Between £20,00 to £40,000	0.4564	2.2	0.5007	3.07
More than £40,000	0.5921	2.3	0.6565	3.07
Education attainment of household head (base is below				
O and A level degree	0.1474	0.7	0.3055	1.97
Graduate degree	0.7576	3.3	0.7157	4.09
Post-graduate degree	1.4780	4.6	0.8822	3.09
Number of employees in the household (base is no				
One employee	-0.1324	-0.5	0.1369	0.79
More than one employees	-0.3334	-1.2	-0.3538	-1.74
Length of current job of household head	-0.0166	-1.5	-0.0263	-2.6
Presence of senior adult (>75 years)	0.2877	1.1	-0.078	-0.36
Dwelling level characteristics				
Tenure type (base is owned house)				
Rented social housing	-0.6642	-2.2	-0.4976	-2.38
Rented private housing	1.3763	6.4	1.2952	7.52
Crowd (household size\number of rooms)	1.0070	3.6	0.972	2.99
Life course events				
Having child in last one year	0.4402	1.3	0.5132	1.91
Changed job in last one year	0.2133	1.1	0.468	3.16
Location characteristics				
Metropolitan area (base is other than London)				
London	1.6700	8.7	0.9065	5.5
Measures of model fit				
Number of observations	24718		50282	
Initial LL	-34266.4		-69705.7	
Final LL	-5210.8		-12800.4	

Appendix E Discussion on state-dependence

Detailed data analysis is conducted to see the extent of state-dependence. Findings are also cross compared with those reported by other researchers who have used longitudinal data and potentially encountered similar issues. The findings are summarized below.

Firstly, in our data, the likelihood of changing behaviour is investigated in a row and observed that indeed a very few respondents have changed their residential location, car ownership and travel mode in two consecutive years (Table D1).

Table E1 Percentage of respondents who changed their behaviour in two subsequent years

Model components	Behaviour Year t	Year t+1	Respondents in %
Residential mobility	Moved	Moved at the local level	0.3%
		Moved at the regional level	0.1%
		Moved at the national level	0.1%
Car ownership change	Gained car	Gained car	0.3%
	Lost car	Lost car	0.1%
Travel mode switching	Changed travel mode	Changed to public transport	0.3%
		Changed to car	0.9%
		Changed to active travel	0.1%

Two potential approaches are evaluated to capture this effect is in the model:

1. Using the lagged dependent variable as an explanatory variable in the model
2. Using 'stay length' as an explanatory variable

Using lagged dependent variable refers to directly acknowledging that the impact of the decision at t affects the decision at t+1. A review of literature revealed that in case of modelling residential relocation, the lagged dependent variable has rarely been used in literature. To the best of my knowledge, only McHugh, Gober and Reid (1990) used lagged variable (recent movers as a dummy) in residential relocation choice modelling and found counter-intuitive result that recent movers are likely to move again. A series of lagged variables are considered in order to capture behaviour at time t-1, t-2, etc, but this has led to well-known issues related to multicollinearity, driven by the fact that rare events are modelled here.

The duration of stay has been used as an independent variable in several previous papers on residential location choice (e.g. Davies and Pickles 1985; McHugh, Gober and Reid 1990; Habib 2009; Clark and Lisowski 2017). Duration of stay (as in common practice in literature) is tested as an independent variable in the model of residential mobility decision. The parameter of stay duration gave a negative estimation which is consistent with the finding in the literature (Davies and Pickles 1985; McHugh, Gober and Reid 1990; Habib 2009; Clark and Lisowski 2017). The results are presented in Table D2.

However, the results indicate that the inclusion of the stay-length variable reduces the explanatory power of other important variables that represent the behaviour of the larger community. Further, though supported by literature, the negative sign indicating the longer one stays, less likely (s)he is to move is misleading. Because, on average, households in England change their home in every 8 years (Randall 2011). Therefore, this table is added in the Appendix as opposed to the main text.

The use of lagged variables to capture the behavioural dynamics in car ownership change and travel mode switching behaviour is quite slim. However, in the literature, the number of cars at time t has been used in car ownership change models (Oakil et al. 2014) and travel mode at time t has been used in travel mode switching model (Fatmi and Habib 2017). This has already been captured in the models in this study as the directionalities of the behavioural changes have been investigated which depends on the car ownership or travel mode at year t (for example, the behaviour of shifting from non-car ownership to car ownership state has been captured for the households who did not have a car in year t).

It should also be noted that the inclusion of state dependence has a potentially detrimental impact on models that is often ignored by analysts. Indeed, by including past choices in the utility for behaviour at time t, the behaviour is explained on the basis of past behaviour rather than explanatory variables. This creates issues with endogeneity (as the past behaviour is driven by the same underlying factors) and also removes explanatory power from the remaining variables in the model.

Table E2 MMNL estimation results of the residential mobility decision

Parameters	Household behaviour					
	Moved at Local level (LL)		Moved at regional level (RL)		Moved at national level (NL)	
	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat
Alternative specific constants						
(not moved is the base alternative)						
Mean	-4.9070	-18.8	-5.8119	-14.8	-5.8242	-11.0
Standard deviation	-0.5939	-6.5	0.8362	6.7	0.6173	3.2
Household level characteristics						
Household type (base is couple with child)						
Single member household	0.8130	4.9	0.8045	2.9	0.4464	1.3
Couple without child	0.3699	2.9	0.7161	3.3	0.5127	2.1
Lone parents	0.5184	3.4	0.0172	0.1	-0.3526	-1.0
Household income (base is less than £20,000)						
Between £20,00 to £40,000	0.1868	1.6	0.1890	1.0	0.4449	2.1
More than £40,000	0.1373	0.9	0.1813	0.7	0.5986	2.3
Education attainment of household head (base is below O level)						
O and A level degree	0.1152	1.1	0.6619	3.5	0.0722	0.4
Graduate degree	0.1153	0.8	0.5966	2.4	0.6058	2.6
Post-graduate degree	0.6935	2.9	0.9187	2.2	1.1954	3.6
Number of employees in the household (base is no employee)						
One employee	0.0094	0.1	-0.0095	-0.1	-0.1995	-0.9
More than one employees	0.2512	1.7	-0.3074	-1.3	-0.4132	-1.6
Length of current job of household head	-0.0044	-0.8	-0.0269	-2.3	-0.0126	-1.1
Presence of senior adult (>75 years)	-0.4600	-2.8	-0.5639	-1.7	0.3207	1.3
Dwelling level characteristics						
Tenure type (base is owned house)						
Rented social housing	0.5314	4.3	-0.0753	-0.3	-0.7224	-2.3
Rented private housing	1.7415	13.3	1.9428	10.5	1.2013	5.6
Crowd (household size\number of rooms)	1.2011	7.9	0.8225	3.2	0.8360	2.5
Life course events						
Having child in last one year	0.3062	1.8	0.5366	1.9	0.3276	1.0
Changed job in last one year	0.1146	1.0	0.0700	0.4	0.2123	1.0
Location characteristics						
Metropolitan area (base is other than London)						
London	0.0710	0.5	-2.5246	-3.6	1.5921	9.1
Stay length						
Linear	-0.0331	-2.7	-0.0318	-1.5	-0.0487	-2.7
Square	0.0001	0.4	0.0001	0.1	0.0006	1.7
Measures of model fit						
Number of observations	24718					
Initial LL	-34266.4					
Final LL	-5178.5					

Appendix F Testing correlations across the random terms in car ownership change model

Different nesting structures are tested in car ownership change model. For instance, nesting of alternatives owning the first car and additional cars vs nesting of losing first and additional cars, nesting of owning and losing of the first car vs nesting of owning and losing of additional cars, etc. Goodness fit of all the models captured different forms of nested structures are found better than the corresponding MNL estimation (Table E1 -Table E3) and found poor fit compared to the heteroscedastic model presented in chapter 4 (Table 4-8). The lower triangular matrix of Cholesky decomposition has been considered to capture the full range of correlations among the random parameters. A larger fit of the estimated model with Cholesky decomposition is observed compare to the models captured other forms of nesting structures. However, the improvement in the fit of the model with Cholesky decomposition is small compared to the heteroscedastic model. In case of all the estimated models, signs and magnitudes of the estimated parameters are found consistent.

Table F1 Estimation results for car ownership change model which captures the correlation between gaining and losing of first car, gaining and losing of additional cars

Variables	Changes in household car ownership level							
	Gained car				Lost car			
	0 to 1		1 to 2		1 to 0		2 to 1	
	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat
Alternative specific constants (no changes in car ownership is the base alternative)	-4.2709	-17.7	-4.8564	-33.8	-4.9878	-22.0	-1.9473	-12.2
Household level characteristics								
Household income	-0.0095	-1.4	0.0260	9.6	-0.0160	-2.3	-0.0116	-5.0
Change in household income (base is no change)								
Income increased	0.0202	0.1	0.5249	7.2	-	-	-	-
Income decreased	-	-	-	-	0.3980	2.7	0.3167	3.6
Household size	0.3475	5.4	0.2781	7.8	0.1428	2.5	-0.0917	-2.2
Change in household size (base is no change)								
Household size increased	1.6242	7.2	1.1689	9.6	-	-	-	-
Household size decreased	-	-	-	-	1.6567	8.4	2.0303	16.1
No of employees in the household	0.4472	4.2	0.7820	15.1	-0.0385	-0.4	-0.0571	-1.1
Change in number of employment (base is no change)								
Number of employment increased	1.0164	5.2	1.0744	10.7	-	-	-	-
Number of employment decreased	-	-	-	-	0.6846	3.5	0.6077	5.6
Presence of senior adults	-0.7073	-3.0	-0.8829	-4.3	0.8920	5.0	-0.3771	-1.7
Less educated people (below O level)	-0.6392	-3.5	-0.2733	-2.7	0.0714	0.4	0.4211	3.8
Dwelling characteristics								
Tenure type (base is owned house)								
Rented social housing	-0.4078	-2.2	-0.5851	-3.6	1.1986	6.7	0.9284	4.4
Rented private housing	0.2016	0.8	-0.5621	-3.2	0.8242	3.7	0.3203	1.4
Life course events								
Moved house								
Moved at local level	-0.0371	-0.1	0.3453	1.9	0.8510	3.2	0.6057	2.9
Moved at regional level	0.2445	0.6	0.7427	2.8	-0.1908	-0.3	-0.1337	-0.4
Moved at national level	2.0468	4.4	0.9021	3.2	0.8017	1.9	0.2142	0.6
Householder changed employer	0.0406	0.2	0.0882	0.9	0.0442	0.2	0.0038	0.0
Travel characteristics								
Travel distance	0.0113	2.3	-0.0076	-3.0	0.0024	0.5	-0.0017	-0.7
Change in travel distance (base is no change)								
Travel distance increased	0.4329	1.7	0.0163	0.2	-	-	-	-
Travel distance decreased	-	-	-	-	-0.1668	-0.7	0.2374	2.0
Correlated random parameters								
$\sigma_{01\text{-}10}$	1.3613	16.23						
$\sigma_{12\text{-}21}$	0.9569	21.12						
Measures of model fit								
Number of observations	24718							
Initial LL	-21985.800							
Final LL	-8038.660							

Table F2 Estimation results for car ownership change model which captures the correlation between gaining of first and additional cars, losing of first and additional cars

Variables	Changes in household car ownership level							
	Gained car				Lost car			
	0 to 1		1 to 2		1 to 0		2 to 1	
	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat
Alternative specific constants (no changes in car ownership is the base alternative)	-3.7992	-15.7	-5.6717	-29.27	-4.5697	-22.89	-1.1755	-7.65
Household level characteristics								
Household income	0.0047	0.88	0.0343	9.93	-0.0158	-2.45	-0.0142	-6
Change in household income (base is no change)								
Income increased	0.0957	0.62	0.5692	6.9	-	-	-	-
Income decreased	-	-	-	-	0.3726	2.6	0.3237	3.7
Household size	0.515	7.2	0.3544	7.65	0.162	3.12	-0.1361	-3.19
Change in household size (base is no change)								
Household size increased	1.8756	7.87	1.1816	8.4	-	-	-	-
Household size decreased	-	-	-	-	1.5771	8.65	1.9793	15.61
No of employees in the household	0.4113	3.73	1.0028	15.17	-0.1072	-1.17	-0.0989	-1.85
Change in number of employment (base is no change)								
Number of employment increased	1.1371	5.69	1.2198	10.57	-	-	-	-
Number of employment decreased	-	-	-	-	0.732	3.99	0.6428	5.91
Presence of senior adults	-1.0328	-4.34	-1.2887	-5.21	0.7409	4.51	-0.1277	-0.56
Less educated people (below O level)	-0.7354	-3.61	-0.588	-3.95	0.0937	0.68	0.4778	4.36
Dwelling characteristics								
Tenure type (base is owned house)								
Rented social housing	-0.6438	-3.04	-1.0559	-4.86	1.3648	8.53	1.0702	5.02
Rented private housing	-0.0583	-0.2	-0.9711	-4.24	1.0112	5	0.3025	1.34
Life course events								
Moved house								
Moved at local level	0.1525	0.48	0.4077	2.03	0.9255	3.79	0.5182	2.45
Moved at regional level	0.4451	0.92	0.8137	2.52	-0.3243	-0.85	-0.1355	-0.33
Moved at national level	2.1111	4.38	1.0759	3.6	0.8386	2.28	0.1868	0.52
Householder changed employer	-0.0084	-0.04	0.0972	0.88	0.0648	0.37	-0.0089	-0.09
Travel characteristics								
Travel distance	0.0259	5.18	-0.0084	-2.81	0.0041	0.94	-0.003	-1.27
Change in travel distance (base is no change)								
Travel distance increased	0.7131	2.91	0.081	0.72	-	-	-	-
Travel distance decreased	-	-	-	-	-0.1786	-0.76	0.1892	1.58
Correlated random parameters								
σ_{01-12}	1.8376	21.61						
σ_{10-21}	0.8446	13.4						
Measures of model fit								
Number of observations	24718							
Initial LL	-21985.800							
Final LL	-7966.170							

Table F3 Estimation results for car ownership change model which captures the full range of correlation by means of Cholesky decomposition

Variables	Changes in household car ownership level							
	Gained car				Lost car			
	0 to 1		1 to 2		1 to 0		2 to 1	
	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat
Alternative specific constants (no changes in car ownership is the base alternative)	-4.381	-15.5	-5.6702	-26.67	-5.0673	-18.77	-1.2666	-7.86
Household level characteristics								
Household income	-0.0043	-0.58	0.0361	10.14	-0.0108	-1.49	-0.0133	-5.71
Change in household income (base is no change)								
Income increased	0.063	0.4	0.6005	7.03	-	-	-	-
Income decreased	-	-	-	-	0.357	2.26	0.3041	3.5
Household size	0.4501	6.22	0.3835	7.88	0.1118	1.71	-0.0998	-2.42
Change in household size (base is no change)								
Household size increased	1.7046	6.84	1.227	8.46	-	-	-	-
Household size decreased	-	-	-	-	1.7063	8.12	1.9577	15.54
No of employees in the household	0.4125	3.47	1.0085	14.03	-0.0553	-0.52	-0.1112	-2.11
Change in number of employment (base is no change)								
Number of employment increased	1.033	4.96	1.2298	10.12	-	-	-	-
Number of employment decreased	-	-	-	-	0.7099	3.49	0.6287	5.76
Presence of senior adults	-1.1299	-3.91	-1.0731	-4.35	0.8836	4.52	-0.0876	-0.4
Less educated people (below O level)	-0.6941	-3.43	-0.4576	-2.97	0.1451	0.78	0.3738	3.65
Dwelling characteristics								
Tenure type (base is owned house)								
Rented social housing	-0.3634	-1.68	-0.8845	-3.8	1.4688	6.83	0.8803	4.19
Rented private housing	-0.1839	-0.59	-0.8753	-3.57	0.7441	2.95	0.0388	0.17
Life course events								
Moved house								
Moved at local level	0.1444	0.44	0.4809	2.3	0.9156	3.23	0.5672	2.64
Moved at regional level	0.4711	0.66	0.6737	2.04	-0.263	-0.43	-0.1454	-0.34
Moved at national level	2.2247	4.11	1.0286	2.87	0.8173	1.72	0.2089	0.59
Householder changed employer	0.0285	0.14	0.1364	1.19	0.0763	0.4	-0.0064	-0.06
Travel characteristics								
Travel distance	0.016	2.89	-0.0068	-2.1	0.004	0.69	-0.0028	-1.2
Change in travel distance (base is no change)								
Travel distance increased	0.4022	1.43	0.0819	0.71	-	-	-	-
Travel distance decreased	-	-	-	-	-0.1432	-0.55	0.1714	1.44
Correlated random parameters								
$\sigma_{01\text{-}01}$	1.6469	10.1						
$\sigma_{12\text{-}01}$	0.1761	1.07						
$\sigma_{12\text{-}12}$	2.0305	20.14						
$\sigma_{10\text{-}01}$	0.8313	7.32						
$\sigma_{10\text{-}12}$	0.2461	1.64						
$\sigma_{10\text{-}10}$	1.401	7.74						
$\sigma_{21\text{-}01}$	0.3498	3.16						
$\sigma_{21\text{-}12}$	0.0888	1.03						
$\sigma_{21\text{-}10}$	0.6826	8.21						
$\sigma_{21\text{-}21}$	Fixed	-						
Measures of model fit								
Number of observations	24718							
Initial LL	-21985.800							
Final LL	-7845.820							

Appendix G Testing alternative variable specifications in car ownership change model

Before developing the models, the potential impact of the attributes on the choice outcomes is hypothesised based on literature survey. It is assumed that changes in the household's circumstances may only influence a few specific directions of car ownership change behaviour. For instance, in the case of the income parameter, it is assumed that an increase in household income may increase the propensity of gaining a car but may not have any influence on the likelihood of losing a car. On the other hand, a decrease in income may increase the probability of losing a car but may not increase the car owning propensity. The estimation results support the hypothesis and found insignificant estimates in a few directions of interdependencies. For example, the impact of an increase in household income on the alternatives for losing cars (both first and additional) are found statistically insignificant (Table E4). Therefore, the final model excludes the parameters having insignificant connections.

It is also tested whether the model without the residential mobility parameters is not statistically different from the model that includes these parameters. The chi-square test rejects the null hypothesis (LR=55.21, Chi-square stat=32.91 degree of freedom=12, confidence interval = 99.9%). The result indicates that the model without residential mobility parameters is significantly worse. Outputs of the estimated model without the residential mobility parameters are presented in Table G2.

Table G1 Estimation results for car ownership change model where all directions of relation between income parameter and choices are tested.

Variables	Changes in household car ownership level							
	Gained car				Lost car			
	0 to 1		1 to 2		1 to 0		2 to 1	
	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat
Alternative specific constants (no changes in car ownership is the base alternative)								
Mean	-4.2618	-14.0	-5.7411	-26.3	-5.1482	-18.4	-1.2575	-8.0
Standard deviation	1.7832	11.5	1.9806	20.4	-1.7524	-10.3	-0.7859	-11.0
Household level characteristics								
Household income	-0.0034	-0.5	0.0356	9.5	-0.0140	-2.0	-0.0135	-5.8
Change in household income (base is no change)								
Income increased	0.0387	0.2	0.5897	6.0	-0.0420	-0.3	-0.0432	-0.5
Income decreased	0.0433	0.2	0.0172	0.4	0.3556	2.1	0.2938	2.8
Household size	0.4328	5.3	0.3706	7.5	0.1429	2.2	-0.1129	-2.7
Change in household size (base is no change)								
Household size increased	1.8194	7.3	1.2424	8.6	-	-	-	-
Household size decreased	-	-	-	-	1.6332	7.9	1.9480	15.5
No of employees in the household	0.5185	4.2	1.0474	14.7	-0.1167	-1.1	-0.0996	-1.9
Change in number of employment (base is no change)								
Number of employment increased	1.2714	6.2	1.2785	10.6	-	-	-	-
Number of employment decreased	-	-	-	-	0.7059	3.4	0.6471	5.9
Presence of senior adults	-0.8065	-3.0	-1.2586	-4.9	1.0048	4.9	-0.0415	-0.2
Less educated people (below O level)	-0.4512	-2.0	-0.5039	-3.2	0.3142	1.7	0.4251	4.1
Dwelling characteristics								
Tenure type (base is owned house)								
Rented social housing	-0.6416	-2.8	-0.7210	-3.2	1.5736	7.3	0.8765	4.5
Rented private housing	-0.1126	-0.3	-0.7946	-3.4	1.1128	4.5	0.2291	1.1
Life course events								
Moved house								
Moved at local level	0.1567	0.5	0.4425	2.2	0.9303	3.6	0.5200	2.4
Moved at regional level	0.4598	0.9	0.6632	2.3	-0.2250	-0.4	-0.1652	-0.4
Moved at national level	2.2364	7.2	0.9813	2.7	0.8597	2.0	0.1804	0.5
Householder changed employer	0.0169	0.1	0.1333	1.2	0.0525	0.3	0.0525	0.3
Travel characteristics								
Travel distance	0.0183	3.1	-0.0053	-1.7	0.0060	1.2	-0.0027	-1.2
Change in travel distance (base is no change)								
Travel distance increased	0.4523	1.7	0.1084	0.9	-	-	-	-
Travel distance decreased	-	-	-	-	-0.0829	-0.3	0.1780	1.5
Measures of model fit								
Number of observations	24718.000							
Initial LL	-21985.800							
Final LL	-7851.020							

Table G2 Estimation results of car ownership change model without the residential mobility parameters.

Variables	Changes in household car ownership level							
	Gained car				Lost car			
	0 to 1		1 to 2		1 to 0		2 to 1	
	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat
Alternative specific constants (no changes in car ownership is the base alternative)								
Mean	-4.176	-14.4	-5.709	-26.3	-5.149	-18.1	-1.268	-8.4
Standard deviation	1.763	11.5	1.974	20.5	1.764	10.3	0.787	11.0
Household level characteristics								
Household income	-0.002	-0.3	0.036	9.8	-0.013	-1.8	-0.014	-5.8
Change in household income (base is no change)								
Income increased	0.043	0.3	0.580	6.8	-	-	-	-
Income decreased	-	-	-	-	0.376	2.4	0.320	3.7
Household size	0.427	5.3	0.371	7.5	0.151	2.3	-0.111	-2.7
Change in household size (base is no change)								
Household size increased	1.795	6.9	1.282	8.8	-	-	-	-
Household size decreased	-	-	-	-	1.627	7.7	1.965	15.7
No of employees in the household	0.510	4.2	1.039	14.5	-0.140	-1.3	-0.105	-2.0
Change in number of employment (base is no change)								
Number of employment increased	1.229	5.8	1.267	10.5	-	-	-	-
Number of employment decreased	-	-	-	-	0.770	3.8	0.650	6.0
Presence of senior adults	-0.827	-3.1	-1.288	-4.6	0.987	4.8	-0.050	-0.2
Less educated people (below O level)	-0.480	-2.2	-0.513	-3.2	0.289	1.5	0.424	4.0
Dwelling characteristics								
Tenure type (base is owned house)								
Rented social housing	-0.671	-3.0	-0.694	-3.0	1.585	7.0	0.884	4.3
Rented private housing	0.003	0.1	-0.622	-2.7	1.266	5.1	0.282	1.3
Life course events								
Moved house								
Moved at local level	-	-	-	-	-	-	-	-
Moved at regional level	-	-	-	-	-	-	-	-
Moved at national level	-	-	-	-	-	-	-	-
Householder changed employer	-0.037	-0.2	0.151	1.3	0.057	0.3	0.016	0.2
Travel characteristics								
Travel distance	0.017	2.9	-0.005	-1.6	0.006	1.2	-0.003	-1.1
Change in travel distance (base is no change)								
Travel distance increased	0.589	2.2	0.122	1.1	-	-	-	-
Travel distance decreased	-	-	-	-	-0.035	-0.1	0.180	1.5
Measures of model fit								
Number of observations	24718.000							
Initial LL	-21985.800							
Final LL	-7880.240							

Appendix H Testing correlations across the random terms in travel mode switching models

Correlations among the different switching options have also been investigated using nesting structures. Potential nesting structures are: nesting of the alternatives of switching to car, public transport and active travel; nesting of the alternatives of switching to public transport and private travel (car and active travel); nesting of the alternatives of switching to motorized travel (car, public transport) and non-motorized travel (active travel), etc. The models captured different forms of nested structures mentioned above are found to have a better fit compared to the corresponding MNL estimation (Table F1-Table F4) and a poor fit is observed compared to the heteroscedastic model presented in chapter 4 (Table 4-9). The estimated model with Cholesky decomposition is found to have a better fit compare to the other models. However, the number of parameters required to estimate to capture the full range of correlation using Cholesky decomposition increases significantly with the increase of the number of alternatives in the choice set which is cumbersome in many cases. For J number of alternatives, total additional parameters to estimate for Cholesky decomposition is $\frac{J*(J-1)}{2} - 1$ (e.g. 20 additional parameters are required to estimate for 7 choices).

Table H1 Estimation results for travel mode switching behaviour that captures the correlation of switching to motorized vehicles (public transport and car) and non-motorized vehicle (active travel).

Variables	Travel mode switching behaviour					
	Switched to CT (from PT & AT)		Switched to PT (from CT & AT)		Switched to AT (from CT & PT)	
	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat
Alternative specific constants						
(no changes in travel mode is the base alternative)						
Switched from car travel (CT)	-	-	-5.7178	-10.6	-5.4066	-7.2
Switched from public transport (PT)	-7.2472	-16.5	-	-	-4.0216	-8.1
Switched from active travel (AT)	-5.1319	-16.6	-4.7813	-16.6	-	-
Household owns car	1.6595	7.3	-2.4140*	-5.3	-0.7902*	-1.2
Changes in car ownership						
Household acquired car	2.1376	8.4	-2.5851*	-3.2	-0.8801*	-1.7
Household relinquished car	0.2505	0.7	0.9562*	3.1	1.2132*	4.0
Moved house						
Moved at local level	-0.0270	-0.1	-0.2828	-0.5	-0.3801	-0.7
Moved at regional level	0.1257	0.2	1.959	3.8	1.2099	1.8
Moved at national level	1.4619	2.7	1.7593	2.8	-0.2621	-0.2
Householder changed employer	0.1801	0.8	-0.0956	-0.4	-0.0767	-0.3
Travel distance	-0.0077	-1.2	0.028	5.9	-0.0765	-7.8
Changes in travel distance						
Travel distance increased	3.7495	15.0	3.3034	11.8	-1.882	-1.3
Travel distance decreased	3.4644	11.7	1.1759	3.2	3.9807	11.2
Correlated random parameters						
$\sigma_{CT\text{-}PT,AT\text{-}PT,PT\text{-}CT,AT\text{-}CT}$	1.5988	11.4				
$\sigma_{PT\text{-}AT,CT\text{-}AT}$	1.5368	7.1				
Measures of model fit						
Number of observations	10704					
Initial LL	-11759.5					
Final LL	-1804.9					

* parameters represent switching from car travel only

Table H2 Estimation results for travel mode switching behaviour that captures the correlation of switching to public transport and private travel options (car and active travel).

Variables	Travel mode switching behaviour					
	Switched to CT (from PT & AT)		Switched to PT (from CT & AT)		Switched to AT (from CT & PT)	
	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat
Alternative specific constants						
(no changes in travel mode is the base alternative)						
Switched from car travel (CT)	-	-	-5.7923	-7.48	-5.5528	-8.12
Switched from public transport (PT)	-5.9517	-15.83	-	-	-3.7492	-8.91
Switched from active travel (AT)	-5.7141	-17.29	-6.7452	-12.67	-	-
Household owns car	1.4812	6.68	-2.8793*	-4.73	-0.5772*	-0.89
Changes in car ownership						
Household acquired car	2.141	8.56	-2.7055*	-3.39	-0.8865*	-1.78
Household relinquished car	0.3772	1.11	0.9209*	2.69	1.1195*	3.99
Moved house						
Moved at local level	0.0689	0.19	-0.1764	-0.28	-0.3205	-0.61
Moved at regional level	0.3463	0.35	2.2441	3.23	1.0867	1.62
Moved at national level	1.5804	2.99	1.5034	1.83	-0.0991	-0.08
Householder changed employer	0.2531	1.19	-0.1096	-0.4	-0.1733	-0.71
Travel distance	-0.0116	-2	0.0296	5.18	-0.0687	-7.73
Changes in travel distance						
Travel distance increased	3.5744	14.29	3.2491	10.5	-1.8576	-1.65
Travel distance decreased	3.3303	11.44	1.1363	2.85	3.8817	12.06
Correlated random parameters						
$\sigma_{CT\text{-}PT,AT\text{-}PT}$	2.3936	9.04				
$\sigma_{PT\text{-}CT,AT\text{-}CT,PT\text{-}AT,CT\text{-}AT}$	1.3227	11.28				
Measures of model fit						
Number of observations	10704					
Initial LL	-11759.5					
Final LL	-1803.1					

* parameters represent switching from car travel only

Table H3 Estimation results for travel mode switching behaviour that captures the correlation of switching to public transport, car and active travel.

Variables	Travel mode switching behaviour					
	Switched to CT (from PT & AT)		Switched to PT (from CT & AT)		Switched to AT (from CT & PT)	
	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat
Alternative specific constants						
(no changes in travel mode is the base alternative)						
Switched from car travel (CT)	-	-	-5.6515	-8.11	-5.2659	-7.63
Switched from public transport (PT)	-6.6634	-15.05	-	-	-4.0682	-9.59
Switched from active travel (AT)	-5.5516	-15.17	-6.5926	-13.93	-	-
Household owns car	1.8354	6.8	-2.8099*	-5.55	-0.7494*	-1.15
Changes in car ownership						
Household acquired car	2.2961	8.29	-2.9164*	-3.91	-0.8530*	-1.71
Household relinquished car	0.609	1.68	0.8360*	2.45	1.1548*	3.97
Moved house						
Moved at local level	-0.2017	-0.53	-0.2141	-0.39	-0.3348	-0.77
Moved at regional level	-0.4239	-2.74	2.1083	3.47	1.1837	2.63
Moved at national level	1.5235	2.93	1.4349	2.98	-0.1336	-0.16
Householder changed employer	0.1713	0.73	-0.1029	-0.38	-0.0032	-0.01
Travel distance	-0.0074	-1.19	0.0295	5.64	-0.0766	-8.25
Changes in travel distance						
Travel distance increased	3.908	13.43	3.186	10.52	-1.8928	-1.61
Travel distance decreased	3.606	11.08	1.1267	2.93	3.9585	12.49
Correlated random parameters						
$\sigma_{PT\text{-}CT,AT\text{-}CT}$	1.7025	8.84				
$\sigma_{CT\text{-}PT,AT\text{-}PT}$	2.2507	9.72				
$\sigma_{PT\text{-}AT,CT\text{-}AT}$	1.3522	7.08				
Measures of model fit						
Number of observations	10704					
Initial LL	-11759.5					
Final LL	-1798.6					

* parameters represent switching from car travel only

Table H4 Estimation results for travel mode switching behaviour that captures the full range of correlation by means of Cholesky decomposition

Variables	Travel mode switching behaviour					
	Switched to CT (from PT & AT)		Switched to PT (from CT & AT)		Switched to AT (from CT & PT)	
	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat
Alternative specific constants						
(no changes in travel mode is the base alternative)						
Switched from car travel (CT)	-	-	-4.5688	-8.9	-5.2743	-7.3
Switched from public transport (PT)	-6.8547	-16.0	-	-	-4.1382	-8.6
Switched from active travel (AT)	-5.5714	-16.3	-6.6264	-11.2	-	-
Household owns car	1.6474	6.7	-2.8414*	-6.5	-0.3733*	-0.6
Changes in car ownership						
Household acquired car	2.2533	8.5	-2.5432*	-3.6	-1.0003*	-2.0
Household relinquished car	0.6706	2.0	0.7981*	2.7	1.0152*	3.8
Moved house						
Moved at local level	-0.0671	-0.2	-0.0401	-0.1	-0.2699	-0.5
Moved at regional level	-0.0689	-0.1	2.0313	3.9	1.1120	1.9
Moved at national level	1.6490	2.7	1.5417	2.5	-0.3794	-0.4
Householder changed employer	0.2523	1.1	-0.1370	-0.6	0.0814	0.4
Travel distance	-0.0017	-0.3	0.0269	6.4	-0.0815	-8.1
Changes in travel distance						
Travel distance increased	3.6253	14.7	3.0802	10.6	-2.0464	-1.6
Travel distance decreased	3.1513	10.6	1.0961	2.9	3.8018	11.6
Correlated random parameters						
$\sigma_{PT\text{-}CT,PT\text{-}CT}$	1.8619	7.0				
$\sigma_{PT\text{-}AT,PT\text{-}CT}$	0.6348	2.6				
$\sigma_{PT\text{-}AT,PT\text{-}AT}$	1.4785	4.0				
$\sigma_{CT\text{-}PT,PT\text{-}CT}$	0.9799	4.7				
$\sigma_{CT\text{-}PT,PT\text{-}AT}$	0.4772	2.4				
$\sigma_{CT\text{-}PT,CT\text{-}PT}$	0.5749	2.9				
$\sigma_{CT\text{-}AT,PT\text{-}CT}$	0.3890	2.2				
$\sigma_{CT\text{-}AT,PT\text{-}AT}$	0.1924	0.7				
$\sigma_{CT\text{-}AT,CT\text{-}PT}$	0.8512	4.1				
$\sigma_{CT\text{-}AT,CT\text{-}AT}$	0.3488	1.4				
$\sigma_{AT\text{-}PT,PT\text{-}CT}$	1.3793	4.4				
$\sigma_{AT\text{-}PT,PT\text{-}AT}$	1.2662	2.9				
$\sigma_{AT\text{-}PT,CT\text{-}PT}$	0.4141	1.2				
$\sigma_{AT\text{-}PT,CT\text{-}AT}$	1.4570	4.3				
$\sigma_{AT\text{-}PT,AT\text{-}PT}$	0.6198	1.8				
$\sigma_{AT\text{-}CT,PT\text{-}CT}$	0.1116	0.7				
$\sigma_{AT\text{-}CT,PT\text{-}AT}$	0.1239	0.7				
$\sigma_{AT\text{-}CT,CT\text{-}PT}$	0.9581	4.7				
$\sigma_{AT\text{-}CT,CT\text{-}AT}$	0.2000	0.9				
$\sigma_{AT\text{-}CT,AT\text{-}PT}$	0.5535	2.9				
$\sigma_{AT\text{-}CT,AT\text{-}CT}$	Fixed	-				
Measures of model fit						
Number of observations	10704					
Initial LL	-11759.5					
Final LL	-1768.4					

* parameters represent switching from car travel only

Appendix I Testing alternative variable specifications in commute mode switching model

Model is estimated dropping the residential mobility parameters. The hypothesis is that "the model without the residential mobility parameters is not statistically different from the model that includes these parameters". The chi-square test rejects the null hypothesis (LR=30.62, Chi-square stat=27.88, degree of freedom=9, confidence interval=99.9 %). The result indicates that the model without residential mobility parameters is significantly worse. Outputs of the estimated model without the residential mobility parameters are presented in Table I2.

Table I1 Estimation results of commute mode switching model without the residential mobility parameters.

Variables	Travel mode switching behaviour					
	Switched to CT (from PT & AT)		Switched to PT (from CT & AT)		Switched to AT (from CT & PT)	
	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat
Alternative specific constants						
(no changes in travel mode is the base alternative)						
Mean						
Switched from car travel (CT)	-	-	-5.0072	-7.5	-5.6956	-6.9
Switched from public transport (PT)	-6.6654	-13.5	-	-	-3.6910	-8.3
Switched from active travel (AT)	-5.3495	-15.3	-6.5443	-5.5	-	-
Standard deviation						
Switched from car	-	-	-1.8540	-5.2	-1.5470	-5.7
Switched from public transport	-2.4463	-7.0	-	-	0.7570	1.6
Switched from active travel	0.7691	3.1	2.5338	2.7	-	-
Household owns car	1.7398	6.6	-2.8694*	-4.9	-0.5399*	-0.7
Changes in car ownership						
Household acquired car	2.2215	8.3	-2.3334*	-2.4	-0.9097*	-1.6
Household relinquished car	0.5235	1.4	0.8240*	2.6	1.1963*	4.0
Moved house						
Moved at local level	-	-	-	-	-	-
Moved at regional level	-	-	-	-	-	-
Moved at national level	-	-	-	-	-	-
Householder changed employer	0.1491	0.7	-0.1190	-0.5	-0.0384	-0.2
Travel distance	-0.0027	-0.4	0.0257	5.2	-0.0726	-6.9
Changes in travel distance						
Travel distance increased	3.7838	14.4	3.2274	10.7	-1.8269	-1.1
Travel distance decreased	3.1498	9.5	1.2220	3.2	3.9606	11.5
Measures of model fit						
Number of observations	10704					
Initial LL	-11759.5					
Final LL	-1799.6					

* parameters represent switching from car travel only

Appendix J Conversion of WAD and TAZ data

J1 Conversion of WAD data from new ward boundaries to old ward boundaries

Greater London area electoral ward boundaries have been changed significantly in 2002. New electoral ward boundaries were considered in the WAD (after the changes in 2002) and old ward boundaries (before the changes in 2002) were considered in the LHSD. The layer function in ArcGIS was used to investigate the physical changes of new ward boundaries after 2002. Three scenarios are observed during the conversion

Scenario 1

No change or minimal changes in old ward boundaries after 2002. For example, the old ward named Park (P16) is renamed as Noel Park (N15) in 2002 but the boundary of the ward remains the same (Figure G1).

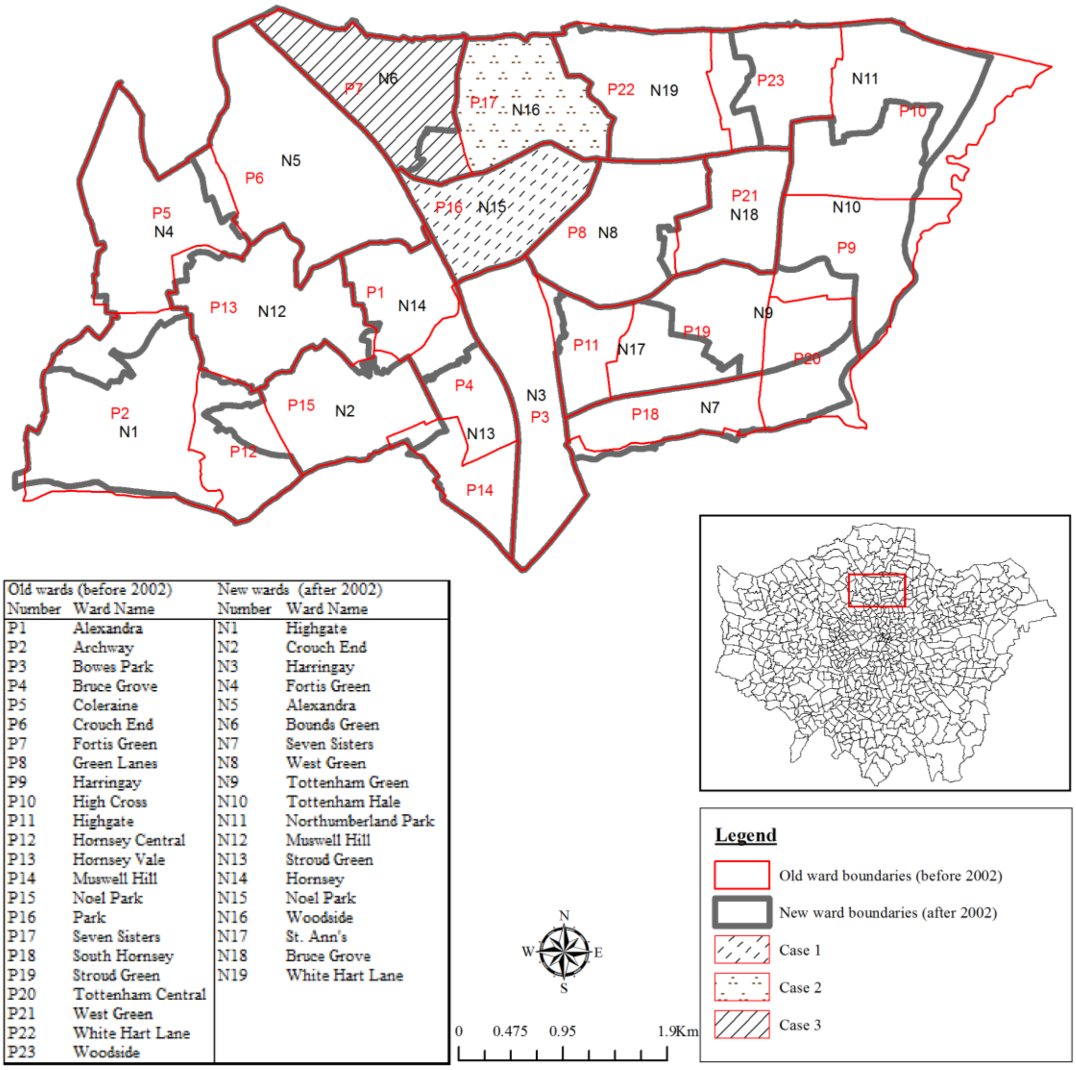
Scenario 2

The old ward area consists of part of a new ward area. For example, the ward called seven sister (P17) has formed part of the new area ward called Woodside (N16). Although, the new ward Woodside also consists of part of the old ward called Fortis Green (P7).

Scenario 3

The old ward area shared multiple new ward areas. For example, the old ward Fortis Green (P7) has shared a part of a new ward called Bounds Green (N6) and rest from Woodside (N16).

Since old ward boundary constitutes of whole or part of a new ward boundary in case of scenarios1 and 2, the attributes of the new wards under scenarios1 and 2 are used as attributes for the corresponding old wards. In case of scenario 3, where the old ward area was found to be shared across multiple new ward areas, the weighted averages of shared new wards attributes were estimated for the old ward. For example, if the old ward area P7 comprised of 20% area from new ward N16 and rest 80% area from new ward N6, for attributes like crime rate, the crime rate at ward P7 was calculated as the sum of 0.20*crime rate at N16 and 0.8*crime rate at N6. The attributes are assumed to be constant within each new ward.

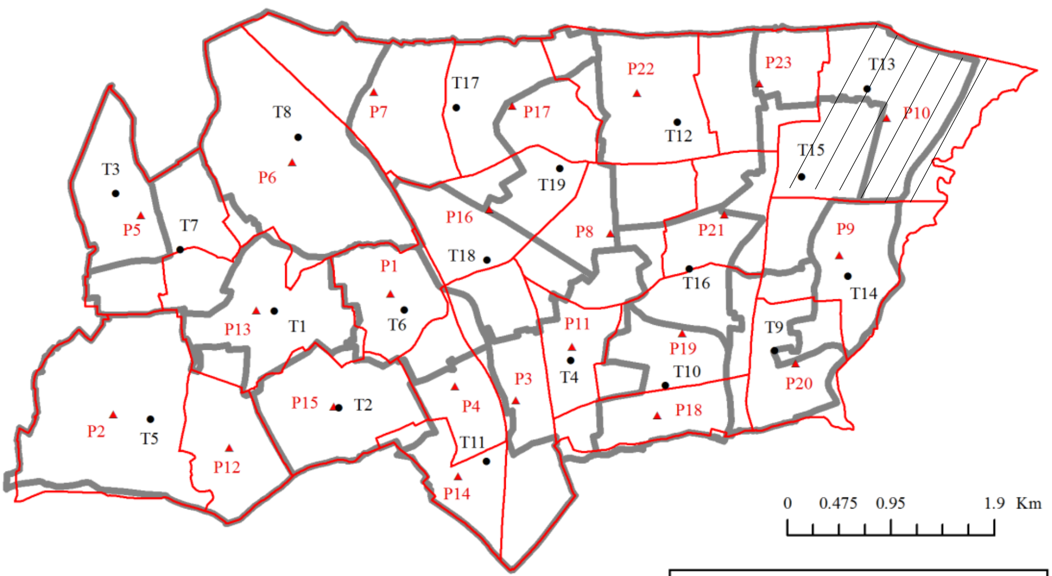


Old wards (before 2002) Number	Ward Name	New wards (after 2002) Number	Ward Name
P1	Alexandra	N1	Highgate
P2	Archway	N2	Crouch End
P3	Bowes Park	N3	Harringay
P4	Bruce Grove	N4	Fortis Green
P5	Coleraine	N5	Alexandra
P6	Crouch End	N6	Bounds Green
P7	Fortis Green	N7	Seven Sisters
P8	Green Lanes	N8	West Green
P9	Harringay	N9	Tottenham Green
P10	High Cross	N10	Tottenham Hale
P11	Highgate	N11	Northumberland Park
P12	Hornsey Central	N12	Muswell Hill
P13	Hornsey Vale	N13	Stroud Green
P14	Muswell Hill	N14	Hornsey
P15	Noel Park	N15	Noel Park
P16	Park	N16	Woodside
P17	Seven Sisters	N17	St. Ann's
P18	South Hornsey	N18	Bruce Grove
P19	Stroud Green	N19	White Hart Lane
P20	Tottenham Central		
P21	West Green		
P22	White Hart Lane		
P23	Woodside		

Figure J1 Conversion of new wards to equivalent old wards

J2 Conversion of LTSM data from TAZ boundaries to old ward boundaries

The distances of alternative locations (wards) from the individual workplace (ward) and CBD are extracted from LTSM. LTSM consists of distance matrix at the Level of TAZs which are different from the old ward boundaries. Therefore, the TAZs are converted to equivalent old wards based on maximum overlapping of the areas using Arc GIS. Centre to centre distances between the overlapped wards and zones are used to identify the best match between wards and corresponding TAZs. For example, the old ward High Cross (P10) has formed parts of TAZs called Tottenham East (T13) and Tottenham Hale West (T15) (Figure G2). Since the centre of the High Cross (P10) is closed to the centre of the TAZ Tottenham East (T13), Tottenham East (T13) is replaced by High Cross (P10). After converting the TAZ to the equivalent old ward, TAZ pair distance matrix is converted to ward pair (old ward) distance matrix. Centre to centre distance is considered. Some inaccuracy might be involved in this process since the centre of the ward boundaries and the equivalent TAZ boundaries are not the same or close in many cases.



Old wards (before 2002)		TAZ (2011)	
Number	Ward Name	Number	TAZ Name
P1	Alexandra	T1	Cranley Gardens
P2	Archway	T2	Crouch End
P3	Bowes Park	T3	Fortis Green North
P4	Bruce Grove	T4	Haringey
P5	Coleraine	T5	Highgate
P6	Crouch End	T6	Hornsey
P7	Fortis Green	T7	Muswell Hill
P8	Green Lanes	T8	Muswell Hill North and Alexandria
P9	Harringay	T9	South Tottenham
P10	High Cross	T10	Stamford Hill North
P11	Highgate	T11	Stroud Green and Finsbury Park
P12	Hornsey Central	T12	Tottenham
P13	Hornsey Vale	T13	Tottenham East
P14	Muswell Hill	T14	Tottenham Hale East
P15	Noel Park	T15	Tottenham Hale West
P16	Park	T16	Tottenham South
P17	Seven Sisters	T17	Wood Green
P18	South Hornsey	T18	Wood Green South
P19	Stroud Green	T19	Wood Green West
P20	Tottenham Central		
P21	West Green		
P22	White Hart Lane		
P23	Woodside		

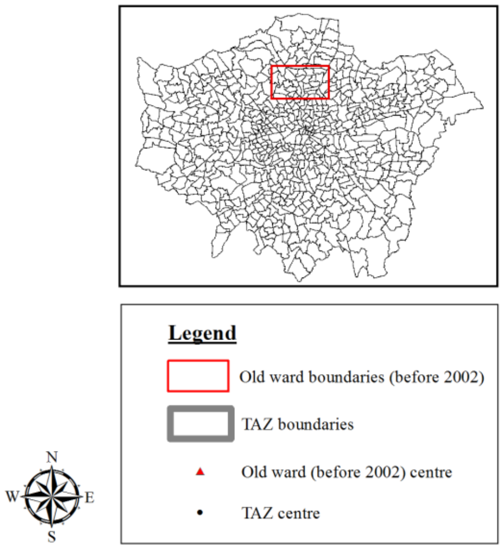


Figure J2 Conversion of Traffic analysis zones (TAZs) to equivalent old wards

Appendix K Distribution of the characteristics of the households in the LHSD and in the estimation subsample

Table K1 Characteristics of the households in the subsample and the full sample

Variables	Subsample	Full sample
Socio-demographic characteristics		
Annual household income		
Less than £30,000	41.6%	64.0%
Between £30,000 to £60,000	29.9%	18.2%
More than £60,000	28.5%	17.8%
Average household size (members in the household)	2.8	2.6
Household composition		
Married couple with and without kids	45.5%	35.0%
Cohabiting couple with and without kids	15.0%	9.9%
Single member household	25.6%	43.2%
Household having more than one member	13.9%	11.8%
Ethnic composition		
White people	78.6%	76.6%
Asian people	13.1%	10.7%
Black people	8.2%	12.7%
Location and dwelling features		
Residential location		
Inner London	34.5%	40.6%
Outer London	65.5%	59.4%
Average dwelling size (number of bedrooms)		
Inner London	2.5	2.5
Outer London	2.8	2.6
Average tenure length (in years)		
Inner London	7.5	10.1
Outer London	9.0	12.6
Travel behaviour		
Car ownership		
Inner London	69.2%	49.0%
Outer London	85.0%	70.0%
Sample size	2700	8159

Appendix L Correlation matrix of the independent variables

Variables	Commute distance	Flat house in inner London	Flat house in outer London	Detached house in inner London	Detached house in outer London	Ratio of black people	Ratio of asian people	Ratio of white people	Commer cial land area	Domestic land area in outer London	Domestic land area in inner London	Land use mix	School quality	Crime rate	Dwelling density in inner London	Dwelling density in outer London	Public transport acc_car owner	Public transport acc_non car owner	Distance from CBD	Housing cost for high income people	Housing cost for medium income people	Housing cost for low income people	House hold size	Employ ment density
Commute distance	1																							
Flat house in inner London	0.018	1																						
Flat house in outer London	-0.048	-0.178	1																					
Detached house in inner London	-0.016	-0.273	0.06	1																				
Detached house in outer London	-0.051	0.067	0.106	-0.045	1																			
Ratio of black people	0.02	0.015	0.012	0.057	0.052	1																		
Ratio of asian people	-0.001	-0.037	-0.093	0.089	0.027	-0.035	1																	
Ratio of white people	0.027	0.021	0.043	-0.096	-0.057	0.113	0.111	1																
Commerial land area	-0.008	0.08	0.05	-0.048	0.013	0.078	0.162	-0.13	1															
Domestic land area in outer London	0.033	-0.029	-0.352	-0.011	0.253	0.016	0.049	-0.05	-0.084	1														
Domestic land area in inner London	0.021	-0.367	-0.038	0.061	-0.025	0.046	0.002	-0.016	-0.135	-0.06	1													
Land use mix	0.001	-0.013	0.067	-0.037	-0.138	-0.014	-0.072	0.013	0.472	0.013	-0.138	1												
School quality	0.008	-0.007	0.04	0.048	0.139	-0.065	0.064	0.006	-0.029	0.194	0	-0.048	1											
Crime rate	0.019	0.042	0.094	-0.022	0.027	0.037	0.057	-0.066	0.017	-0.032	0.129	0.089	-0.087	1										
Dwelling density in inner London	-0.015	0.377	0.011	-0.181	-0.007	-0.019	0.005	-0.008	0.166	0.037	0.684	-0.034	-0.038	-0.121	1									
Dwelling density in outer London	-0.041	-0.016	0.462	0.007	-0.318	0.012	0	-0.003	0.057	0.653	0.043	-0.013	-0.156	0.005	-0.063	1								
Public transport acc_car owner	-0.019	0.165	0.21	0.027	-0.07	-0.047	0.049	-0.043	0.2	0.147	0.102	0.038	0.079	0.184	0.067	-0.019	1							
Public transport acc_non car owner	-0.002	-0.026	-0.056	0.011	0.029	-0.018	-0.037	0.034	0.06	-0.043	-0.036	-0.062	0.005	0.01	0.036	0.033	0.293	1						
Distance from CBD	0.19	-0.189	-0.213	0.138	0.166	-0.125	-0.051	0.241	0.06	-0.005	-0.06	-0.091	0.018	-0.027	0.044	-0.067	-0.125	0.013	1					
Housing cost for high income people	-0.029	-0.008	-0.005	-0.014	0.069	-0.069	-0.05	0.202	0.012	0.019	0.037	-0.048	0.074	-0.015	0.03	-0.008	0.061	-0.041	-0.141	1				
Housing cost for medium income people	-0.04	0	0	-0.022	0.089	-0.086	-0.06	0.188	0.013	0.032	0.036	-0.048	0.092	-0.028	0.022	-0.02	0.035	0.01	-0.166	-0.145	1			
Housing cost for low income people	-0.038	0.005	0.007	-0.028	0.096	-0.109	-0.065	0.187	0.009	0.045	0.033	-0.049	0.081	-0.04	0.019	-0.033	0.022	0.046	-0.188	-0.145	-0.151	1		
Household size	-0.011	0.011	0.016	-0.01	-0.009	-0.008	-0.198	0	-0.013	-0.005	-0.004	0.008	0.03	0.006	-0.005	0.001	0.008	-0.047	0.008	0.006	-0.008	-0.02	1	
Employment density	-0.033	-0.02	-0.055	0.006	-0.058	-0.048	-0.094	0.089	0.262	0.047	-0.069	-0.21	0.073	0.808	-0.053	-0.072	-0.073	0.027	0.042	0.065	0.054	0.058	-0.003	1

Appendix M Differences in the preferences of households moved in the different time periods

To investigate the differences in the residential location choice behaviour of the households who moved in the different time periods and still prefer their current places to live, the dataset is divided into four subgroups based on the year households moved to the current locations. The four subgroups are names as TP1 (households moved before 1990), TP2 (households moved between 1990-1996), TP3 (households moved between 1997-2000) and TP4 (households moved between 2001-2002). It may be noted that it was difficult to identify any intuitive breakpoints of the sample subdivision. Since many households, specifically renters, moved into their current locations in the recent past, the sample has been subdivided considering longer length for the older time periods (TP1 and TP2) to ensure representative sample in all the time periods (TP). Separate models are estimated for the subgroups using the MMNL technique explained in chapter 5. These models are estimated using the same specification of the final pooled model in chapter 5. It may be noted that given the pooled structure of the model, the effect of the disproportional share of owners and renters in the sub-datasets on the estimation results is expected to be minimal, given that the majority of parameters are pooled across owners and renters. The model findings are presented in Table J1.

It is observed that the estimated parameters are different in different models and many parameters are statistically significant in one model and insignificant in another model. To compare the estimated parameters of these separate models, elasticity analysis is conducted which is presented in Table 5-5. The elasticities of the parameters in the one model are found to be different in another model and some cases the sign has changed. However, some parameters show a trend (either increasing or decreasing) of changing the elasticity or the sensitivity over time (from TP1 to TP4). A few parameters which had shown a trend of changing sensitivities considerably over the years are presented graphically in Figure 5-2 and further discussed below

Table M1 Estimation of models for households moved in different time periods (TP)

Parameters	Year household moved: TP1 (Before 90) Coeff.	t-stat	TP2 (1990-1996) Coeff.	t-stat	TP3 (1997-2000) Coeff.	t-stat	TP4 1 (2001-2002) Coeff.	t-stat	t diff. TP1 and TP4
Constants									
Central London	0.303	1.2	0.066	0.3	0.334	1.7	0.098	0.5	0.6
South London	-0.124	-0.7	0.266	1.4	0.447	2.9	0.400	2.7	-2.3
North London	0.727	3.4	0.537	2.5	0.508	2.8	0.059	0.3	2.3
East London	0.520	2.6	0.646	3.3	0.371	2.2	0.097	0.6	1.7
Dwelling characteristics									
Dwelling cost of owners (price × 0.0001)									
Household income less than £30,000	-0.325	-4.3	-0.527	-3.6	-0.802	-5.9	-1.275	-4.4	3.2
Household income between £30,000 to £60,000	-0.374	-3.6	-0.285	-2.5	-0.588	-4.7	-0.506	-3.0	0.7
Household income more than £60,000	-0.230	-3.2	-0.119	-2.9	-0.297	-5.0	-0.278	-3.2	0.4
Missing values	-0.034	-0.9	0.026	0.6	-0.020	-0.4	-0.041	-0.6	0.1
Dwelling cost of renters (monthly rent × 0.01)									
Household income less than £30,000	-0.243	-1.3	-0.184	-1.4	-0.198	-3.7	-0.225	-5.8	-0.1
Household income between £30,000 to £60,000	-0.204	-2.5	-0.156	-1.0	-0.181	-2.6	-0.166	-4.1	-0.4
Household income more than £60,000	-0.069	-1.1	-0.136	-0.5	-0.077	-1.9	-0.070	-3.3	0.0
Missing values	0.002	0.1	-0.038	-0.8	-0.089	-2.5	-0.037	-2.5	1.6
Dwelling type									
Detached house in inner London (owners)	-0.084	-3.1	-0.154	-3.9	-0.144	-4.2	-0.216	-3.1	1.8
Detached house in inner London (renters)	-0.308	-1.2	-0.050	-0.5	0.018	0.6	-0.050	-1.6	-1.0
Detached house in outer London (owners)	-0.036	-4.3	-0.026	-3.1	-0.025	-3.1	-0.023	-1.8	-0.8
Detached house in outer London (renters)	0.047	1.4	-0.075	-1.4	-0.020	-0.9	0.013	1.2	0.9
Flat in inner London	0.025	3.6	0.038	5.1	0.038	6.4	0.037	6.1	-1.4
Flat in outer London (owners)	-0.006	-4.0	-0.005	-1.0	-0.008	-1.7	-0.007	-1.2	0.2
Flat in outer London (renters)	-0.007	-0.2	-0.028	-1.6	-0.001	-0.1	0.012	1.8	-0.5
Location and land use characteristics									
Land use type									
Residential land area in inner London	0.178	5.9	0.138	4.3	0.178	8.4	0.175	8.0	0.1
Residential land area in outer London	0.268	7.5	0.238	6.9	0.236	7.5	0.219	6.6	1.0
Commercial land area in inner and outer london	-0.053	-2.7	-0.059	-2.7	-0.028	-1.9	-0.094	-6.0	1.6
Land use mix	1.134	2.0	0.871	1.3	1.084	2.0	2.940	4.2	-2.0
Ethnic composition									
Ratio of white people × white dummy	0.015	4.5	0.023	5.6	0.020	6.0	0.015	4.6	0.0
Ratio of asian people × asian dummy	0.044	8.2	0.037	6.3	0.033	5.5	0.039	5.4	0.6
Ratio of black people × black dummy	0.061	4.9	0.055	4.3	0.043	4.7	0.049	4.9	0.7
Dwelling density									
Inner London	-0.034	-3.0	-0.020	-1.8	-0.023	-3.1	-0.027	-4.1	-0.5
Outer London	-0.138	-8.3	-0.115	-7.5	-0.129	-9.0	-0.118	-8.6	-0.9
School quality	0.003	1.0	0.009	3.3	0.009	3.4	0.011	3.2	-1.7
Crime rate	-0.098	-1.0	-0.031	-0.3	-0.085	-1.2	-0.215	-3.3	1.0
Household size	-0.394	-2.5	-0.514	-3.5	-0.380	-2.6	-0.126	-0.8	-1.2

Table M1 Estimation of models for households moved in different time periods (TP) (cont.)

Parameters	Year household moved								t diff. TP1 and TP4
	TP1 (Before 90)		TP2 (1990-1996)		TP3 (1997-2000)		TP4 1 (2001-2002)		
	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat	
Employment opportunity	0.148	1.6	0.128	1.5	0.121	1.9	0.288	4.6	-1.3
Distance from CBD	0.050	3.4	0.061	3.9	0.038	2.8	0.028	1.9	1.1
Transport and travel characteristics									
Public transport accessibility									
Households own car (owners)	-0.061	-0.7	-0.318	-3.4	-0.295	-3.7	-0.125	-1.2	0.5
Households own car (renters)	0.449	1.4	0.176	0.7	0.000	0.0	0.018	0.2	1.3
Households do not own car	0.165	1.3	0.375	2.8	0.374	3.4	0.428	5.2	-1.7
Commute distance (random parameter)									
Owners									
Mean	-0.242	-11.4	-0.180	-18.0	-0.178	-19.6	-0.139	-12.1	-4.2
Standard deviation	0.142	2.7	0.017	2.1	0.023	2.1	0.020	2.1	2.3
Renters									
Mean	-0.481	-3.1	-0.254	-5.9	-0.246	-12.3	-0.204	-15.0	-1.8
Standard deviation	0.169	2.8	0.082	3.6	0.060	3.4	0.056	2.5	1.7
Measures of model fit									
Number of observations	645		532		726		647		
Initial LL	-4005.84		-3304.04		-4508.90		-4018.26		
Final LL	-2872.07		-2510.27		-3483.56		-3204.74		
Adjusted ρ2	0.273		0.228		0.218		0.192		

(a) Parameters where owners and renters have different sensitivities

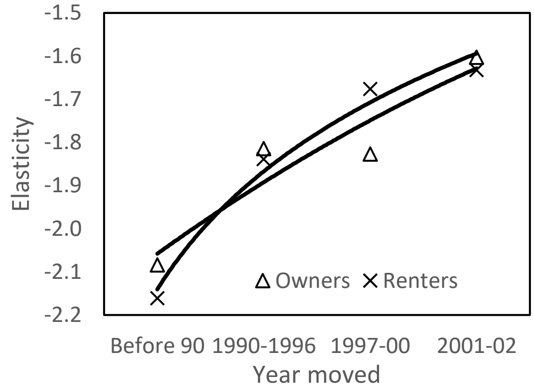
- The elasticity of the parameter commute distance is found to decrease between the Model TP 1 to TP4 but there is no significant level of difference between owners and renters.
- The housing cost elasticity of low-income owners is found to increase significantly between the Model TP 1 to TP4.
- The elasticity of the parameter percentage of detached houses in Inner London has changed between TP1 and TP4.

(b) Parameters where owners and renters have the same sensitivity

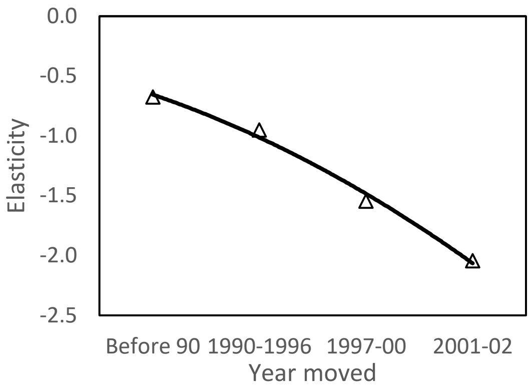
- The preference for the areas with better school quality for kids gives an increasing trend between the household moved in TP1 to TP4.
- The elasticity of the parameter mix land use pattern is found positive and has shown an increasing trend.
- The choices of households having no cars are found to become more elasticity to public transport accessibility over time.

Table M2 Elasticities of the parameters in the models for the households moved in different time periods (TP)

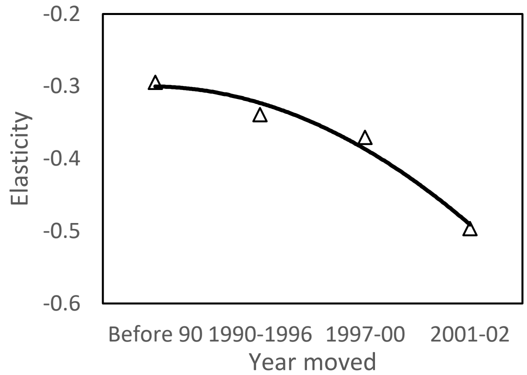
Parameters	Year household moved			
	TP1 (Before 90)	TP2 (1990-1996)	TP3 (1997-2000)	TP4 (2001-2002)
Dwelling characteristics				
Dwelling cost of owners (price × 0.0001)				
Household income less than £30,000	-0.676	-0.951	-1.546	-2.044
Household income between £30,000 to £60,000	-0.783	-0.604	-1.191	-0.934
Household income more than £60,000	-0.636	-0.355	-0.799	-0.717
Dwelling cost of renters (monthly rent × 0.01)				
Household income less than £30,000	-1.955	-1.131	-1.141	-1.438
Household income between £30,000 to £60,000	-1.661	-1.125	-1.195	-1.158
Household income more than £60,000	-0.705	-0.912	-0.618	-0.623
Dwelling type				
Detached house in inner London (owners)	-0.295	-0.339	-0.370	-0.496
Detached house in inner London (renters)	-0.502	-0.194	0.046	-0.125
Detached house in outer London (owners)	-0.279	-0.220	-0.202	-0.175
Detached house in outer London (renters)	0.377	-0.473	-0.174	0.102
Flat in inner London	1.604	2.560	2.679	2.675
Flat in outer London (owners)	-0.269	-0.180	-0.287	-0.262
Flat in outer London (renters)	-0.286	-0.999	-0.039	0.517
Location and land use characteristics				
Land use type				
Residential land area in inner London	2.752	2.086	2.867	2.766
Residential land area in outer London	2.717	2.468	2.337	2.310
Commercial land area in inner and outer london	-0.293	-0.326	-0.181	-0.670
Land use mix	0.711	0.697	0.889	2.449
Ethnic composition				
Ratio of white people × white dummy	1.145	1.778	1.508	1.111
Ratio of asian people × asian dummy	1.388	1.025	0.869	0.997
Ratio of black people × black dummy	1.139	0.859	0.724	0.743
Dwelling density				
Inner London	-1.513	-0.930	-1.172	-1.365
Outer London	-2.819	-2.417	-2.605	-2.615
School quality	0.873	2.660	2.632	3.232
Crime rate	-0.122	-0.039	-0.115	-0.326
Household size	-0.476	-0.683	-0.408	-0.130
Employment opportunity	0.075	0.062	0.070	0.214
Distance from CBD	0.875	1.048	0.613	0.413
Transport and travel characteristics				
Public transport accessibility				
Households own car (owners)	-0.198	-1.034	-0.980	-0.416
Households own car (renters)	1.786	0.552	0.010	0.069
Households do not own car	0.652	1.458	1.697	1.932
Commute distance (random parameter)				
Owners				
Mean	-2.084	-1.814	-1.827	-1.602
Standard deviation	0.775	0.137	0.187	0.183
Renters				
Mean	-2.161	-1.839	-1.676	-1.632
Standard deviation	0.589	0.464	0.323	0.358



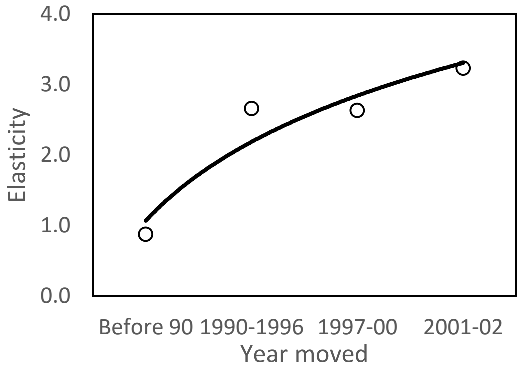
a. Commute distance



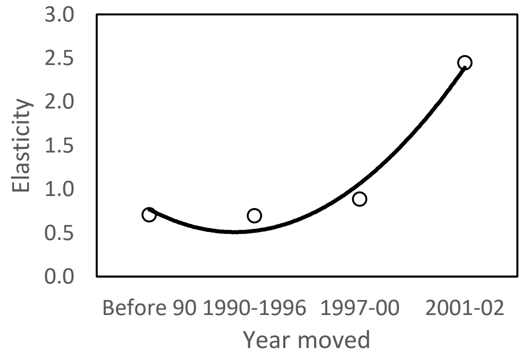
b. Housing cost for owners (income<£30,000)



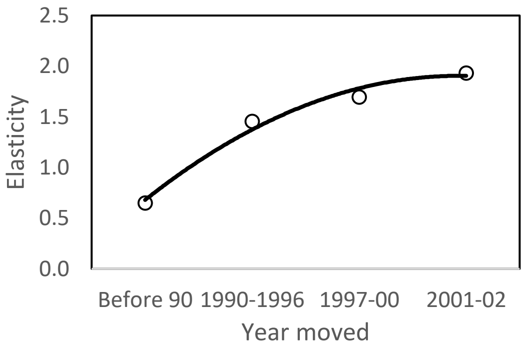
c. Detached house in inner London (owners)



d. School quality



e. Land use mix



f. PT accessibility (non-car owners)

Figure M1 Elasticity changes of parameters between TP1 and TP4

Appendix N Models estimated using subset of data for validation

N1 Estimated models for the subsets of data for owners

Table N1 Residential location choice parameters for estimation subset 1 of owner's dataset.

Parameters	CMNL		rCMNL		ICMNL	
	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat
Constants						
Central London	0.334	2.2	0.353	2.3	0.406	2.6
South London	0.375	3.2	0.376	3.2	0.399	3.1
North London	0.608	4.4	0.624	4.7	0.659	4.4
East London	0.728	5.4	0.731	5.3	0.839	5.6
Dwelling characteristics						
Dwelling cost (price* 0.0001)						
Income less than £30,000	-0.493	-5.7	-0.494	-5.7	-0.526	-5.7
Income between £30,000 to £60,000	-0.479	-5.8	-0.469	-5.6	-0.513	-5.9
Income more than £60,000	-0.236	-4.0	-0.237	-4.1	-0.248	-4.0
Missing values	-0.028	-0.8	-0.026	-0.7	-0.037	-1.0
Dwelling type						
Detached house in inner London	-0.126	-4.3	-0.127	-4.5	-0.120	-4.0
Detached house in outer London	-0.025	-4.8	-0.025	-4.9	-0.027	-5.1
Flat in inner London	0.026	5.2	0.026	5.4	0.025	5.1
Flat in outer London	-0.011	-3.4	-0.013	-3.5	-0.010	-3.2
Location and land use characteristics						
Land use type						
Residential land area in inner London	0.140	7.8	0.140	7.8	0.130	7.2
Residential land area in outer London	0.212	8.4	0.224	8.4	0.211	8.4
Commercial land area in inner and outer London	-0.061	-4.8	-0.060	-4.7	-0.056	-4.4
Land use mix	1.365	3.5	1.349	3.6	1.151	3.0
Ethnic composition						
Ratio of white people × white dummy	0.018	7.1	0.017	7.1	0.019	7.2
Ratio of asian people × asian dummy	0.030	6.5	0.028	6.6	0.028	5.8
Ratio of Black people × black dummy	0.051	5.7	0.050	5.7	0.048	5.2
Dwelling density						
Inner London	-0.020	-3.3	-0.020	-3.3	-0.019	-3.0
Outer London	-0.110	-9.3	-0.120	-9.3	-0.110	-9.3
School quality	0.007	3.6	0.007	3.7	0.006	3.2
Crime rate	-0.111	-1.8	-0.110	-1.8	-0.112	-1.8
Household size	-0.329	-3.0	-0.329	-2.9	-0.326	-2.9
Employment opportunity	0.193	3.5	0.194	3.6	0.202	3.6
Distance from CBD	0.083	8.5	0.082	8.6	0.101	9.8
Transport and travel characteristics						
Public transport accessibility						
Households own car	-0.189	-3.4	-0.188	-3.3	-0.202	-3.6
Households do not own car	0.132	1.6	0.133	1.5	0.109	1.3
Commute distance	-0.147	-25.6	-0.146	-25.7	-0.152	-23.8
Penalty parameter (μ)						
Distance from past home	0.198	34.3	0.198	34.5	0.019	30.9
Measures of model fit						
Number of observations	1500		1500		1500	
Initial LL	-9315.9001		-9315.9001		-9315.9001	
Final LL	-6150.5050		-6149.0500		-6033.8810	
Adjusted ρ2	0.337		0.337		0.349	

Table N2 Residential location choice parameters for estimation subset 2 of owner’s dataset.

Parameters	CMNL		rCMNL		ICMNL	
	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat
Constants						
Central London	0.389	2.6	0.391	2.7	0.469	3.0
South London	0.297	2.5	0.298	2.5	0.325	2.5
North London	0.759	5.7	0.762	5.7	0.835	5.6
East London	0.807	6.1	0.808	6.2	0.942	6.3
Dwelling characteristics						
Dwelling cost (price* 0.0001)						
Income less than £30,000	-0.506	-5.8	-0.507	-5.8	-0.545	-6.1
Income between £30,000 to £60,000	-0.487	-5.8	-0.488	-5.8	-0.517	-5.9
Income more than £60,000	-0.213	-3.9	-0.214	-3.9	-0.224	-4.0
Missing values	-0.072	-1.8	-0.073	-1.8	-0.085	-2.0
Dwelling type						
Detached house in inner London	-0.126	-4.3	-0.125	-4.3	-0.120	-4.0
Detached house in outer London	-0.027	-5.2	-0.028	-5.3	-0.029	-5.5
Flat in inner London	0.027	5.4	0.027	5.4	0.027	5.3
Flat in outer London	-0.014	-4.5	-0.014	-4.5	-0.014	-4.4
Location and land use characteristics						
Land use type						
Residential land area in inner London	0.133	7.5	0.134	7.5	0.124	6.9
Residential land area in outer London	0.210	8.4	0.210	8.4	0.208	8.3
Commercial land area in inner and outer London	-0.056	-4.5	-0.055	-4.4	-0.051	-4.1
Land use mix	0.998	2.6	0.999	2.6	0.808	2.3
Ethnic composition						
Ratio of white people × white dummy	0.020	7.7	0.020	7.7	0.020	7.8
Ratio of asian people × asian dummy	0.027	5.9	0.028	6.0	0.025	5.2
Ratio of Black people × black dummy	0.047	5.0	0.047	5.0	0.044	4.6
Dwelling density						
Inner London	-0.021	-3.5	-0.022	-3.5	-0.020	-3.3
Outer London	-0.110	-9.4	-0.111	-9.5	-0.110	-9.4
School quality	0.009	4.9	0.009	4.9	0.009	4.6
Crime rate	-0.047	-0.8	-0.049	-0.6	-0.042	-0.7
Household size	-0.297	-2.7	-0.296	-2.7	-0.292	-2.6
Employment opportunity	0.116	2.1	0.116	2.1	0.117	2.1
Distance from CBD	0.084	8.7	0.085	8.8	0.100	9.8
Transport and travel characteristics						
Public transport accessibility						
Households own car	-0.142	-2.5	-0.144	-2.5	-0.156	-2.8
Households do not own car	0.234	2.9	0.233	2.9	0.211	2.6
Commute distance	-0.146	-25.7	-0.146	-25.6	-0.151	-24.0
Penalty parameter (μ)						
Distance from past home	0.196	33.3	0.196	33.2	0.019	30.6
Measures of model fit						
Number of observations	1500		1500		1500	
Initial LL	-9315.9001		-9315.9001		-9315.9001	
Final LL	-6175.5150		-6175.9400		-6072.8420	
Adjusted ρ2	0.334		0.334		0.345	

Table N3 Residential location choice parameters for estimation subset 3 of owner's dataset.

Parameters	CMNL		rCMNL		lCMNL	
	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat
Constants						
Central London	0.374	2.5	0.377	2.6	0.464	3.0
South London	0.429	3.6	0.433	3.6	0.489	3.9
North London	0.740	5.5	0.750	5.5	0.801	5.6
East London	0.839	6.2	0.838	6.2	0.970	6.7
Dwelling characteristics						
Dwelling cost (price* 0.0001)						
Income less than £30,000	-0.481	-5.6	-0.480	-5.5	-0.513	-5.7
Income between £30,000 to £60,000	-0.426	-5.4	-0.428	-5.4	-0.463	-5.6
Income more than £60,000	-0.224	-3.8	-0.226	-3.9	-0.236	-3.8
Missing values	-0.028	-0.8	-0.028	-0.8	-0.041	-1.1
Dwelling type						
Detached house in inner London	-0.124	-4.2	-0.123	-4.2	-0.117	-3.9
Detached house in outer London	-0.030	-5.6	-0.031	-5.6	-0.032	-5.9
Flat in inner London	0.027	5.4	0.027	5.5	0.027	5.2
Flat in outer London	-0.011	-3.6	-0.011	-3.6	-0.011	-3.5
Location and land use characteristics						
Land use type						
Residential land area in inner London	0.151	8.3	0.150	8.3	0.142	7.7
Residential land area in outer London	0.196	7.8	0.197	7.8	0.194	7.7
Commercial land area in inner and outer London	-0.075	-5.9	-0.074	-5.8	-0.070	-5.3
Land use mix	1.586	4.3	1.585	4.2	1.371	3.3
Ethnic composition						
Ratio of white people × white dummy	0.016	6.4	0.016	6.4	0.017	6.5
Ratio of asian people × asian dummy	0.033	7.1	0.032	7.1	0.031	6.3
Ratio of Black people × black dummy	0.052	5.8	0.051	5.8	0.049	5.3
Dwelling density						
Inner London	-0.021	-3.3	-0.021	-3.4	-0.020	-3.1
Outer London	-0.108	-9.1	-0.107	-9.1	-0.107	-9.1
School quality	0.008	4.6	0.008	4.6	0.008	4.3
Crime rate	-0.058	-0.9	-0.058	-0.9	-0.054	-0.8
Household size	-0.193	-1.8	-0.192	-1.8	-0.185	-1.6
Employment opportunity	0.159	2.9	0.160	2.9	0.163	2.9
Distance from CBD	0.089	9.1	0.090	9.2	0.106	10.3
Transport and travel characteristics						
Public transport accessibility						
Households own car	-0.146	-2.6	-0.146	-2.6	-0.159	-2.8
Households do not own car	0.117	1.4	0.116	1.4	0.084	1.0
Commute distance	-0.148	-26.2	-0.149	-26.2	-0.155	-24.5
Penalty parameter (μ)						
Distance from past home	0.197	33.8	0.196	33.6	0.020	30.8
Measures of model fit						
Number of observations	1500		1500		1500	
Initial LL	-9315.9001		-9315.9001		-9315.9001	
Final LL	-6150.3080		-6151.0300		-6037.0380	
Adjusted ρ2	0.337		0.337		0.349	

Table N4 Residential location choice parameters for estimation subset 4 of owner's dataset.

Parameters	CMNL		rCMNL		ICMNL	
	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat
Constants						
Central London	0.493	3.1	0.494	3.3	0.572	3.7
South London	0.402	3.1	0.405	3.2	0.439	3.4
North London	0.686	4.6	0.687	4.6	0.743	5.0
East London	0.747	5.0	0.748	5.0	0.875	6.0
Dwelling characteristics						
Dwelling cost (price* 0.0001)						
Income less than £30,000	-0.571	-6.2	-0.572	-6.3	-0.611	-6.4
Income between £30,000 to £60,000	-0.442	-5.5	-0.443	-5.5	-0.476	-5.6
Income more than £60,000	-0.213	-3.7	-0.213	-3.7	-0.219	-3.7
Missing values	-0.027	-0.7	-0.028	-0.8	-0.039	-1.0
Dwelling type						
Detached house in inner London	-0.131	-4.5	-0.130	-4.5	-0.125	-4.3
Detached house in outer London	-0.029	-5.3	-0.029	-5.3	-0.031	-5.6
Flat in inner London	0.023	4.7	0.023	4.7	0.022	4.5
Flat in outer London	-0.011	-3.3	-0.011	-3.3	-0.010	-3.2
Location and land use characteristics						
Land use type						
Residential land area in inner London	0.135	7.6	0.134	7.6	0.126	7.0
Residential land area in outer London	0.202	8.1	0.202	8.1	0.199	8.0
Commercial land area in inner and outer London	-0.067	-4.9	-0.067	-4.9	-0.062	-4.6
Land use mix	1.210	2.8	1.210	2.8	1.014	2.4
Ethnic composition						
Ratio of white people × white dummy	0.018	7.0	0.018	7.0	0.018	7.1
Ratio of asian people × asian dummy	0.033	7.0	0.033	6.9	0.032	6.4
Ratio of Black people × black dummy	0.050	5.7	0.050	5.7	0.047	5.2
Dwelling density						
Inner London	-0.018	-3.0	-0.018	-3.0	-0.017	-2.7
Outer London	-0.105	-9.0	-0.105	-9.0	-0.104	-8.9
School quality	0.007	3.7	0.007	3.7	0.007	3.5
Crime rate	-0.057	-0.8	-0.056	-0.8	-0.051	-0.7
Household size	-0.268	-2.4	-0.267	-2.4	-0.262	-2.3
Employment opportunity	0.107	1.8	0.107	1.8	0.109	1.8
Distance from CBD	0.080	8.0	0.082	8.2	0.097	9.4
Transport and travel characteristics						
Public transport accessibility						
Households own car	-0.210	-3.7	-0.211	-3.7	-0.226	-4.0
Households do not own car	0.111	1.3	0.111	1.3	0.092	1.1
Commute distance	-0.144	-25.5	-0.145	-25.4	-0.150	-23.6
Penalty parameter (μ)						
Distance from past home	0.198	33.9	0.197	33.8	0.021	30.8
Measures of model fit						
Number of observations	1500		1500		1500	
Initial LL	-9315.9001		-9315.9001		-9315.9001	
Final LL	-6175.3840		-6174.8100		-6056.9180	
Adjusted ρ2	0.334		0.334		0.347	

Table N5 Residential location choice parameters for estimation subset 5 of owner's dataset.

Parameters	CMNL		rCMNL		ICMNL	
	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat
Constants						
Central London	0.479	3.1	0.480	3.2	0.533	3.2
South London	0.274	2.2	0.274	2.2	0.273	1.9
North London	0.653	4.5	0.655	4.5	0.691	4.3
East London	0.726	5.1	0.726	5.1	0.834	5.2
Dwelling characteristics						
Dwelling cost (price* 0.0001)						
Income less than £30,000	-0.519	-6.0	-0.520	-6.0	-0.555	-6.2
Income between £30,000 to £60,000	-0.580	-6.3	-0.583	-6.4	-0.620	-6.4
Income more than £60,000	-0.223	-3.9	-0.224	-3.9	-0.237	-4.0
Missing values	-0.044	-1.1	-0.045	-1.1	-0.057	-1.4
Dwelling type						
Detached house in inner London	-0.117	-4.1	-0.116	-4.0	-0.110	-3.8
Detached house in outer London	-0.026	-4.9	-0.026	-4.9	-0.028	-5.1
Flat in inner London	0.026	5.3	0.026	5.3	0.026	5.1
Flat in outer London	-0.012	-3.6	-0.012	-3.5	-0.011	-3.5
Location and land use characteristics						
Land use type						
Residential land area in inner London	0.151	8.3	0.150	8.2	0.141	7.7
Residential land area in outer London	0.242	9.7	0.241	9.6	0.240	9.6
Commercial land area in inner and outer London	-0.060	-4.6	-0.060	-4.6	-0.054	-4.2
Land use mix	1.277	3.0	1.275	3.0	1.065	2.6
Ethnic composition						
Ratio of white people × white dummy	0.017	6.6	0.017	6.6	0.017	6.7
Ratio of asian people × asian dummy	0.029	6.3	0.029	6.3	0.027	5.5
Ratio of Black people × black dummy	0.046	5.3	0.046	5.3	0.043	4.8
Dwelling density						
Inner London	-0.024	-3.8	-0.024	-3.8	-0.023	-3.6
Outer London	-0.120	-10.2	-0.120	-10.2	-0.120	-10.2
School quality	0.006	3.4	0.006	3.4	0.006	3.1
Crime rate	-0.092	-1.4	-0.091	-1.4	-0.089	-1.4
Household size	-0.318	-2.9	-0.318	-2.9	-0.310	-2.8
Employment opportunity	0.166	2.9	0.167	2.9	0.172	3.0
Distance from CBD	0.075	7.4	0.076	7.5	0.092	8.7
Transport and travel characteristics						
Public transport accessibility						
Households own car	-0.230	-4.1	-0.231	-4.1	-0.244	-4.3
Households do not own car	0.191	2.4	0.190	2.4	0.164	2.0
Commute distance	-0.147	-25.3	-0.147	-25.3	-0.151	-23.4
Penalty parameter (μ)						
Distance from past home	0.204	33.8	0.203	33.7	0.018	30.8
Measures of model fit						
Number of observations	1500		1500		1500	
Initial LL	-9315.9001		-9315.9001		-9315.9001	
Final LL	-6129.6750		-6129.4300		-6018.5830	
Adjusted ρ2	0.339		0.339		0.351	

N2 Estimated models for the subsets of data for renters

Table N6 Residential location choice parameters for estimation subset 1 of renter's dataset.

Parameters	CMNL		rCMNL		ICMNL	
	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat
Constants						
Central London	0.7376	2.6	0.7370	2.6	0.7904	2.7
South London	0.5827	2.2	0.5835	2.2	0.6234	2.3
North London	0.8328	2.5	0.8339	2.5	0.8907	2.6
East London	1.2668	4.2	1.2692	4.2	1.3519	4.4
Dwelling characteristics						
Dwelling cost (monthly rent* 0.01)						
Income less than £30,000	-0.2816	-4.9	-0.2814	-4.9	-0.2891	-5.0
Income between £30,000 to £60,000	-0.1048	-1.8	-0.1046	-1.8	-0.1035	-1.8
Income more than £60,000	-0.0931	-2.6	-0.0935	-2.6	-0.0968	-2.7
Missing values	-0.0967	-2.8	-0.0975	-2.8	-0.1016	-2.9
Dwelling type						
Detached house in inner London	-0.0421	-0.9	-0.0421	-0.9	-0.0420	-0.9
Detached house in outer London	-0.0279	-1.6	-0.0281	-1.6	-0.0288	-1.6
Flat in inner London	0.0353	3.8	0.0353	3.8	0.0355	3.8
Flat in outer London	0.0051	0.6	0.0052	0.6	0.0049	0.6
Location and land use characteristics						
Land use type						
Residential land area in inner London	0.1637	5.5	0.1630	5.5	0.1600	5.3
Residential land area in outer London	0.2477	4.2	0.2473	4.2	0.2459	4.1
Commercial land area in inner and outer London	-0.0691	-3.3	-0.0682	-3.3	-0.0663	-3.1
Land use mix	2.8181	2.9	2.8160	2.9	2.6396	2.7
Ethnic composition						
Ratio of white people × white dummy	0.0205	3.4	0.0206	3.4	0.0208	3.4
Ratio of asian people × asian dummy	0.0316	3.0	0.0316	3.0	0.0313	2.9
Ratio of Black people × black dummy	0.0314	2.0	0.0304	2.0	0.0287	1.8
Dwelling density						
Inner London	-0.0162	-1.9	-0.0162	-1.9	-0.0161	-1.9
Outer London	-0.1250	-4.7	-0.1249	-4.7	-0.1245	-4.7
School quality	0.0087	1.5	0.0086	1.5	0.0079	1.4
Crime rate	-0.3025	-2.9	-0.3027	-2.9	-0.3016	-2.9
Household size	-0.0717	-0.3	-0.0767	-0.1	-0.3000	-0.1
Employment opportunity	0.3659	3.8	0.3660	3.8	0.3678	3.8
Distance from CBD	0.1010	4.5	0.1018	4.5	0.1075	4.7
Transport and travel characteristics						
Public transport accessibility						
Households own car	0.2194	1.9	0.2171	1.9	0.2141	1.8
Households do not own car	0.3534	3.1	0.3536	3.1	0.3438	3.0
Commute distance	-0.1758	-13.6	-0.1762	-13.6	-0.1795	-13.3
Penalty parameter (μ)						
Distance from past home	0.1668	13.0	0.1669	13.0	0.0174	13.1
Measures of model fit						
Number of observations	305		305		305	
Initial LL	-1894.233		-1894.233		-1894.233	
Final LL	-1345.8530		-1345.9800		-1336.1580	
Adjusted ρ2	0.274		0.274		0.279	

Table N7 Residential location choice parameters for estimation subset 2 of renter's dataset.

Parameters	CMNL		rCMNL		ICMNL	
	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat
Constants						
Central London	0.6479	2.4	0.6495	2.4	0.7029	2.5
South London	0.3048	1.2	0.3072	1.2	0.3227	1.2
North London	0.5004	1.6	0.5010	1.6	0.5501	1.7
East London	0.9161	3.2	0.9172	3.2	1.0181	3.5
Dwelling characteristics						
Dwelling cost (monthly rent* 0.01)						
Income less than £30,000	-0.2614	-4.7	-0.2618	-4.7	-0.2720	-4.8
Income between £30,000 to £60,000	-0.1611	-2.5	-0.1604	-2.5	-0.1561	-2.4
Income more than £60,000	-0.1038	-2.6	-0.1048	-2.6	-0.1081	-2.6
Missing values	-0.0816	-2.8	-0.0817	-2.8	-0.0855	-2.9
Dwelling type						
Detached house in inner London	-0.0196	-0.4	-0.0196	-0.4	-0.0230	-0.5
Detached house in outer London	-0.0082	-0.5	-0.0083	-0.5	-0.0086	-0.6
Flat in inner London	0.0409	4.3	0.0408	4.3	0.0415	4.3
Flat in outer London	0.0079	1.0	0.0079	1.0	0.0075	1.0
Location and land use characteristics						
Land use type						
Residential land area in inner London	0.1857	5.9	0.1853	5.9	0.1829	5.7
Residential land area in outer London	0.2537	4.4	0.2525	4.4	0.2466	4.2
Commercial land area in inner and outer London	-0.0826	-3.9	-0.0824	-3.9	-0.0800	-3.7
Land use mix	3.0890	3.4	3.0898	3.4	2.9014	2.9
Ethnic composition						
Ratio of white people × white dummy	0.0176	3.0	0.0179	3.0	0.0180	3.0
Ratio of asian people × asian dummy	0.0422	3.7	0.0426	3.7	0.0433	3.7
Ratio of Black people × black dummy	0.0472	2.8	0.0475	2.8	0.0455	2.6
Dwelling density						
Inner London	-0.0281	-3.1	-0.0281	-3.1	-0.0282	-3.1
Outer London	-0.1267	-4.8	-0.1264	-4.8	-0.1245	-4.8
School quality	0.0077	1.4	0.0075	1.4	0.0067	1.2
Crime rate	-0.3271	-3.2	-0.3279	-3.2	-0.3304	-3.2
Household size	-0.0004	0.1	-0.0006	0.1	0.0141	0.1
Employment opportunity	0.4106	4.3	0.4115	4.3	0.4174	4.3
Distance from CBD	0.0899	4.2	0.0895	4.2	0.1017	4.6
Transport and travel characteristics						
Public transport accessibility						
Households own car	0.1783	1.5	0.1780	1.5	0.1680	1.4
Households do not own car	0.4171	3.7	0.4164	3.7	0.4082	3.6
Commute distance	-0.1803	-14.5	-0.1812	-14.5	-0.1903	-14.3
Penalty parameter (μ)						
Distance from past home	0.1692	12.6	0.1693	12.6	0.0250	13.0
Measures of model fit						
Number of observations	305		305		305	
Initial LL	-1894.233		-1894.233		-1894.233	
Final LL	-1326.3080		-1326.1500		-1311.8180	
Adjusted ρ2	0.284		0.284		0.292	

Table N8 Residential location choice parameters for estimation subset 3 of renter's dataset.

Parameters	CMNL		rCMNL		ICMNL	
	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat
Constants						
Central London	0.4314	1.6	0.4322	1.6	0.4523	1.6
South London	0.2846	1.1	0.2849	1.1	0.2858	1.0
North London	0.3968	1.3	0.3978	1.3	0.4178	1.3
East London	0.7659	2.7	0.7664	2.7	0.8097	2.7
Dwelling characteristics						
Dwelling cost (monthly rent* 0.01)						
Income less than £30,000	-0.2550	-4.6	-0.2553	-4.6	-0.2631	-4.6
Income between £30,000 to £60,000	-0.1340	-2.2	-0.1335	-2.2	-0.1322	-2.1
Income more than £60,000	-0.0825	-2.3	-0.0827	-2.3	-0.0843	-2.3
Missing values	-0.0761	-2.3	-0.0760	-2.3	-0.0773	-2.3
Dwelling type						
Detached house in inner London	-0.0269	-0.6	-0.0272	-0.6	-0.0290	-0.6
Detached house in outer London	-0.0049	-0.3	-0.0049	-0.3	-0.0051	-0.3
Flat in inner London	0.0323	3.4	0.0323	3.4	0.0324	3.4
Flat in outer London	0.0089	1.1	0.0089	1.1	0.0088	1.1
Location and land use characteristics						
Land use type						
Residential land area in inner London	0.1563	5.1	0.1560	5.1	0.1528	4.9
Residential land area in outer London	0.2395	4.1	0.2391	4.1	0.2342	4.0
Commercial land area in inner and outer London	-0.0720	-3.3	-0.0718	-3.3	-0.0706	-3.2
Land use mix	2.4761	2.7	2.4740	2.7	2.3667	2.3
Ethnic composition						
Ratio of white people × white dummy	0.0178	3.0	0.0178	3.0	0.0179	3.0
Ratio of asian people × asian dummy	0.0463	4.4	0.0463	4.4	0.0473	4.4
Ratio of Black people × black dummy	0.0417	2.6	0.0417	2.6	0.0399	2.5
Dwelling density						
Inner London	-0.0163	-1.8	-0.0163	-1.8	-0.0159	-1.8
Outer London	-0.1319	-5.0	-0.1317	-5.0	-0.1302	-5.0
School quality	0.0055	0.9	0.0054	0.9	0.0049	0.8
Crime rate	-0.1890	-1.8	-0.1898	-1.8	-0.1916	-1.8
Household size	-0.2190	-0.9	-0.2184	-0.9	-0.2163	-0.9
Employment opportunity	0.2864	2.9	0.2866	2.9	0.2924	3.0
Distance from CBD	0.0631	2.9	0.0637	3.0	0.0710	3.2
Transport and travel characteristics						
Public transport accessibility						
Households own car	-0.0155	-0.1	-0.0159	-0.1	-0.0213	-0.2
Households do not own car	0.3003	2.6	0.3000	2.6	0.2956	2.6
Commute distance	-0.1917	-15.0	-0.1919	-15.0	-0.1969	-15.0
Penalty parameter (μ)						
Distance from past home	0.1587	12.1	0.1586	12.2	0.0232	12.8
Measures of model fit						
Number of observations	306		306		306	
Initial LL	-1900.4436		-1900.4436		-1900.4436	
Final LL	-1340.4150		-1340.1800		-1333.1460	
Adjusted ρ2	0.279		0.279		0.283	

Table N9 Residential location choice parameters for estimation subset 4 of renter's dataset.

Parameters	CMNL		rCMNL		ICMNL	
	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat
Constants						
Central London	0.5941	2.1	0.5947	2.1	0.6304	2.2
South London	0.4085	1.5	0.4093	1.5	0.4202	1.5
North London	0.4144	1.3	0.4149	1.3	0.4399	1.3
East London	0.8287	2.8	0.8286	2.8	0.8858	2.8
Dwelling characteristics						
Dwelling cost (monthly rent* 0.01)						
Income less than £30,000	-0.2285	-4.5	-0.2294	-4.5	-0.2371	-4.5
Income between £30,000 to £60,000	-0.1386	-2.3	-0.1384	-2.3	-0.1361	-2.2
Income more than £60,000	-0.0968	-2.3	-0.0969	-2.3	-0.1022	-2.4
Missing values	-0.0739	-2.5	-0.0743	-2.5	-0.0770	-2.5
Dwelling type						
Detached house in inner London	-0.0007	-0.1	-0.0006	-0.1	-0.0023	-0.1
Detached house in outer London	-0.0048	-0.3	-0.0048	-0.3	-0.0049	-0.3
Flat in inner London	0.0339	3.6	0.0340	3.6	0.0343	3.6
Flat in outer London	0.0036	0.5	0.0036	0.5	0.0036	0.5
Location and land use characteristics						
Land use type						
Residential land area in inner London	0.1490	4.9	0.1488	4.9	0.1459	4.8
Residential land area in outer London	0.2058	3.6	0.2050	3.6	0.2008	3.5
Commercial land area in inner and outer London	-0.0664	-3.3	-0.0663	-3.3	-0.0646	-3.2
Land use mix	2.2152	2.4	2.2147	2.4	2.0948	2.3
Ethnic composition						
Ratio of white people × white dummy	0.0190	3.1	0.0190	3.1	0.0193	3.2
Ratio of asian people × asian dummy	0.0436	4.5	0.0437	4.5	0.0443	4.5
Ratio of Black people × black dummy	0.0365	2.3	0.0363	2.3	0.0349	2.2
Dwelling density						
Inner London	-0.0210	-2.4	-0.0210	-2.4	-0.0210	-2.4
Outer London	-0.1215	-4.7	-0.1214	-4.7	-0.1202	-4.7
School quality	0.0081	1.5	0.0081	1.5	0.0075	1.3
Crime rate	-0.2160	-2.3	-0.2163	-2.3	-0.2186	-2.3
Household size	-0.0902	-0.4	-0.0897	-0.4	-0.0850	-0.4
Employment opportunity	0.2968	3.3	0.2966	3.3	0.3023	3.4
Distance from CBD	0.0682	3.1	0.0680	3.1	0.0767	3.4
Transport and travel characteristics						
Public transport accessibility						
Households own car	0.1709	1.5	0.1706	1.5	0.1639	1.4
Households do not own car	0.3818	3.4	0.3813	3.4	0.3735	3.3
Commute distance	-0.1868	-15.0	-0.1875	-15.0	-0.1935	-15.0
Penalty parameter (μ)						
Distance from past home	0.1598	12.7	0.1597	12.8	0.0221	13.2
Measures of model fit						
Number of observations	306		306		306	
Initial LL	-1900.4436		-1900.4436		-1900.4436	
Final LL	-1334.8470		-1334.7500		-1325.6790	
Adjusted ρ2	0.282		0.282		0.287	

Table N10 Residential location choice parameters for estimation subset 5 of renter's dataset.

Parameters	CMNL		rCMNL		ICMNL	
	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat
Constants						
Central London	0.2589	0.9	0.2601	0.9	0.2900	1.0
South London	0.3921	1.4	0.3929	1.4	0.4129	1.5
North London	0.3765	1.1	0.3773	1.1	0.4214	1.3
East London	0.6786	2.1	0.6792	2.1	0.7329	2.5
Dwelling characteristics						
Dwelling cost (monthly rent* 0.01)						
Income less than £30,000	-0.2661	-4.7	-0.2666	-4.7	-0.2724	-4.7
Income between £30,000 to £60,000	-0.1394	-2.2	-0.1390	-2.2	-0.1364	-2.2
Income more than £60,000	-0.1038	-2.5	-0.1039	-2.5	-0.1076	-2.5
Missing values	-0.0703	-2.2	-0.0703	-2.2	-0.0704	-2.2
Dwelling type						
Detached house in inner London	-0.0667	-1.3	-0.0670	-1.3	-0.0695	-1.4
Detached house in outer London	-0.0058	-0.4	-0.0059	-0.4	-0.0065	-0.4
Flat in inner London	0.0042	0.5	0.0042	0.5	0.0039	0.5
Flat in outer London	0.0253	2.7	0.0253	2.7	0.0255	2.7
Location and land use characteristics						
Land use type						
Residential land area in inner London	0.1472	4.8	0.1470	4.8	0.1443	4.8
Residential land area in outer London	0.1981	3.6	0.1978	3.6	0.1932	3.5
Commercial land area in inner and outer London	-0.0688	-3.3	-0.0689	-3.3	-0.0674	-3.2
Land use mix	2.3235	2.5	2.3227	2.5	2.2161	2.5
Ethnic composition						
Ratio of white people × white dummy	0.0174	2.9	0.0174	2.9	0.0176	2.9
Ratio of asian people × asian dummy	0.0466	4.6	0.0467	4.6	0.0475	4.6
Ratio of Black people × black dummy	0.0402	2.7	0.0399	2.7	0.0386	2.5
Dwelling density						
Inner London	-0.0168	-1.9	-0.0168	-1.9	-0.0166	-1.8
Outer London	-0.1092	-4.4	-0.1089	-4.4	-0.1075	-4.4
School quality	0.0096	1.6	0.0095	1.6	0.0090	1.5
Crime rate	-0.2429	-2.4	-0.2430	-2.4	-0.2442	-2.4
Household size	0.0109	0.1	0.0111	0.1	0.0124	0.1
Employment opportunity	0.3317	3.6	0.3319	3.6	0.3361	3.6
Distance from CBD	0.0639	2.8	0.0642	2.9	0.0719	3.2
Transport and travel characteristics						
Public transport accessibility						
Households own car	0.1128	1.0	0.1125	1.0	0.1085	0.9
Households do not own car	0.3931	3.4	0.3930	3.4	0.3862	3.4
Commute distance	-0.1899	-15.1	-0.1901	-15.1	-0.1952	-15.0
Penalty parameter (μ)						
Distance from past home	0.1596	12.2	0.1595	12.3	0.0208	12.7
Measures of model fit						
Number of observations	306		306		306	
Initial LL	-1900.4436		-1900.4436		-1900.4436	
Final LL	-1337.7990		-1337.9100		-1330.6520	
Adjusted ρ2	0.280		0.280		0.284	