Quantification of the influences of built-form upon travel of employed adults: new models based on the UK National Travel Survey

Kaveh Jahanshahi

Department of Architecture

University of Cambridge

This dissertation is submitted for the degree of
Doctor of Philosophy

Clare Hall

May 2017

I would like to dedicate this thesis to my loving parents and beloved brother

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements. This dissertation contains fewer than 80,000 words including appendices, bibliography, footnotes, tables and equations and has fewer than 150 figures.

Acknowledgment

I would like to express my sincere gratitude to my supervisor, Dr Ying Jin for his continuous support, patience, motivation, and immense knowledge. His guidance, advice, and encouragement has been priceless throughout this research and for writing of this thesis. I could not have imagined having a better mentor for my PhD study. I also wish to express my appreciation to my advisor, Ian Williams, for his unfailing advice and comments and continuous encouragement and support. In addition to learning from Ian as my advisor over this PhD course; I have benefited from his theoretical, technical and practical knowledge and expertise over many years I was working with him.

I also wish to acknowledge the support of EPSRC Doctoral Training Grant (EPSRC reference EP/P505445/1) for funding my PhD, MPLUS software developers for their technical support in using the software, and the UK Data Archive and the DfT NTS team for providing access to the UK NTS data.

Last but not the least, I would like to thank my beloved family and friends without whom the completion of this undertaking would have not been possible. I deeply appreciate the encouragement and moral support of many people whose name may not all be enumerated. In particular, I would like to thank my best and beloved friend, Anoosheh Fallahfard, for her unfailing spiritual support and encouragement throughout my PhD course; my beloved friends and college mates, Hamid Hazby, Milad Shokoohi, Zeynab Raeesy, Ahoora Baranian, Khashayar Ghafarzadeh, Hanifeh Zarezadeh, Laura Panades, Marina Riabiz, Bjoern Zeeb, and many more whose name may not all be listed in this short space.

Abstract

After decades of research, a host of analytical difficulties is still hindering our understanding of the influences of the built form on travel. The main challenges are (a) assembling good quality data that reflects the majority of the known influences and that supports continuous monitoring, and (b) making sense methodologically of the many variables which strongly intercorrelate. This study uses the UK national travel survey (NTS) data that is among the most comprehensive of its form in the world. The fact that it has rarely been used so far for this purpose may be attributable to the methodological difficulties.

This dissertation aims to develop a new analytical framework based on extended structural equation models (SEMs) in order to overcome some of the key methodological difficulties in quantifying the influences of the built form on travel, and in addition to provide a means to continuously monitor any changes in the effects over time. The analyses are focused on employed adults, because they are not only the biggest UK population segment with the highest per capita travel demand, but also the segment that are capable of adapting more rapidly to changing land use, built form and transport supply conditions.

The research is pursued through three new models. Model 1 is a path diagram coupled with factor analyses, which estimates continuous, categorical and binary dependent variables. The model estimates the influences on travel distance, time and trip frequency by trip purpose while accounting for self-selection, spatial sorting, endogeneity of car ownership, and interactions among trip purposes. The results highlight stark differences among commuters, particularly the mobility disadvantages of women, part time and non-car owning workers even when they live in the most accessible urban areas.

Model 2 incorporates latent categorisation analyses in order to identify a tangible typology of the built form and the associated variations in impacts on travel. Identifying NTS variables as descriptors for tangible built form categories provides an improved basis for investigating land use and transport planning interventions. The model reveals three distinct built form categories in the UK with striking variations in the patterns of influences.

Model 3 further investigates the variations across the built form categories. The resulting random intercept SEM provides a more precise quantification of the influences of self-selection and spatial sorting across the built form categories for each socioeconomic group.

Four research areas are highlighted for further studies: First, new preference, attitude and behavioural parameters may be introduced through incorporating non-NTS behavioural surveys; Second, the new SEMs provide a basis for incorporating choice modelling where the utility function is defined with direct, indirect and latent variables; Third, conceptual and methodological developments – such as non-parametric latent class analysis, allow expanding the current model to monitor changes in travel behaviour as and when new NTS or non NTS data become available. Fourth, the robustness of the inferences regarding causal or directional influences may require further quantification through designing new panel data sets, building on the findings above.

Table of Contents

List of Figures

List of Tables

1. Introduction

Accelerated urbanization coupled with fast economic growth and forces of globalization across the globe has been the source of the growing global issues such as inequity, congested traffic, environmental degradation, lack of infrastructure, and housing shortage (Ichimura, 2003); this has led to increasing advocacy for a better understanding of the role of the built form[1] in promoting sustainable development (Berke and Conroy, 2000, Lawrence and Low, 1990).

Travel patterns is known as one of the major influences from built form. There is a long history of more than 60 years of research on the influence of built form on travel, which could be traced back to Mitchell and Rapkin (1954). However, there is a consensus that much is still remaining to be done to achieve a comprehensive understanding of influences, let alone any break in the trends (Næss, 2012, Aditjandra et al., 2012, Cao and Chatman, 2016). The main challenges are twofold: (a) it is very difficult and costly to assemble a high quality dataset which can cover the majority of socioeconomic and built form variables which are known to play important roles in influencing the travel outcomes such as trip frequency, travel distance and travel time even for one specific urban area for one point in time, let alone so doing for a whole country, year in year out; (b) it seems even more challenging for existing research methodologies to tackle analytical complexities once such a dataset has become available.

The comments above on Challenge (b) appear to be borne out by the fact that the UK national travel survey (NTS) which is a well-designed survey that meets a significant range of the data requirements mentioned in Challenge (a) has so far only been sparingly used for investigating the influences of the built form on travel. It seems particularly important to address the gaps in the analytical methodology because many new data sources are emerging which could potentially answer to Challenge (a).[2]

---

[1] In transport planning and urban modelling literature, the terms built form and land use have often been used interchangeably. This is rather unfortunate, because the two terms have distinct meanings. In this dissertation, we follow the definitions from urban planning and urban morphology, to define 'built form' as the combined physical characteristics of an urban, suburban or rural area that are shaped by land use planning and urban design, whereas we use the term 'land use' to mean the uses of land or the distinct activities occurring on the land.
[2] Significant efforts to develop such data collection and analytics are emerging, e.g. Bohte and Maat (2009) and Pawlak et al. (2015).

One of the striking features of the existing research literature is the uneven coverage of the effects. Hitherto, the existing studies tend to explore built form influences on car use and ownership (e.g. Dargay and Hanly, 2004, Giuliano and Dargay, 2006, Cao et al., 2007a, Dargay, 2007, Huang, 2007), and travel distance (e.g. Cervero, 1996, Banister et al., 1997, Stead, 2001, Axhausen, 2003). A limited few evaluate the influences on trip frequency (e.g. Weis and Axhausen, 2009, Silva et al., 2012) and very few studies exist on the effects upon travel time (e.g. Susilo and Kitamura, 2008, Cervero and Duncan, 2006). To the best of our knowledge, no study has yet compared the influences systematically on all the above outcomes. This would appear to be a major gap in the literature, because for tackling the challenges to providing sustainable, equitable and efficient transport in today's city regions it would require us to monitor both accessibility and mobility patterns, which implies that we must have a sound understanding of not only the evolution of car use and travel distance, but also that of trip frequency and travel time (Preston and Rajé, 2007). Preferably, there should be a structured system-level understanding (Schwanen et al., 2015).

This study uses the UK national travel survey (NTS) data that is arguably the most comprehensive household and individual travel survey dataset in the world and does meet many of the demanding requirements for such analysis. As stated in Le Vine and Jones (2012), NTS is the only rich national dataset available to decompose aggregate travel patterns among British residents, and in particular to focus on changing trends over a decade-long timescale. The distinctive advantages of NTS in including all major travel indicators, travel modes, and trip purposes, and the consistent approach used for survey design and data collection after the year 2002 (Hayllar et al., 2005), have made it an ideal dataset for analysing built form influence and its trend of changes. However, the methodological difficulties for dealing with the strong correlations and endogeneity among its variables might have so far hindered its usage for exploiting its full potential.

This research aims to respond to the above analytical challenge. The main research objectives are formulated as follows: a) quantifying the effects of built form characteristics on travel behaviour (i.e. travel distance, time and frequency) after controlling for the main endongeneities in the NTS data, such as self-selection and spatial sorting, car ownership status, and interactions among travel purposes; b) measuring the scale of changes in travel behaviour over time to identify potential trend breaks. The first objective requires developing a new framework for quantifying the influences while controlling for the strong correlations among the variables. Formulating a method for the second objective could lead to an approach to continuous monitoring of the effects into the foreseeable future[3].

This research builds on more than a decade of progress in structural equation modelling (SEM) in the field (Notably Cao et al., 2007b, Gao et al., 2008, Cervero and Murakami, 2010). and constructs a general purpose, robust approach to understanding the complex web of influences as recorded in the NTS data. It further expands the analysis to make a novel use of more recent developments in traditional SEM by allowing combining distribution assumptions and applying novel approaches for modelling influences arising from different categories of built form. This is accommodated through developing three sets of models, each aims to highlight some unique features of built form influences.

The first model is close to a conventional structural equation model which is formed by a path diagram coupled with factor analyses. It aims to understand average built form influences after controlling for self-selection and car ownership endogeneity. Built form is modelled as a continuous latent variable which is handled as a common factor of three highly correlated built form features. The model structure is set to account for self-selection and spatial sorting by modelling the interrelations between socioeconomic and built form latent variables in addition to their influences on three travel outcomes: trip frequency, travel distance and travel time. Car ownership is modelled as an intervening variable where its influences on travel is estimated conditional on car ownership being a function of socioeconomic and built from characteristics.

The second model expands the first by relaxing the assumptions for linearity of built form influences through incorporating latent categorisation analyses (LCA). This model aims to provide insights into the typology of built form in the UK and the variations in their impact on travel patterns. Built form is modelled as a categorical rather than a continuous latent variable in order to identify tangible built form types. It shows that analysing variations in the type of residents and the influences on travel across built form clusters can provide valuable insights for urban and transport planners in adapting policy interventions to built form typologies.

[3] So far as we know the UK Department of Transport intends to continue with the survey into the foreseeable future, in England as least, although there had been several discussions in the recent past to reduce or terminate the annual survey. Besides the academic aims of this dissertation, it is hoped that the work presented here would add to the list of evidence that help convince the government that the survey should be maintained and enhanced, rather than diminished.

The third model further extends the analyses through incorporating random intercepts in the equations. Its aim is to understand more precisely the role of self-selection and spatial sorting vis-à-vis that of specific built form characteristics. This method also identifies the specific tendencies for self-selection and spatial sorting among the socioeconomic groups of the population. Here, instead of modelling built form characteristics as continuous or categorical latent variable, an alternative approach is used which allows the SEM intercepts vary across built form clusters. The process of estimating the socioeconomic influences which explain this variation and subtracting them from the total variations of random intercepts provides a new measure that goes one step closer to measuring specific built form effects.

We use the Mplus software[4] for this dissertation. . It is not currently the widest used software for SEM modelling in transport[5], but it provides a number of specific estimation options that are unique. Mplus allows modelling robust cluster errors which is helpful to control for unobserved correlation among household members within NTS dataset. It is also capable of modelling latent and observed variables with different distribution assumptions. This feature is particularly used in the second model where We have incorporated latent categorical variable into SEM framework, but it becomes also helpful for modelling binary variables such as car ownership in conjunction with normally distributed or count variables (e.g. travel time and trip frequency). We provide more detailed explanations on Mplus's unique features in Section 3.

Employed adults are set as the target population for this study. This is specifically in response to the growing interest in understanding the influences on productivity growth, job access, social equity and environmental sustainability. Employed adults are the biggest UK population segment with the largest extent of mobility. Moreover, commuting and business form the majority of trips in the UK (refer to NTS technical report; Morris et al., 2014). consequently, understanding the travel behaviour of employed adults and determining the built form influences on that is of high policy interest.

Although this study has looked into the accessibility and mobility of less advantaged groups (e.g. low income or part time workers), it has not covered the potential travel needs of pensioners and economically inactive segments. Ideally, the study should have covered the whole population. However, the analysis of that extend would have required segmenting into

[4] See https://www.statmodel.com, accessed 15 May 2015
[5] Van Acker et al (2014) is a recent precedent of applying Mplus in transport research.

more homogenous sub-groups and developing separate models for each to ensure robust statistical inference[6]. This has been beyond the scope of this dissertation. However, the methodology developed in this study can be extended to cover other population groups.

[6] For instance, economically inactive groups are supposed to make no commuting or business trips. Including them with the rest of the population in analysis makes it difficult to separate those with low probability of making trips from those who are not making a trip (i.e. structural zeros).

2. Literature Review

2.1. Introduction

The intellectual and practical interests in the complex built form influences on travel have a long history (the significant works in recent decades being Cervero, 1996, Cervero and Kockelman, 1997, Banister, 1997, Newman and Kenworthy, 1999, Crane, 2000, Ewing and Cervero, 2001, Stead, 2001). However, a comprehensive mapping of the effects is still emerging. First, the empirical datasets that include a wide range of relevant variables are difficult to assemble. Second, the analytical challenges that arise from model specification issues such as endogeneities among variables cast doubt on many estimates (Boarnet, 2004, Cao et al., 2007a, Silva et al., 2012). Third, the economic, social, cultural and physical circumstances within which travel is undertaken are shifting substantially through time; regular and timely updates on the effects – which could provide fundamental insights into the changing travel behaviour – prove particularly difficult to achieve given the data and analytical challenges just mentioned.

Whilst data collection and assembly are largely dependent on funding, skills and the perceived payback, remarkable progress has been made in model specification in recent years. In particular, there is a growing body of literature that aims to isolate the built form effect after controlling the endogeneities among different factors such as the interdependencies between travel patterns, travel attitudes, built form characteristics, and car ownership (Handy et al., 2005, Van Acker et al., 2007, Mokhtarian and Cao, 2008, Gao et al., 2008, Bohte et al., 2009, Cao et al., 2009, Sun et al., 2009, Cervero and Murakami, 2010, Silva et al., 2012, Sun et al., 2012, Zegras et al., 2012)

In order to provide a comprehensive review of the studies on complex influences of built form, the rest of this chapter is organized as follows. Section 2.2 reviews the lineage of research on the influences of built form parameters and socioeconomic characteristics; Section 2.3 discuss approaches in handling the issues of endogeneity and intercorrelation; finally Section 2.4 underlines the research gaps in this context.

2.2. Early studies on the influence of built form and socioeconomic characteristics

The interest in potential built form influences goes back to 1950s when Mitchell and Rapkin (1954) present the first scientific framework of the concept and terminology of "land use and transport interaction". They argue that travel is not an aim by itself and movement system is secondary derived from built form patterns. Their landmark study on urban traffic not only established a link between built form and transport, but also called for comprehensive framework for understanding travel behaviour. Since then it has been a long history of discussions on the extent and strength of the built form influences. Two main strands of classic literature formed the earlier arguments on travel policies: those who believe in considerable role of urban planning and design and the effectiveness of more compact cities (e.g. Bourne, 1992, Cervero, 1996, Cervero and Kockelman, 1997, Banister, 1997, Newman and Kenworthy, 1999), and the sceptics who gives the main role to the enforcing socioeconomic patterns and market mechanism (e.g. Gordon and Richardson, 1989, Breheny, 1992, Farthing et al., 1996, Levinson and Kumar, 1994, Gordon and Richardson, 1997).

The key study which triggers the debates and stimulates further research on the built form influences on travel is that of Newman and Kenworthy. In their book- 'sustainability and cities: overcoming automobile dependence (Newman and Kenworthy, 1999)'- they explore the influences on automobile dependency in 46 cities in the USA, Australia, Canada, Western Europe and Asia. The project which took 7 years to complete is the update and expansion of their earlier work where they define the term 'car dependency' and examine that in 32 cities across the world (Newman and Kenworthy, 1989). Their research presents the broad patterns of car dependency and establish a link between that and built form factors. Their analysis is mainly at aggregate level and descriptive but reveals some important associations. Their study shows the crucial role of built form indicators, particularly population density, in reducing car dependency and the distance that people need to travel. Their analysis of fuel consumption also gives a direct comparison of carbon dioxide emissions around the world and shows the prominent role of built form indicators on that.

Cervero (1996) is another good example of studies claiming significant influence of built form on travel. He examines the job-housing balance in 23 large San Francisco Bay Area cities. The study uses the 1980 and 1990 census data on place of residents and employment and explores the trends of changes in the ratio of jobs to employed residents (they called this jobs-housing balance). They have then explored the influence of change in job-housing balance on that in vehicle miles travelled (VMT). The analysis of jobs-housing balance shows that the rate of imbalances is generally declined in Bay area cities. However, this is mainly derived by the attraction of jobs toward the dormitory communities in 1980, otherwise, the job surplus cities such as Silicon Valley have experienced worsening in imbalances. In addition, the association between balance and self-containment (i.e. living and working in the same area) is shown to be week. Further regression analysis of the influences of housing supply and price shows that the lack of affordable housing in job-rich communities prevent people from living near their workplace. This is consequently the main contributor to growth in vehicle miles travelled (VMT). The author believes this is more of a planning failure rather than a market one. He calls for more of regional or state pressure and incentive for encouraging development of new houses in job rich areas.

The former assertions are largely disputed by the latter group of researchers who argue that households' socio-economic characteristics and life style are the main identifiers of their travel patterns, and further, that the adjustment between living and working places –i.e. reduction in residents' commuting distance and costs- occur over time as the result of market mechanism without the requirement of urban planning intervention.

Gordon and Richardson (1989)'s blistering criticism of Newman and Kenworthy's original thesis (i.e. Newman and Kenworthy, 1989):

> "This reply will argue that Newman and Kenworthy's analysis is faulty, that the problems are wrongly diagnosed, and that their policy and planning prescriptions are inappropriate and infeasible …The idea of planners turning our lives upside down in pursuit of a singleminded goal is as horrible as it is alien. Newman and Kenworthy's world is the Kafkaesque nightmare that Hayek always dreaded, a world in which consumers have no voice, relative prices have no role, and planners are tyrants … NK also make a plea for the cheapest and most fuel efficient transportation modes, walking and biking. Are they recommending the Beijingization of US cities to reinforce their Maoist planning methods? They argue for limits on automobile use, trying to cripple the more efficient mode, but via the adoption of direct controls rather than road congestion pricing. They claim that pain and loss of rights will be offset by more equity, a stronger sense of community, and "real gains in access through new transit systems

> and new, more centrally located housing." But in the United States most poor households need cars, walkers in many locations are more concerned with safety than a sense of community, new transit systems are too expensive and too inconvenient, and developers exploit the availability of central city housing subsidies to build high-income housing."

In addition to the ideological debate between free and socialist economy which is the reflection of the 1990s atmosphere, the basis of Gordon and Richardson's argument is the role of economic factors, life style and socioeconomic characteristics which are largely ignored in Newman and Kenworthy's study. They believe that people's decision on where to live and work should be mainly left to the market mechanism and individuals' life style. , they argue that even in a conservative scenario where the reduction in gasoline consumption is required this can be much better accommodated by increase in gasoline taxation rather than redesigning metropolitan areas or developing expensive transit lines.

In line with increasing awareness of the sustainability issues arisen from motorized travel and car dependency, Gordon and Richardson's later papers recognize travel negative externalities but still see the solution in travel demand management and soft policies rather than urban planning intervention.

> "There are inevitable problems with how we manage the highway system, negative externalities and urban service delivery systems. Approaches to some of these shortcomings, such as time of day road pricing, tradable development rights, and fully portable education vouchers, have been discussed elsewhere (Richardson and Gordon, 1993). Spelling out the details of why such policies are cost-efficient remains a research challenge; but the alternative of attempting a reversal of existing urban development trends is neither feasible nor desirable- Gordon and Richardson, 1997"

The aforementioned earlier studies on built form influences have been mainly at aggregate level and tend to overlook the role of socioeconomic factors; this forms the basis of debates which are discussed above. Over the last decade many studies attempt to respond to this issue by simultaneous analysis of built form and socioeconomic characteristics (Banister, 2000, Stead, 2001, Van Wee et al., 2002, Dargay and Hanly, 2004) though they also produce some seemingly contradictory findings.

Cervero and Kockelman (1997) expands Cervero (1996) by incorporating the socioeconomic characteristics which was missing from Cervero, 1996 and classifying built form into three principal dimensions named 3Ds- Density, Diversity and Design- and test how these dimensions affect trip rates and mode choice of residents in San Francisco Bay Area. 1990-91 BATS (Bay Area Travel Survey) is combined with the Census Transportation Planning Package in order to assemble the required built form, travel and socioeconomic information. However, as a consequence of combining the two dataset, the time dimension and so the capability for analysing the change in trends is lost. Factor analysis is used to form a set of built form factors. The constructed built form factors are modelled alongside the socioeconomic control variables in a multiple regression in order to evaluate influences on vehicle miles travelled per person. To understand the role of built form characteristics, this model is then compared with a reference one where the built form factors are excluded. The study concludes that density, built form diversity and pedestrian oriented design reduce the number of trips and promote non-auto travel; even though for marginal amount. The findings from this study should be interpreted as associative rather than causal due to the use of cross sectional data. In addition, there are some other points which should be considered: a) the study does not control for potential interrelations between socioeconomic and built form characteristics; b) the endogeneity caused by the interrelations between car ownership control variable with other socioeconomic parameters are not considered.

Analysing the NTS data for 1989-1991, Stead (2001) illustrates the important role of population density, mixed-use development and settlement size in explaining travel patterns even after accounting for socioeconomic characteristics; however, the extent of built form influence is lower than what indicated in earlier studies with no control for the effect of socioeconomic factors. Multivariate regression is used for the analyses at individual level. Socioeconomic variables include gender, age, employment status, household size and composition, but also household car ownership which is in general highly interrelated with other socioeconomic variables. Built form characteristics such as proximity to local facilities, high street shops, bus stops and railway stations, population density at local authority and ward level, and settlement size are also modelled as exogenous determinants alongside the socioeconomic variables. He concludes that socioeconomic characteristics can explain half the variation in travel distance and car use while built form parameters can maximum explain one third of changes in travel distance. He also argues that there is some interactions between socioeconomic characteristics and built form factors; however, he does not examine the extent and form of the interrelations. Stead's study is one of the first and few study which uses the time series of UK National Travel Survey dataset to analyse the built form influences. The study has also attempted to control for a large number of the socioeconomic variables. However, the study is not controlling for potential interactions across the independent variables including socioeconomic, built form, and car ownership.

Van Wee (2002) adds the preferences for travel modes into multivariate regression model of built form, personal and household variables- including car ownership- on travel distance. The study is based on the questionnaire of 446 individuals collected across three neighbourhoods in towns close to the Dutch city of Utrecht. The study concludes that preferences for travel modes has an important role to play and its exclusion can result in overestimating other influences including built form parameters. However, he concludes that built form patterns has still significant effect which should not be neglected. The author highlights the requirement for more rigorous research which control for the interactions between preference, socioeconomic factors, built form, and travel patterns and suggest the use of structural equation models and longitudinal data as future potential analytical methods.

This strand of studies, however, produces some seemingly contradictory results. For instance, Stead (2001) concludes that there is no strong relationship between settlement size and travel distance. This outcome is also confirmed by Hickman et al (2004) who find no general increase in travel associated with increase in population size in the county of Surrey. On the contrary, using the same Dataset as Stead (2001)- i.e. NTS, 1989-1991, Dargay et al (2003) show association between increase in the size of municipality and decline in travel distance especially within London and metropolitan areas. Their follow up analysis-i.e. Dargay et al (2004), shows similar influences on car ownership and mode share. This contradictory findings from analysing the same dataset might be explained by variations in analytical methods. Notably, while Stead (2001) has included car ownership along other socioeconomic variables in his analysis, Dargay et al (2003) has excluded that from their regression model.

Another example is population density for which there is a larger agreements on its important influence on travel behaviour (e.g. Cervero, 1996, Cervero and Kockelman, 1997, Banister, 1997, Stead, 2001, Van Wee et al., 2002, Maat et al., 2004, Dargay and Hanly, 2004). However, it is believed that the form and extent of its influence can be affected by the potential interactions with socioeconomic characteristics, individual preferences, and among travel outcomes. Van Wee (2002) has highlighted the potential influence from interactions between number of trips and total travel distance; higher density may offer the possibility for travelling less due to improved accessibility. However, this might also encourage participation in more number of activities which might consequently increase the travel distance. Applying structural equation model on a two-day travel diary in Netherlands, Maat et al (2004) highlight the interactions between individual characteristics and built form parameters; the households with children tend to live in lower-dense areas while preferences for residing in denser areas increase with income. They also tentatively conclude that the effect of density on total travel time is only indirect and through the number of trips-i.e. extra trips in higher dense area results in longer travel time.

The manifold explanation for the contradictory findings above can be classified into two main groups: first, the target travel outcomes and the explanatory variables vary across studies; second, the analytical challenges arising from endogeneities among highly interrelated socioeconomic, preferences, and built form variables has made it difficult to gain a robust judgement of the effects; third, the nature and magnitudes of the influences are expected to shift substantially through time. While it is hard to assemble a comprehensive dataset to explore and compare wide range of influences on the main travel outcomes and track them over time, there has been extensive progress in model specifications and development to deal with the issue of endogeneity. Section 2.3 below provides a review of main recent literature attempting to tackle the issue of endogeneity by making use of more advanced econometric methods.

2.3. Endogenous effects

2.3.1. What is the problem?

In econometrics, the *endogeneity bias* occurs when an independent covariates in the model is correlated with the model error term. *Endogeneity* refers to the presence of an unobserved (endogenous) factor which is not accounted in the model (Wooldridge, 2009). When this endogenous factor is correlated with an explanatory variable as well as the model outcome, we see a correlations between the error term and the explanatory variable. This results in a biased estimation of parameters and ambiguous causal links. In other words, the estimated relations between the explanatory variables and the outcome of the model might not be the true causal but mainly an association from the effect of the underlying unobserved factor.

There are a wide variety of sources of endogeneity which results in the term to be used for a wide range of problems, some with very different remedy approaches. The major source of endogeneity problem are *model misspecification* (the issue of omitted variable), *simultaneity*, and *Measurement error* (Wooldridge, 2009).

Misspecification means the model does not account for some important variables which are correlated with one or more model independent variables as well as the model outcome. Consequently, we face an ambiguity on the causal effect, make it difficult to infer the marginal effects of the included variables. Putting this in a methodological context, the omitted variable forms part of the model error term, make that correlated with one or more explanatory variables. This is the violation of the OLS assumption which results in bias estimation.

Simultaneity refers to two interrelated variables, each one affecting the other. This is also called the issue of reverse causality. For instance, car ownership can be considered as an explanatory variable for travel distances with those having access to car are more likely to travel longer. However, the reverse can be also valid with those who need to travel longer (e.g. living further from their workplace) tend to acquire car. This particular simultaneity issue can be a potential problem in examining built form influences and is discussed further in Section 5.1.

Measurement errors happen when the true and the measured value of a variable are not the same. This normally deals with the data collection and the study design which should be properly randomized. When the measurement error is *random*, it will not cause endogeneity problem as is not correlated with the dependant variable. However, the endogeneity problem can arise when the measurement error is *systematic* (e.g. when part of population is measured differently from the rest) and correlated with the dependant variable.

Earlier studies on built form influences discussed in Section 2.2 have not controlled for potential endogeneities. Even more recent ones, e.g. Stead, 2001, which include socioeconomic characteristics (partly to deal with omitted variable biased from excluding them) fail to account for potential interactions between built form and socioeconomic variables. The rest of this chapter review more recent studies which aim to account for these shortcomings.

2.3.2. The endogeneity issues in examining built form influences

There is a growing body of literature that aims to isolate the built form effect after controlling the endogeneities among different factors such as the interdependencies between travel patterns, travel attitudes, built form characteristics, and car ownership (Handy et al., 2005, Giuliano and Dargay, 2006, Silva et al., 2007, Van Acker et al., 2007, 2014, Mokhtarian and Cao, 2008, Cao et al., 2007a, Cao et al., 2007b, 2009, Gao et al., 2008, Bohte et al., 2009, Sun et al., 2009, Cervero and Murakami, 2010, Silva et al., 2012, Sun et al., 2012, Zegras et al., 2012). Notably, Gao et al (2008) analyse the connections between job accessibility, workers per capita, income per capita and cars per capita with census tract data for Sacramento, CA by employing a Structural Equation Model (SEM) to capture endogeneity effects. They find that the error terms of many variables strongly correlate and a multivariate regression model would overestimate the significance of their influences.

Residential self-selection or sorting effect is one of the endogeneities which has attracted a great deal of attention. As outlined by Cao et al (2007b), the question is whether neighbourhood design independently influences travel behaviour or whether preferences for travel options affect residential choice. Using a self-administered twelve-page survey of 1682 respondents from eight neighbourhoods in Northern California, Cao,et al (2007a, 2007b) and Handy et al (2005, 2006) analyse the factors affecting car ownership. Explanatory Factor Analysis (EFA) is employed to construct latent variables from responses to neighbourhood characteristics, neighbourhood preferences, and travel attitude. The data and constructed factors are used to explore the role of the self-selection effect in explaining travel patterns. Notably, Cao et al (2007a) examine the influences of neighbourhood characteristics, neighbourhood preferences, travel attitudes, and socio-demographics on car ownership. The findings from cross sectional analysis show that the correlation between neighbourhood characteristics and car ownership is primarily the result of self-selection.

Cervero and Murakami (2010) represents an important landmark in tackling both the data and model specification challenges through assembling a very large dataset from 370 US urban areas in 2003 and employing an extensive SEM to examine the effects of density, diversity, destination accessibility and design on vehicle miles travelled (VMT), building on analyses of the first three Ds in Cervero and Kockelman (1997). They analyse a complex web of interactions among built form characteristics, average household income and travel demand, where travel demand is represented as VMT, percentage of commute trip by private car and rail passenger miles per capita. Their findings, after evaluating the interrelation between road density and population density, suggest that the largest reduction in vehicle travel distance comes from the combination of compact design and below-average roadway provision.

Silva et al (2012) is one of a limited few examples which have examined car ownership as an intervening variable in influencing total kilometer travelled and trip frequency. In addition, they control for self-selection effects by modelling concentration, density and diversity as a function of socioeconomic attributes in their SEM framework. Their results suggest that after controlling for socioeconomic self-selection effect, built form variables significantly affect commuting distance and car ownership.

The studies that control for endogeneities tend to focus on the effects on distances travelled. A small number of studies have been able to examine the influences on trip frequency (e.g. Weis and Axhausen, 2009, Silva et al., 2012), and on travel time (e.g. Giuliano and Small, 1993, Cervero and Wu, 1997, Cervero and Duncan, 2006, Susilo and Kitamura, 2008). In particular, Weis and Axhausen (2009) construct a pseudo-panel dataset based on the Swiss National Travel Survey to examine the aggregate effects of generalised travel costs upon the number of trips and journeys conducted, the resulting total out-of-home times as well as distances travelled. They use an SEM to control for self-selection effects. They confirm that the generalised cost and accessibility elasticities are substantial after controlling for age, cohort and other socio-demographic factors. However, they find it surprising that the model reports no significant income and car ownership influences on travel.

The study of temporal changes is so far focused on causal quantification of the effects from quasi-panel data sets. Figure 1 illustrates the conceptual model of Cao et al (2007b) who use a quasi-longitudinal data of movers (688 respondents who changed their residential locations over the previous year) to analyse the interdependencies between socioeconomic factors and built form characteristics. Their study is able to identify a similar to or larger influence of built form- specifically accessibility- on travel behaviour than those of socio-demographic.

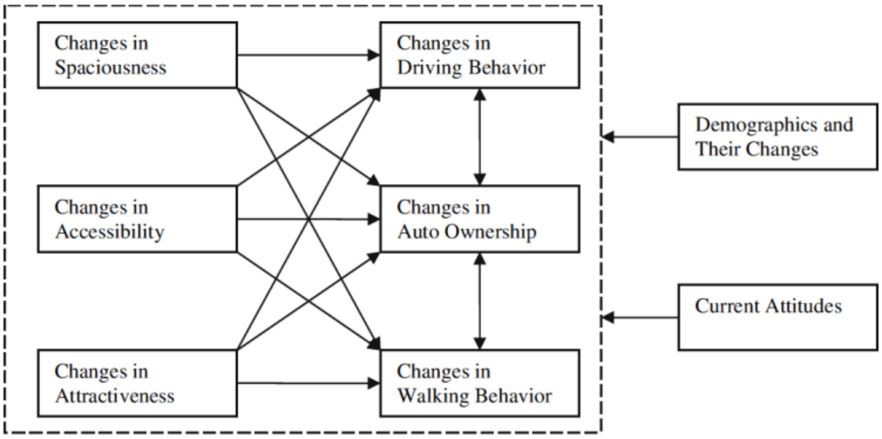

Figure 1 Conceptual SEM Source: (Cao et al., 2007b)

Cao et al (2007a) compare cross sectional and quasi-panel analyses of influences on car ownership. The cross sectional analysis find no correlation between neighbourhood characteristics and car ownership- specifically after inclusion of preferences- but some marginal influence of built form constructs are found when quasi-longtitudinal approach is adopted.

Adopting a quasi-longitudinal SEM approach, Aditjandra et al (2012) report similar conclusions of the impact of neighbourhood design (e.g. accessibility, safety, attractiveness) upon the amount of private car travel after controlling for self-selection. Using Tyne and Wear metropolitan area as their case study, this is one of the first studies of this kind which has used British metropolitan data. It is also a recent study which has controlled for the endogeneity of car ownership in influencing travel; they conclude that neighbourhood design affects travel behaviour through their influence on car ownership.

In addition to the SEM studies discussed above, recent years have seen rapid progress in choice modelling techniques to deal with interrelations across influences. These include relaxing one or more IID error structure assumptions in MNL, and/or unobserved response homogeneity assumption through allowing random coefficients for attributes. Also, more studies are incorporating latent clustering and factor analysis into choice modelling.

Although, choice modelling is not directly employed for this study, some of the advances in that (such as random coefficients or latent clustering) are incorporated in SEMs for this study. As such, we have provided a brief review of recent developments in this area in Appendix A.

2.4. Research gaps

The above review of literature has underlined some significant gaps on which this study attempts to shed a light on. We categorize them into two specific classes: the quality of data, and analytical methods. Data availability often pre-determines the study method so the two are highly related. Below we are going through them in turn.

2.4.1. Quality of data

It would seem that in two aspects the current studies still have under-tapped potential which can be associated with the prevailing data difficulties. First, most studies reveal insights into the influences on distance travelled, but so far only very few do so on trip frequency and travel time; this limits the understanding of influences on travel accessibility and leaves an apparent gap on mobility. Secondly, few existing studies except the census-based longitudinal work could easily provide regular updates going forward without major data efforts. This is foremost a data issue, as few researchers would disregard particular explanatory variables or travel outcomes if suitable data is available.

This issue is well outlined in Preston and Rajé (2007) where they discuss inequality in transport. They show that it is necessary to understand both accessibility and mobility patterns in order to design effective policy responses. The problems of the disadvantaged cannot be analysed in isolation from the rest of society, not least because we need to have a clear understanding of how big the disadvantage gaps are among them. For rich countries where the majority of travel is suburb to suburb, some enjoy fast and smooth car or rail journeys whilst others rely on infrequent, expensive and poorly connected public transport. Such differences could arise from a wide range of causes, such as demographic-socioeconomic circumstances, built form, gender, life-cycles, lifestyles, ownership/access to car, social and environmental attitudes, etc. Furthermore, the circumstances and attitudes could evolve rapidly, given the momentous changes in labour market and wider society.

In this context, it is a little surprising that the potential of UK National Travel Survey (NTS) has not been more fully investigated for this purpose. The NTS has been collecting an extensive household sample dataset since 1965, and since 1988 the survey has been carried out every year. The survey is conducted as home interviews of all household members, recording a detailed one-week travel diary together with carefully selected personal, household and circumstantial variables that are thought to influence travel behaviour. The data is weighted to provide annual updates on all main purposes of domestic travel in terms of travel distances, times and frequency. The list of the variables is arguably the most comprehensive among nation-wide travel surveys.

Hitherto, there are only very limited attempts to relate travel patterns to the extensive range of the NTS variables (Stead and Marshall, 2001, Stead, 2001, Dargay and Hanly, 2004, Jahanshahi et al., 2009, Jahanshahi et al., 2013, Susilo, 2015), and none except the last one have made use of the improved time series of survey results since 2002. Methodological limitations tend to be the main reason that has held back a fuller exploitation of the comprehensive list of NTS variables; the personal, households and circumstantial variables are highly intercorrelated because of self-selection, spatial sorting and other endogeneities. In addition, there may be interactions among trip purposes and travel outcomes (e.g. long distance commuters might travel less frequently and forgo some other trips). Unlocking insights in the data would require robust models that can cope with such complexity.

2.4.2. Conceptual and methodological gaps

There is ample room for improvement in conceptual design and analytical methods regarding the influence of built form influence on travel. First, existing studies collectively suggest that significant endogenous interactions exist among the influencing factors like travellers' socioeconomic and demographic profiles, residents' self-selection and spatial sorting, built form and to some extent car ownership; however, few if any studies have examined this whole range of influences in one model.

Secondly, there are some potentially important interactions that have been left under- or un-investigated, such as among different trip purposes or different travel outcomes. For instance, would longer commuting be offset by shorter or fewer shopping journeys, or less frequent travel imply longer distances or durations?

Thirdly, recent studies have benefited from improvements in statistical methods to accommodate heterogeneity among individuals or choices, and characterized the latent states behind individuals' decisions; however, we are not aware of any study which aims to accommodate these techniques within the SEM for understanding the built form influences. Categorizing geographical locations through latent class analysis in combination with SEM can better quantify the built form effect to inform built form and transport policies and models. In addition, the incorporation of random intercept models could make it feasible for simultaneously controlling potential endogeneities through SEM and measuring the macro level variations when individuals are nested within more aggregate units. This is the reason for its wide use in other disciplines such as education and health (Cho et al., 2015, Dunn et al., 2015, Marsh et al., 2009, Marsh et al., 2015) where individuals are nested within aggregate units (such as schools in educational studies or built form clusters in our context). It would therefore appear of both theoretical and policy interest to incorporate LCA and random intercept models in SEM for examining the more complex and controversial aspects of influences on travel behaviour.

2.4.3. Summary

The above literature review illustrates the requirement for systematic analyses of comprehensive range of socioeconomic and built form characteristic on all aspects of travel (i.e. accessibility, mobility and frequency of travel) in order to fully comprehend the influences on travel behaviour and measure the extent of built form influences. Not only does this require a high quality travel survey dataset, but also an appropriate methodological approach for modelling highly inter-correlated parameters and accounting for potential endogeneities. This study aims to contribute by employing extended Structural Equation Modelling (SEM) framework to analyse a dataset from the UK National Travel Survey, which is arguably one of the most comprehensive ones, if not the most, in the world. Chapter 3 below provides an extensive explanation of the analytical methods used for this study.

3. Method of Study

3.1. Overview

The literature review in Chapter 2 highlights the main areas that are calling for further studies in exploring influences on travel behaviour. In general, despite relatively rapid methodological developments, there are still few systematic analyses of large and reliable datasets which enable the development of a methodology to account for a complex web of exogenous and endogenous influences and form a robust basis for tracking such influences over time.

In order to fulfil these demanding requirements, SEM in combination with random effect and multi-level analysis is adopted for our study. The review of the major econometric methods in dealing with endogeneity and interrelation problems (refer to Section 3.2) supports the use of SEM for this study. Considering the nature of problem and data in hand, we conclude that SEM provides the most comprehensive control for the interactions among many inter-related variables we are dealing with.

Figure 2 presents the SEM structure which forms the basis for all the developed models in our study. The proposed structure is based on the reviewed literatures and the aim to deal with highlighted issues in section 2 including the endogeneity of car ownership and the built form dependency on socioeconomic characteristics. However, the potential effect of SEM structural ambiguity- in our case the reverse causality from the influence of accessibility (travel distance) on car ownership, should be also considered. This chapter explains the models developed for this dissertation based on the overall structure proposed in Figure 2. The analyses on the potential effect of adopting alternative structure (structural ambiguity) on built form influences are provided in section 5.1.

Three sets of SEM-based models are developed in succession and estimated by employing the expectation–maximization (EM) algorithm which allows estimating latent variables and the joint distribution of variables with different distributing assumptions. Below we have provided a brief technical overview of the three developed models.

First, we formulate a set of extended structural equation models (SEMs) to uncover the influences of latent built form characteristics, indirect influences on car ownership, interactions among trip purposes as well as residents' self-selection and spatial sorting (we call this path-diagram based SEM). In doing so, whilst maintaining statistical rigour, the interactions and endogeneities among the many variables will need to be dealt with. This includes estimating the influence of built form conditional on socioeconomic characteristics; estimating that of car ownership as a function of built form and socioeconomic variables; and controlling for interactions among travel purposes.

To do so, we construct a latent continuous variable (which is termed a 'built form factor' below) - following factor analysis- based on the most relevant built form characteristics from the NTS dataset. The factor is modelled as a function of socioeconomic characteristics.

As discussed in Chapter 2, most studies have acknowledged one way or another that car ownership is an outcome of socioeconomic, location and built form characteristics, but few have made rigorous attempts to model car ownership as such. In this study, we estimate a model of logistic distributed car ownership alongside travel distance, travel time, and number of trips.

Second, we improve the representation of the built form latent variable through formulating a Latent Categorical Analysis within SEM framework (we call this LCA-SEM) where the continuous latent variable for built form in the first model is replaced with a categorical one. This facilitates a tangible typology of the built form across the UK and the varied influences of the built form on travel patterns. Whilst the SEM is particularly helpful in defining direct and indirect influences on travel patterns after controlling for highly correlated built form indicators, the LCA-SEM has the potential to provide categorizations by built form type.

Finally, we further combine the potential of random intercept analysis in modelling macro-level variations with that of SEM in controlling endogeneities through developing a random intercept SEM. This is a two-level SEM where the intercepts can vary across built form clusters, allowing the modelling of the influences both within and between the clusters. Allowing variations in intercepts across built form clusters and estimating the extent to which these variations can be explained by macro-level socioeconomic parameters improves the handling of the self-selection effect at a far more detailed level than hitherto achievable.

The rest of this chapter are organized as follows. Sections 3.3 to 3.5 present technical details of the SEM, LCA-SEM and two-level SEM respectively. Section 3.6 provides a brief overview of tracking changes over time. Section 3.7 provides a summary and consider the more general implications of the methods.

3.2. Main econometric approaches addressing the endogeneity problem

As discussed in Section 2, establishing a causal link between built form and travel patterns is complicated. This is due to the endogeneity influences including that of car ownership and residential self-selection. The major attempts in transport studies to deal with these issues have been highlighted in Section 2.3.2. Here, we provide a broader review of the major econometric methods to control the endogeneity.

In the ideal world, randomized experiments is the best approach to examine causal effects (Angrist and Pischke, 2008). Through randomized allocation, one can omit the selection bias and hence isolate the causal influences. Eliminating the population sorting effects make it possible to evaluate the causal effect of the treatment. Moreover, randomized experiments are fully non parametric approach which facilitates the explanation of results.

With the exception of recent developments in education studies (e.g. Angrist, 2004, Jahanshahi, 2016), randomized experiments are not yet as common in social science, as in medicine (Angrist and Pischke, 2008). Part of the reason is the practical difficulties of implementing randomized trials for social studies. For instance, in analysing the influences of built form on travel patterns, we should have ideally allocated respondents randomly to built form clusters and then study their behaviour for a period of time. Not only would this have eliminated potential influences of residents' preferences and socioeconomic characteristics on travel, but also it would have cancelled the spatial sorting effects: the indirect influences arising from respondents' choice of residential location based on their travel preferences.

It goes without saying that, as in many social studies, randomized experimental approach is not feasible for this research. It is simply not practical to allocate respondents randomly to residential locations and examine their travel behaviour. Therefore, some alternative methods should be adopted to make the best use of the available data while controlling for the effects of the endogeneities.

Non experimental methods can be classified into three prototypical estimators: a) before-after estimators when the dynamic panel data is used to evaluates the same person in the treated and non-treated states (e.g. Autoregressive models, fixed effects); b) cross section estimator where different people are compared at treated and untreated states at a point in time (e.g. regressions, instrumental variables, control functions, regression discontinuity, SEM); c) The hybrid models such as difference in difference estimators where the time dimension in panel data is replaced with grouping data by cohorts. The cohorts are assumed to be homogenous with no within-cohorts variations due to the treatment.

The estimators in group (a) and (c) are not suitable for our study. In our context, the former requires panel data and the latter needs identifying built form clusters with homogenous residents. As we discussed in Chapter 2, we are not aware of any panel data with a decent sample size and comprehensive list of socioeconomic, built form and travel patterns characteristics to support the goals of this study. Alternatively grouping data into homogenous built form cohorts has its own complexities. First, it requires establishment of built form cohorts with reasonably small within-cohorts variations in residents' socioeconomic characteristics and travel patterns. Second, changes in built form characteristics happens over relatively a long period of time which imposes a requirement for long time series information which is not available in a consistent format from NTS.

Consequently, we have decided to make the best use of the available comprehensive cross sectional data and adopt a technique which can best deal with endogeneities in our settings. The rest of this section provides brief explanations of these major techniques.

3.2.1. Linear Regression analysis

In the absence of randomized experiments, linear regression is one of the simplest approach in establishing causal relations. By including control covariates, linear regressions aim to isolate the marginal effect of the treatment. In our study for instance, linear regression can assist in controlling for the influences of socio economic characteristics in examining the effect of built form on travel patterns. However, the regression assumptions can be restrictive.

In addition to the homoscedasticity (refer to e.g. Wooldridge, 2009 for the full definition), linear regressions assume that the error term is uncorrelated with the model covariates. The violation of the latter (i.e. the issue of endogeneity) leads to the biased estimation of the regression parameters.

As discussed in Section 2.3.1, the omitted variables which are correlated with one or more covariates as well as the model outcome is the main source of endogeneity problem. Moreover, the exclusion of the important interaction terms as well as the issue of reverse causality can cause endogeneity.

In conclusion, while the simple regression model helps dealing with some of the problems arising from the absence of randomized experiments, the endogeneity is the main barrier to establish causal effects. In the analysis of travel behaviour, there are large interactions across socioeconomic variables, built form characteristics and travel patterns in addition to the potential issues of reverse causality. Therefore, more advanced statistical methods are required to control for the endogeneity.

3.2.2. Instrumental Variables

To explain how Instrumental Variables can be used to determine the causal effects, consider a simple linear regression below:

$$Y = X\beta + \epsilon$$

The OLS estimator provides the following estimation for β.

$$\beta_{ols} = (X'X)^{-1}X'Y = (X'X)^{-1}X'(X\beta + \epsilon) = \beta + (X'X)^{-1}X'\epsilon$$

β_{ols} is only equal to the true coefficient when the error term is uncorrelated with covariates (i.e. $X'\epsilon$ and so the second term is 0). IV method can be used when this assumption does not hold (i.e. one or more of X variables are endogenous).

The requirement for IV variables, Z are: a) they should be highly correlated with the endogenous X ($E[X'Z] \neq 0$) b) their correlation with the outcome Y should be only through the endogenous variable, X (i.e. $E[Z'Y|X] = 0$), and c) Z should be uncorrelated with the error term (i.e. $E[Z'\epsilon] = 0$) which is the by-product of condition b. Moreover for the model to be identifiable the number of instrumental variables, Z should be at least the same as number of endogenous variables.

Now let's split the covariates into endogenous and exogenous and call them X_1 and X_2 respectively. We would then have:

$$Y = X_1\beta_1 + X_2\beta_2 + \epsilon$$

In order to resolve the endogeneity issue, normally the two-stage procedure (2SLS) will be adopted. First, the endogenous variables are estimated with only instrumental variables, Z, and exogenous regressors, X_2.

$$X_1 = X_2\gamma_1 + Z\gamma_2 + \eta$$

The predicted value of X_1 from the above equation ($\hat{X}_1 = X_2\gamma_1 + Z\gamma_2$) would be then substituted in the original regression. By separating out the η which is the correlated component of X_1 with ϵ, we have eliminated the endogeneity.

3.2.3. Control Functions

Control functions were formally introduced by Heckman and Robb (1985). The approach aims to expand the idea of instrumental variables for nonlinear and non-invertible models such as discrete choice models.

Let's start from a generalized function below where X_1 and X_2 are endogenous and exogenous variables respectively and $f(X_j, j = 1 \text{ } or \text{ } 2)$, a linear or nonlinear function of X_j.

$$Y = f(X_1, X_2, \epsilon)$$

We can also think of Z as an instrument variable (i.e. uncorrelated with ϵ) which is linked to X_1 as stated in the following equation:

$$X_1 = X_2\gamma_1 + Z\gamma_2 + \eta$$

In the instrumental variable approach X_1 is estimated and replaced into the main model of $Y{\sim}f(X_j)$ by using a 2SLS procedure. This eliminates the error term, η, which is the source of correlation between Y and X_1. Control function approach, however, accounts for the correlation, ρ between ϵ and η by modelling that implicitly.

$$\epsilon = \rho\eta + e$$

As $E[\eta e] = 0 \; and \; E[z'e] = 0$, plugging ϵ from the above equation into the model would eliminate the endogeneity. Therefore, model parameters as well as ρ can be estimated consistently.

3.2.4. Structural Equation Model (SEM)

The structures of the expanded SEMs used for this study are discussed in detail in sections 3.3 to 3.6. Here, we provide a brief explanation of SEM and its advantages for our analysis. We also discuss the potential issues which should have been considered.

SEM is essentially a union of path analysis and latent variable analysis. Path analysis is similar to a system of simultaneous regression equations where there can be mediating variables, i.e. an independent variable in one equation is in turn a dependent variable in another. The equations are of a more general form and 'structural' in that the correlations both among measurement errors and between measurement and specification errors can be controlled for and with mediator variables which can help testing the hypothetical mechanism of causal effects[7]. A latent variable is an unobserved one which is represented as a function of observed variables; latent variable analysis is similar to factor analysis (when the latent variable is continuous), except that the modeller can decide in advance what the constituent factors are based on prior hypotheses and explanatory factor analysis (Wang and Wang, 2012). Although the theoretical benefits are understood fairly early (e.g. Golob, 2003), the methods are only made accessible in stages through specialist estimation software, which are still being actively extended.

Through the path analysis, SEM provides the opportunity for systematic study of mediating effects (indirect effects) where the whole or part of the influences on the outcome variable is through one or more additional variables (i.e. mediator variables). In this context, a direct effect is the influence of an explanatory variable on a dependent variable. An indirect effect is the influence of an independent variable on a dependent one through one or more intervening variables along the path diagram. Through estimating all the direct and indirect influences, SEM measures simultaneously the covariance structures of multiple, highly interrelated variables.

[7] Note that mediator variables are different to the interaction terms which can be included in the simple linear regression. The latter does not have a causal interpretation and only moderate the strength of the influences.

The ability of SEM in representing and modelling complex interrelated variables in one combined framework is quite beneficial to our study. As discussed in section 2, the influences on travel patterns are complicated due to the potential self-selection effects which give rise to the endogeneity issues. For instance, a group of people with specific socioeconomic characteristics might choose to reside in certain areas in order to moderate their travel distance. This indirect effect should be taken into account in estimating the direct effect of socioeconomic characteristics on travel distance. This would be further discussed in section 3.3 where our modelling framework is explained. The importance of mediator variables is also presented in section 5.2 where the findings from our path-diagram based SEM are discussed and compared with alternative models with and without mediating terms.

Another main advantage of SEM is its ability in specifying latent variables along with the path analysis. This gives the potential to apply data dimensionality reduction techniques in order to map the highly correlated characteristics into lower dimensional latent constructs. In our study, we use data dimensionality reduction techniques to construct “built form” latent variable. This allows mapping highly correlated built form characteristics into one continuous or categorical latent variable. Moreover, SEM regression can be modelled with random slopes and intercepts which allow more in-depth analysis of interactions. These various techniques are explained in sections 3.3 to 3.5 below.

The main drawback with SEM is the potential structural ambiguity. SEM requires the modeller to provide a conceptual model in the form of a path diagram with any latent variables embedded in them. The path diagram effectively represents the hypothesis of causal effects. It is tested on empirical data to determine how valid the hypotheses are through computation of robust errors. The modeller can reconfigure the paths and variables based on fit and overall model performance. However, one might derive alternative model structures with similar level of goodness of fit to the data. In our case, for instance, the major issue is the directionality of the influences between travel distance and car ownership; whether people decide to acquire a car based on their travel distance or they travel longer distances when they have access to car. One way to deal with this issue is to test how alternative structures affect the influences of interest (built form influences in our case). The effect of structural ambiguity for this study is examined in section 5.1.

3.3. A path-diagram based Structural Equation Model Framework [8]

Building on our prior experience in NTS analysis (Jahanshahi et al., 2009, Jahanshahi et al., 2013, WSP, 2009) and based on extensive review of literatures, we have settled upon a conceptual path diagram that consists of following three types of explanatory variables (cf Figure 2): (1) a long list of demographic and socioeconomic characteristics as exogenous variables; (2) a single, latent variable which we call 'built form' that reflects the composite characteristics of close associations among NTS variables like population density, built form type, levels of access to public transport services, etc. Explanatory Factor Analysis helps constructing "built form" factor which is then used in the overall framework- this will be further discussed in Chapter 5; (3) car ownership as a mediating, endogenous variable that is subject to influences from both (1) and (2). This enables us to test the level of any indirect influences where exist.

For dependent variables, we exploit the fact that the NTS records three travel outcomes: travel distance, travel time and trip frequency by trip purposes. We set up one SEM respectively for each of the three travel outcomes, within which the amounts of travel by trip purpose are defined as separate dependent variables – this will allow me to see if any complementary or substitutive effects exist among the trip purposes for each of the three travel outcomes.

[8] This section is based on Jahanshahi et al (2015).

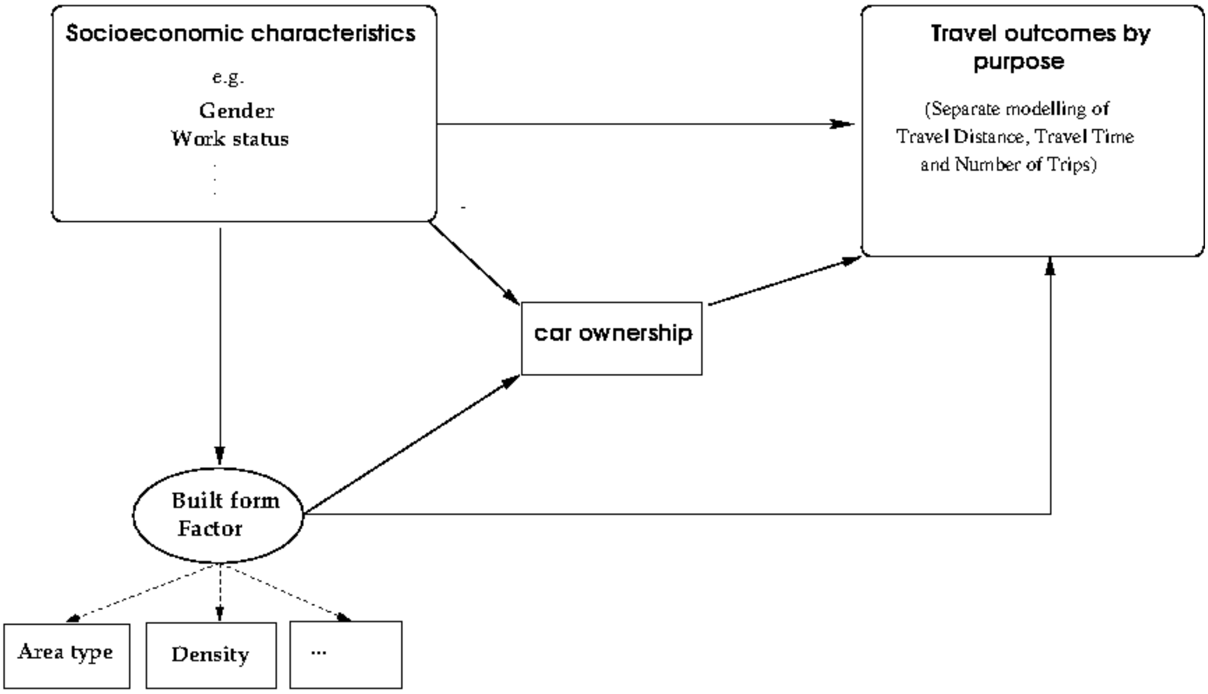

Figure 2 The Conceptual model for testing travel survey data[9]
Note: The rectangular represent the observed and eclipse represents the latent variables

As stated in the introduction, we use the Mplus software which provides some unique features essential for our analysis. First, because the NTS reports travel by all members of a household and there may be unobserved correlations among household members, we use Mplus to control for robust cluster errors. Mplus provides two alternative approach for estimating unobserved correlations. One is to use random effect estimators, where the parameters for within-cluster correlation (i.e. correlation between household members in path-diagram based SEM) is estimated in a hierarchical model structure (we use this feature for two-level model explained in section 3.5). The alternative, which is used for path-diagram based SEM, is the post-estimation of the cluster-robust standard errors using sandwich estimator[10]. Secondly, to estimate the model to compare different time periods, we use a multi-group structure available in the software. Thirdly, unlike other softwares that support only continuous dependent variables, Mplus is capable of analysing observed variables that are continuous (e.g. travel distance and travel time), censored, binary, ordered categorical (ordinal), unordered categorical (nominal), counts, or a combination of the above, which is suited for the NTS dataset where there are categorical variables (e.g. car ownership) and counts (trip frequency). In our model, we use probit regression for estimating binary and categorical dependent variables (car ownership is an example of the former and the indicators of the built form latent variable is an

[9] 'Travel characteristics' mean travel distance, travel time or trip frequency – i.e. a separate SEM is tested for each travel characteristic.
[10] Refer to Cameron and Miller (2015) for more information on cluster-robust interference

example of the latter), multinomial regression for continuous variables (i.e. built form latent variable, travel distances and travel time), and the negative binomial regression for counts (i.e. trip frequency).

Mplus is also capable of analysing zero inflated or truncated models – however, because our subject of analysis is employed adults who tend to travel in a working week, we do not have the problem of zero inflation in travel outcomes; the Mplus feature is nevertheless useful for analysing other types of travellers such as the retired and the elderly following this approach.

Furthermore, we are able to test both weighted least squares (WLS) and maximum likelihood (ML) estimators in Mplus. WLS is more widely used and can produce standardised, unitless coefficients as well as absolute goodness of fit statistics (e.g Chi-square), but ML is now considered more efficient, providing more precise quantification[11]. Also, the more advanced models which we are presenting later can be only estimated using simulation based ML estimator.

We provide technical details in sections 3.3.1 and 3.3.2 below regarding the overall modelling framework and an example of applying it for its most complex use - i.e. carrying out a negative binomial regression for trip frequency with a normally distributed built form latent variable and a probit model of car ownership.

3.3.1. SEM specification and estimation

We have chosen a novel Mplus option that enables an integrated SEM estimation. Here we follow the general notation of Muthén and Asparouhov (2007) in presenting the equations for our model, extending the notation where needed.

The equations for the observed and latent dependent variables for individual i are respectively:

$$\Upsilon_i = \nu + \Lambda\eta_i + \mathrm{K}\mathrm{X}_i + \epsilon_i \quad (1)$$

$$\eta_i = \alpha + \mathrm{B}\eta_i + \Gamma\mathrm{X}_i + \zeta_i \quad (2)$$

where Υ_i refers to a vector of observed dependent variables (e.g. total weekly travel distance, travel time and trip frequency, household car ownership, etc); η_i is a vector of a latent variable, or more specifically the built form latent variable in our model; X_i is a vector of

[11] For more information see discussions on Mplus forum, e.g. http://www.statmodel.com/discussion/messages/11/657.html?1342887417; accessed 14 May 2015.

observed variables; ν and α are vectors of intercepts; Λ, K, B, and Γ are matrices of slope and regression parameters; ϵ_i is a vector of residuals with mean zero and covariance Θ; ζ_i is a vector of normally distributed residuals with covariance matrix Ψ (i.e. the continuous latent variable, built form, in our model is assumed to have a Gaussian distribution) .

Equations (1) and (2) imply that:

$$\Upsilon_i = \nu + \Lambda(\mathrm{I} - \mathrm{B})^{-1}\alpha + \Lambda(\mathrm{I} - \mathrm{B})^{-1}\Gamma \mathrm{X}_i + \mathrm{K}\mathrm{X}_i + \Lambda(\mathrm{I} - \mathrm{B})^{-1}\zeta_i + \epsilon_i \qquad (3)$$

Where the jth element in vector Υ_i (i.e. the jth observed dependent variable), Υ_{ij}, is a normally distributed continuous variable (such as travel distance or time in our analysis), the residual variable ϵ_{ij} is assumed normally distributed. For categorical variable Υ_{ij} (e.g. the car ownership variable), a normality assumption for ϵ_{ij} is equivalent to the probit regression for Υ_{ij} on η_{ij} and X_{ij}. For count data (i.e. trip frequency in our model), the residual, ϵ_{ij}, is assumed to be zero and the dependent variable's link function is in an exponential form.

The model estimates by maximum likelihood estimator using the EM algorithm[12] where the latent variable η_i is treated as missing data. The observed-data likelihood is given by:

$$\prod_i \int f_i(Y_i)\, \phi_i(\eta_i) d_{\eta_i} \qquad (4)$$

where f_i and ϕ_i are likelihood functions respectively for Y_i and η_i.

Numerical integration is used which approximate the likelihood function by

$$\prod_i \sum_r \Pr(\eta_i = \eta_{ri})\, f_i(Y_i) \qquad (5)$$

where η_{ri} are the nodes of numerical integration.

The expected log likelihood can then be given by equation (6) below which should be maximized with respect to model parameters.

$$\sum_{ri} \log(\Pr(\eta_i = \eta_{ri})) + \sum_{ri} \log(f_i(Y_i)) \qquad (6)$$

[12] As implemented in Mplus software (https://www.statmodel.com, accessed 15 May 2015) and explained in Muthen and Asparouhov (2007).

In order to avoid being trapped in a local likelihood maxima, we use many different sets of starting values in the interactive maximization procedure and ensure that the maximized value of the likelihood function is replicated.

As we model the travel of employed adults over a week, we have not encountered data samples with zero trips (this has been subsequently verified using the censored zero inflated model and multinomial regression model tests). However, the modelling methodology can be used where some individuals in the sample make zero trips.

Because the NTS is a very large dataset, we consider the coefficients to be statistically significant only when the estimated coefficients are ≥ a 99% confidence interval (i.e. the respective p-values are $\leq 1\%$).

3.3.2. The SEM for trip frequency analyses

Following a general explanation in section 3.3.1, we explain how the log likelihood function is defined for modelling trip frequency, which is the most complex of the SEM models.

Figure 3 provides a simplified version of our model for trip frequency where Xi, η1i, Y1i, Y2i are respectively representing socioeconomic characteristics, the built form latent variable, trip frequency and car ownership for one trip purpose. All notations are for variable j and individual i.

For simplicity, we assume one travel purposes and three indicators for the latent variable.

$$Y^{*}_{1i} = \upsilon_1 + \Lambda_2\eta_{1i} + K_1X_i + K_3Y_{2i} \tag{7}$$

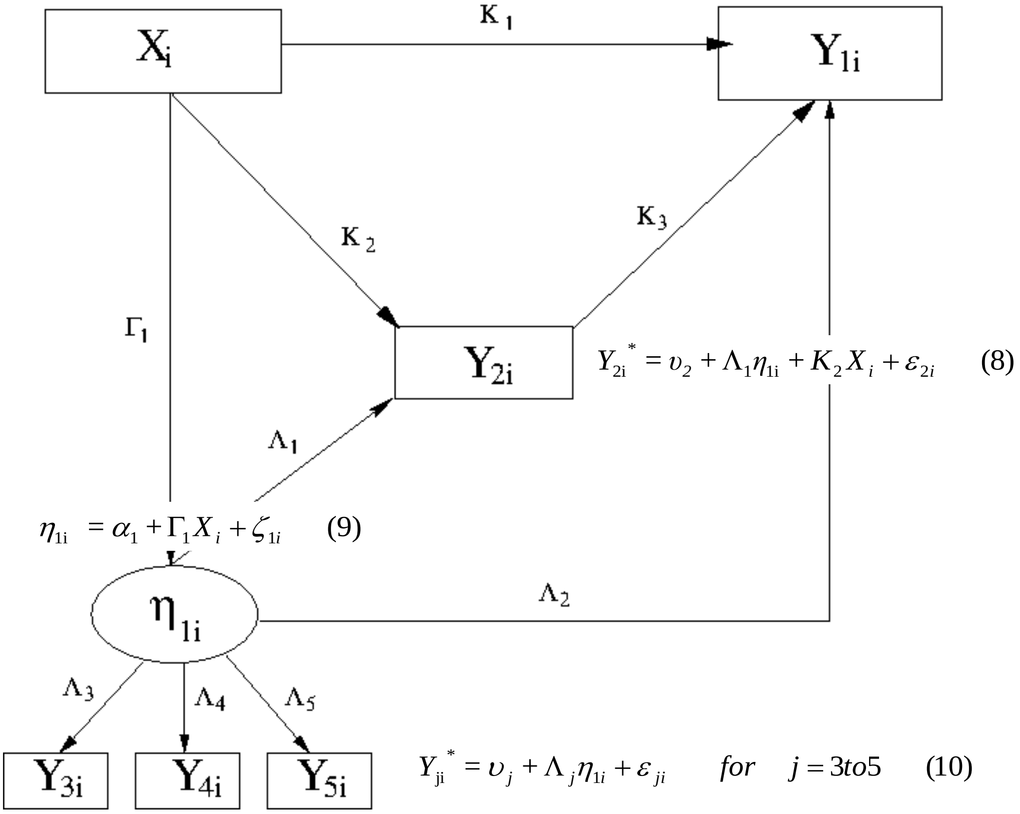

Figure 3 The SEM structure for trip frequency, one trip purpose only (Y1)

Equations (7) and (8) show the regression functions for trip frequency and car ownership respectively. Equation (9) shows the influences on built form latent variable, η_{1i} and equation (10) shows the regression function for the categorical indicators of built form latent variable.

In our analysis, the indicators of built form latent variable (Y_{3i} to Y_{5i}) are ordered categorical. Therefore, we can formulate the link function for variable j by defining an underlying continuous variable Y^{*}_{ji} such that

$$Y_{ji} = s \Leftrightarrow \tau_{j,s} < Y^{*}_{ji} < \tau_{j,s+1} \tag{11}$$

where $\tau_{j,s}$ are threshold parameters.

The normal distribution assumption for ε_{ji} in equation (10) is equivalent to a probit regression for each categorical variable $Y_{ji\,(\,j=3\ to\ 5)}$ on η_{1i} , with the following probability function:

$$f_{\mathrm{i(Y_{ji})}} = \Pr(Y_{ji} = s) = \Phi\left[\left(\tau_{j,s+1} - \upsilon_j - \Lambda_j\,\eta_{1i}\right)\right] - \Phi\left[\left(\tau_{j,s} - \upsilon_j - \Lambda_j\,\eta_{1i}\right)\right] \tag{12}$$

Car ownership, Y_{2i} , is a binary variable with probit distribution; given that $\varepsilon_2 \sim N(0,1)$, we can parameterize its function:

$$Y_{2\mathrm{i}} = \begin{cases} 1 & Y_{2i}^{*} > 0 \quad \textit{i.e.} \quad -\varepsilon_{2i} < \upsilon_2 + \Lambda_1\eta_{1\mathrm{i}} + K_2 X_i \\ 0 & \textit{Otherwise} \end{cases} \tag{13}$$

The likelihood function of Y_{2i} can then be written as

$$f_i(Y_{2i}) = \Phi^{Y_{2i}}(\upsilon_2 + \Lambda_1\eta_{1\mathrm{i}} + K_2 X_i)(1 - \Phi(\upsilon_2 + \Lambda_1\eta_{1\mathrm{i}} + K_2 X_i))^{1-Y_{2i}} \tag{14}$$

The trip frequency variable Y_{1i} in equation (7) is modelled as a count variable, with negative binomial distribution and link function $Y_{1i}^{*} = \ln(\mathrm{Y_{1i}})$ for $Y_{1i} \neq 0$. The likelihood function can then be formulated as:

$$f_i(Y_{1i}) = \exp\{Y_{1i}\ln(\frac{\alpha\mu_i}{1+\alpha\mu_i}) - \frac{1}{\alpha}\ln(1+\alpha\mu_i) + \ln\Gamma(Y_{1i} + \frac{1}{\alpha}) - \ln\Gamma(Y_{1i}+1) - \ln\Gamma(\frac{1}{\alpha})\} \tag{15}$$

where α is an overdispersion factor and μ_i is the expected value of Y_{1i}.

Based on equation (7):

$$\mu_{\mathrm{i}} = \exp\{\upsilon_1 + \Lambda_2\eta_{1\mathrm{i}} + K_1 X_i + K_3 Y_{2i}\} \tag{16}$$

Finally, the likelihood function, marginalizing over the latent variable, is given by:

$$\prod_i \int f_i(\mathrm{Y}_{1i}) f_i(\mathrm{Y}_{2i}) \prod_{j=3}^{5} f_i(\mathrm{Y}_{ji}) \Psi_i(\eta_{1i}) \mathrm{d}_{\eta_{1\mathrm{i}}} \tag{17}$$

Where $f_i(Y_{1i})$, $f_i(Y_{2i})$, and $f_i(Y_{ji\ for\ j=3\ to\ 5})$ are defined in equations (15), (14) and (12) respectively. $\Psi_{\mathrm{i}(\eta_{1\mathrm{i}})}$ is the likelihood function of normally distributed η_{1i}.

The numerical maximization of the above function is implemented as:

$$\prod_i \int f_i(\mathrm{Y}_{1i}) f_i(\mathrm{Y}_{2i}) \prod_{j=3}^{5} f_i(\mathrm{Y}_{ji}) \Psi_i(\eta_{1i}) \mathrm{d}_{\eta_{1\mathrm{i}}} \approx \prod_i \sum_r \Pr(\eta_{1i} = \eta^{r}{}_{1i}) \prod_{j=1}^{5} f_i(\mathrm{Y}_{ji}) \tag{18}$$

Where η_{1i}^{r} is the rth node of the numerical integration.

The EM algorithm is used for model estimation as follows:

First, the posterior distribution for the latent variable is computed for η_{1i}:

$$pr(\eta_{1i} = \eta_{1i}^{r} | *) = \frac{\Pr(\eta_{1i}=\eta_{1i}^{r}) \prod_{ij} f_i(Y_{ji})}{\sum_r \Pr(\eta_{1i}=\eta_{1i}^{r}) \prod_{ij} f_i(Y_{ji})} \quad (19)$$

then the expected complete-data log-likelihood is maximised as:

$$\sum_{r,i} \log(\Pr(\eta_{1i} = \eta_{1i}^{r})) + \sum_{r,i,j} \log(f_i(Y_{ji})) \quad (20)$$

Equation (20) is maximized w.r.t the model parameters. This process continues iteratively till convergence.

3.4. LCA-SEM Method[13]

LCA-SEM is an expansion to the SEM formula presented above. LCA-SEM employs conditional Latent Categorical Analysis (LCA) where we model built form as a categorical latent variable with socio-demographic characteristics of residents as controlling covariates. Identifying tangible built form typology which can be closely associated with NTS built form characteristics improves our understanding of variation in travel behaviour across built form clusters.

Latent class analysis (LCA) involves a set of observed variables which are called indicators (i.e. in our case Area Type, Population Density, and accessibility to public transport stations). The indicators form the basis for estimating latent variables such as the Built form latent variable in Figure 2. The LCA approach shares the same conceptual aim as Explanatory Factor Analysis (EFA): Both LCA and EFA are to construct latent variables from observed indicators. However, the estimated latent variable is continuous for EFA and discrete (or categorical) for LCA - LCA gives rise to a latent class model because the latent variable is discrete; latent class is characterized by a pattern of conditional probabilities that indicate the chance that the variables take on specific values. When it comes to interpretation of results, EFA focuses on grouping contributing variables (such as the contribution of built form area type, density and

[13] This section is based on Jahanshahi and Jin (2015)

public transport access), and can be considered as a variable-centered approach. By contrast, LCA focuses on grouping survey respondents or cases facing distinct patterns of the contributing variables into classes, and is thus a respondent-centered approach (Wang and Chen, 2012).

The statistical estimations are carried out in two stages: Firstly, we use Conditional Latent Categorical Analysis (LCA) to cluster individuals who reside in similar geographical location by estimating simultaneously individuals' built form class membership and their socioeconomic background; Secondly, the SEM is used to account for the inter-correlations among the residents' socioeconomic characteristics, their car ownership status, and the interactions among different journey purposes in the quantification of the direct and indirect influences on the amount of travel carried out for each journey purpose. The second stage estimation is performed conditional on the class membership which is estimated in the first.

To formulate the first stage, let Y_{ij} be the jth indicator variable (i.e. population density, area type etc) of the built form latent categorical variable, C_i for individual i. As all our indicators are ordered categorical variables, we can formulate the link function by defining an underlying continuous variable, ${Y^*}_{ij}$ such that

$$Y_{ij} = s|C_i = c \Leftrightarrow \tau_{cj,s} < {Y^*}_{ij} < \tau_{cj,s+1}) \qquad (21)$$

where C_i, our latent categorical variable (i.e. built form), takes values 1,…,c and $\tau_{cj,s}$ are a set of threshold parameters.

Conditional on regressors X (e.g. our socioeconomic characteristics) we can then present the link function as:

$${Y^*}_{ij}|_{C_i=k,x_i} = \nu_{kj} + K_{kj}X_i + \varepsilon_{ij} \qquad (22)$$

The normal distribution assumption for ε_{ij} is equivalent to a probit regression for categorical variable Y_{ij} on X_i with the following probability function:

$$\Pr\left(Y_{ij} = s\middle|c_i = k\right) = \Phi\left[\left(\tau_{kj,s+1} - \nu_{kj} - K_{kj}X_i\right)\right] - \Phi\left[\left(\tau_{kj,s} - \nu_{kj} - K_{kj}X_i\right)\right] \quad (23)$$

The class membership probability conditional on X is given by multinomial logistic regression with the following formula:

$$\Pr(C_i = k|X_i) = \frac{\exp(\alpha_k+\gamma_k X_i)}{\sum_{s=1}^{k}\exp(\alpha_s+\gamma_s X_i)} \tag{24}$$

The joint probability of indicators or observed data likelihood is then given by:

$$\Pr\left(Y_{i1} \dots Y_{iJ}\right) = \prod_i \sum_{k=1}^{c} \Pr(C_i = k) \prod_j \Pr\left(Y_{ij} = s \middle| c_i = k\right) \tag{25}$$

EM algorithm is then used for estimating the parameters and class membership where the latent variable C_i is treated as missing data. We first compute the posterior distribution for the latent variable. The posterior conditional joint distribution is calculated as:

$$\Pr(C_i = k| *) = \frac{\Pr(C_i=k)\prod_j \Pr\left(Y_{ij} = s \middle| c_i = k\right)}{\sum_{k=1}^{c}\Pr(C_i=k)\prod_j \Pr\left(Y_{ij} = s \middle| c_i = k\right)} \tag{26}$$

which is estimated given the parameters.

Given the class membership, model parameters are then estimated through maximising Equation (25). The model is solved iteratively until reaching convergence.

Equations (27) to (29) specify the structural equation model which is estimated within each latent class for the second stage of our modelling. The subscript for latent class membership is dropped here for simplicity.

$$\mathrm{Y}_{ij} = \nu_j + \mathrm{K}_j \mathrm{X}_{ij} + \epsilon_{ij} \tag{27}$$

where Y_{ij} refers to the ith respondent and jth vector of a dependant variable (e.g. travel distance for commuting to work) and X_{ij} is the vector of all individual level covariates. ν_j and K_j are the vectors of intercepts and the matrices of regression parameters correspondingly.

ϵ_{ij} is a vector of residuals with a mean of zero and covariance Θ. Where the jth observed dependent variable, Y_{ij}, is a normally distributed continuous variable (e.g. the distance travelled by journey purpose), the residual variable ϵ_{ij} is assumed normally distributed. For a dichotomous variable Y_{ij} (i.e. car ownership), a normality assumption for ϵ_{ij} is equivalent to the probit regression for Y_{ij} on X_{ij}[14]

The observed-data likelihood is given by:

$$\prod_{ij} f_{ij}(Y_{ij}) \tag{28}$$

where f_{ij} is the likelihood function for Y_{ij}.

The expected log likelihood is then maximized with respect to model parameter estimation:

$$\sum_{ij} \log(f_{ij}(Y_{ij})) \tag{29}$$

In order to avoid being trapped in a local maxima for log likelihood estimation, we use many different sets of starting values in the iterative maximization procedure. Mplus by default use 100 starting values and then select the best 10 to run to convergence. The software is then reported the maximum likelihood achieved for each converged run. In cases where the maximum likelihood was not replicated, we increase the number of runs to be converged to ensure that the maximized value of the likelihood function is replicated.

3.5. Two Level/Random Intercept SEM[15]

Two level Structural Equation Model is an expansion to the SEM framework shown in Figure 2 which adopts a single level SEM with a built form latent variable to control for self-selection and spatial sorting effects. In two level SEM, the single built form latent variable (of Section3.3) is replaced with random intercepts which can vary across built form categories for each regression equation of the SEM (cf Figure 4). The random intercepts provide more precise quantification of the share of self-selection and spatial sorting effects vis-à-vis built form characteristics' effects in explaining the influences on travel behaviour.

Figure 4 presents graphically the two level SEM with random intercepts. The upper path diagram postulates the overall model structure which consists of the model of within built form clusters together with random intercepts presented by eclipses. The within model is built on

[14] For more information on modelling categorical data in SEM and Mplus, see Muthen (1984)
[15] This section is based on Jahanshahi and Jin (2016).

SEM framework that we have settled upon a conceptual path diagram and consists of the following two types of explanatory variables: (1) a long list of demographic and socioeconomic characteristics as exogenous variables; (2) car ownership as a mediating, endogenous variable that is subject to influences from (1). This enables me to control for car ownership endogeneity. For dependent variables, similar to the SEM model of Section 3.3, we set up models for travel distance and travel time respectively[16]. Regression equations of car ownership, travel distance and travel time by purpose have random intercepts which vary across built form clusters. This between-level variation can be left unrestricted (we call this **Model A** in Section 5.4) or can be conditional on exogenous variables at built form cluster level as is shown in lower diagram in Figure 4 (we name this **Model B** in Section 5.4).

[16] For two-level analysis, we did not model the number of trips for two reasons: a) being a count variable, the complexity of the model would have increased to the state which would have been very difficult to get the model to converge b) SEM analysis shows that built form influences on trip frequency is relatively small

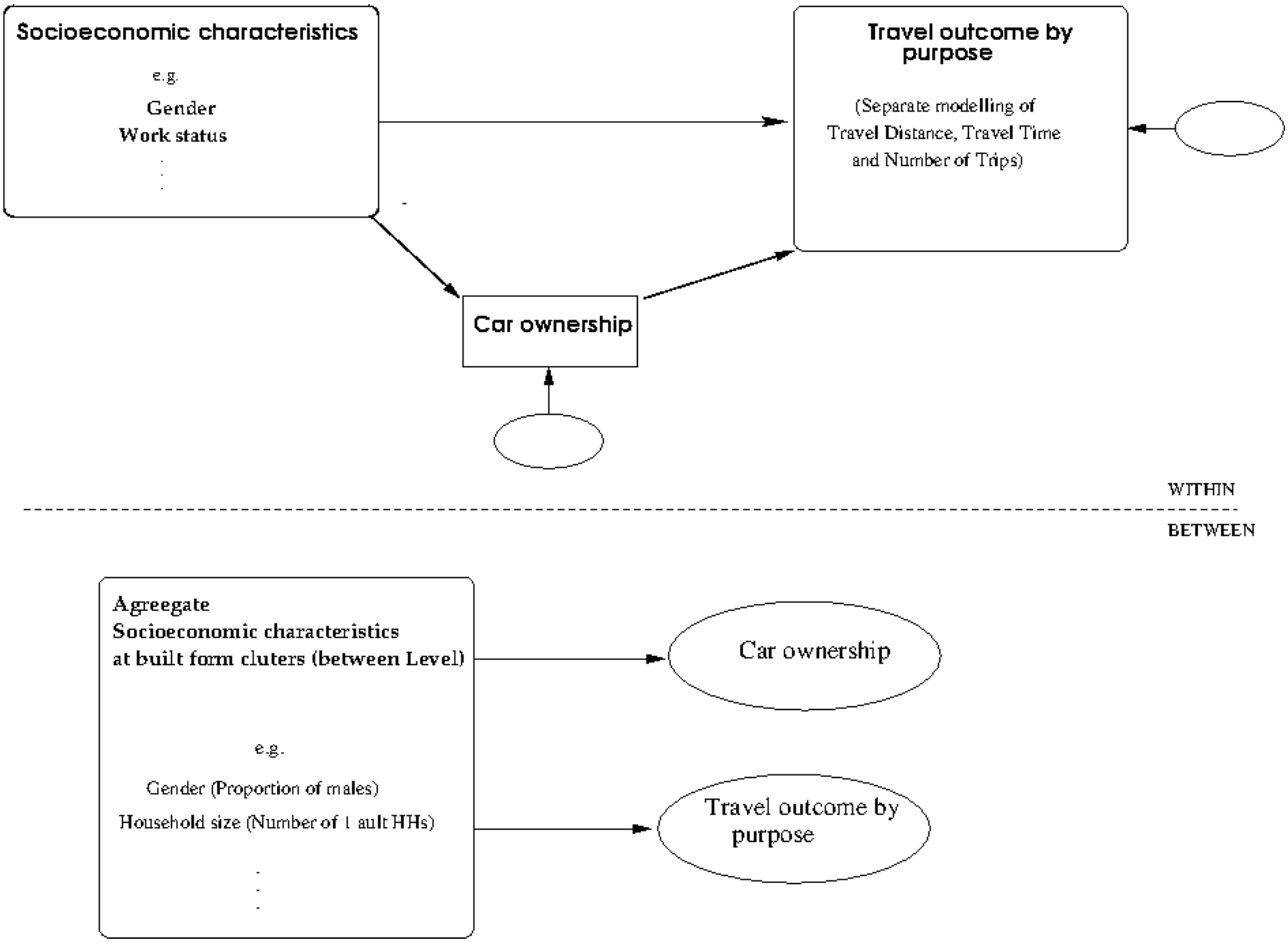

Figure 4 random intercept SEM

Note: The eclipses represent unobserved/latent variables.

The model notations are straightforward. Let Y_{pij} be the p-th dependent variable (i.e. car ownership status, travel distance, travel time by trip purpose) for individual i in built form cluster j. We proceed by defining an underlying normally distributed latent variable Y^*_{pij}. For travel distance and travel time by purposes which are normally distributed, we have $Y_{pij} = Y^*_{pij}$. The car ownership variable is a binary variable with probit distribution, we can parameterize its function as shown in equation (30).

$$y_{\mathrm{pij}} = \begin{cases} 1 & y^*_{pij} > 0 \\ 0 & Otherwise \end{cases} \tag{30}$$

The two level model can be constructed as:

$$y^*_{\mathrm{pij}} = y_{wpij} + y_{bpij} \tag{31}$$

where Y_{wpij} and Y_{bpij} are individual level (within level) and built form cluster level (between-level, i.e. random intercept) components of $Y^*{}_{pij}$ respectively. Both Y_{wpij} and Y_{bpij} are assumed to be normally distributed.

Equation 32 shows the model notation at individual level (within-level model)

$$Y_{wij} = BY_{wij} + \Gamma_w X_{wij} + \varepsilon_{wij} \tag{32}$$

where Y_{wij} and X_{wij} are the vector of within level dependent (Y_{wpij}) and independent variables (X_{wmij})-i.e. individual and household level socioeconomic variables in our model. The random intercepts (i.e. Y_{bpij}) can be modelled unrestricted which follow the notation in equation 33 (i.e. Model A) or as a function of between-level independent variables (i.e. Model B) which is shown in equation 34.

$$Y_{bij} = \alpha_b + \varepsilon_{bij} \tag{33}$$

$$Y_{bij} = \alpha_b + \Gamma_b X_{bij} + \varepsilon_{bij} \tag{34}$$

where Y_{bij} and X_{bij} are the vectors of model random intercepts (i.e. Y_{bpij}) and socioeconomic characteristics at built form clusters (i.e. X_{bnij}) respectively.

$B, \Gamma_w, \alpha_b, \Gamma_b$ are the vectors and matrices of slopes and regression parameters to be estimated. ε_{wij} and ε_{bij} are zero mean normally distributed independent vector variables. For identification purposes, the variance of the error term (ε_{wpij}) associated with car ownership binary variable is fixed to 1.

The model is estimated by a maximum likelihood estimator using an EM algorithm. Where the latent variable Y_{bij} (i.e. random intercept) is treated as missing data. The observed data likelihood is:

$$\mathrm{L} = \int \phi_{\mathrm{b}}(\mathrm{Y_b}) \prod_i \mathrm{f_{wi}}(\mathrm{Y_{wi}}) \mathrm{d}_{\mathrm{Y_b}} \tag{35}$$

Where $\mathrm{f}_{\mathrm{wi}}(Y_{wi})$ is the likelihood function of within level variable, being marginalized over between-level random intercept. Similar to the numerical integration approach explained in section 3.3.1, equation 36 provides the approximate estimation of likelihood function.

$$\hat{L} = \sum_{\mathrm{rb}} \Pr\left(\mathrm{Y_b} = \eta_{\mathrm{rb}}\right) \prod_i \mathrm{f}_{\mathrm{wi}}(Y_{wi}) \tag{36}$$

where η_{rb} is the rth node of the numerical integration for between-level cluster b.

In practice, there are two limitations in two level SEM. First, the degree of freedom of the between-level model is a function of the number of clusters (i.e. 98 built form categories here). This can limit the number of between-level parameters to be estimated. Second, the complexity of between-level model can increase the dimensions of the numerical integration resulting in slow convergence. Such considerations have been factored into model design by assuming fixed coefficients (except for intercepts) rather than more flexible model with random coefficients and by adding only the most significant between-level influences after testing various combinations.

We note here that apart from the methods explained above, a more general models can be developed by combining the underlined approaches. The limitation of time and resources has prevented me from undertaking that task - we will return to this in section 6.4 where we discuss rooms for further studies.

3.6. Prospects for continuous monitoring

As discussed in Section 3.1, one of the goals of this research is to establish a framework for monitoring changes in travel behaviour over time. While there may be scope to test this process to a limited extent for data currently available, there is ample rooms for conceptual and methodological developments – such as non-parametric latent class analysis etc - in future studies when more years of consistent data are available in NTS, with or without additional data from outside the NTS.

The NTS dataset has been enhanced progressively over years. With the current prospects for the continuation of this trend, with more new variables (e.g. attitudinal data) added and more geographically detailed information recorded and reported (such as the built form characteristics at the destination of trips), NTS can build up a substantial and richer time series for monitoring changes in travel behaviour.

In addition, the accelerating access to big data (e.g. data from mobile phones, social medias etc) and enhancement in the technology of analysing them will enable us to assemble systematic datasets such as NTS through data fusion techniques. This should in turn help improving the quality and accuracy of travel behaviour models and enlarge our ability to monitor changes in travel behaviour.

The model developed in this study can become the basis for continuous monitoring through setting up recurring runs every year or every couple of years to update the results and identify any major shifts. Moreover, this study can provide an insight into additional survey variables that may be required for enhancing and revising the existing models.

In addition, this study can highlight the room for technical improvement in further work, which could make the model more suitable for continuous monitoring shifts in travel behaviour. An example is to employ a nonparametric Bayesian approach such as Dirichlet process where the number of latent clusters can be infinite and flexible enough to change as more data become available. This range of machine learning techniques can help establishing a framework with automatic, self-improving analytics. This is further discussed in Section 6.4.

3.7. Summary

This chapter presents the main analytical methods to be used for investigating influences on travel patterns and its changes over time. In particular we have developed three sets of models in succession within the SEM framework: first, a path-diagram based SEM that models a web of interactive influences such as socioeconomic, built form and car ownership on the trip frequency, travel distance and travel time, with the built form variables being represented as a continuous latent variable; second, an LCA-SEM that replaces built form factor with a categorical latent variable allowing for identifying tangible built form clusters in the UK; third, a two level SEM that provides more detailed evaluation of the extent of self-selection and spatial sorting effects through allowing SEM intercepts to vary across built form clusters. The second and third set of models build on the initial one and provide more depth in the analyses of the influences of built form on travel outcomes.

In Chapter 4 we review the UK National Travel Survey dataset which is used in the dissertation to test the models developed above.

4. DATA[17]

4.1. Overview

The National Travel Survey (NTS) is a series of household surveys designed to provide regular, up-to-date data on personal travel and to monitor changes in travel behaviour over time. The survey is conducted as home interviews of all members of the sampled households, recording their personal and household characteristics and a detailed diary of their travel during one week in the year. The survey methodology has been continuously perfected over decades, recording the characteristics of the journeys made and a set of carefully selected personal, household and circumstantial variables that are believed to influence travel behaviour.

The first NTS was commissioned by the UK Ministry of Transport in 1965/66. Further periodic surveys were carried out in 1972/73, 1975/76, 1978/79 and 1985/86. Since July 1988 the NTS has been carried out as a continuous survey with field work being carried out in every month of the year and an annual set sample of over 5000 addresses. Substantial changes were made to the NTS organisation and method just before 2002 (Hayllar et al., 2005). Although it is still possible to include the earlier period from 1995 to 2001 for disaggregate level of analysis for some set of variables (e.g. Jahanshahi et al., 2009), a number of variables we used in our analysis would not be available in a form that was consistent between the two datasets. In addition, drops in response rates after introducing new variables and changes in survey conductors in 2002 might make comparison before and after 2002 less robust (refer to WPS, 2009 on a debate on response rates in NTS). For this study we therefore use the post 2002 NTS data.

The UK department for transport regularly publishes a technical report on the NTS (Morris et al., 2014) which includes detailed information such as sample selection and data collection, fieldwork procedures, response rates, and data processing. Here, we will only provide the most relevant description for the immediate purposes of the analyses.

The rest of this chapter is organised as follows: Section 4.2 provides the lists of tables and individual records of the NTS data as to be used for our analyses; Section 4.3 explains data

[17] This chapter is based on Jahanshahi et al (2015)

preparations; finally Section 4.4 provides the descriptive statistics as the backdrop to the model tests to be reported in Chapter 5.

4.2. NTS records used for analysis

The path-diagram based SEM and LCA-SEM are built upon 2002 to 2010 NTS data which was available at the time when those two sets of analyses where carried out. The two-level SEM, however, is developed later on when the records from the 2011 and 2012 surveys have become available. So the two-level SEM has made use of the data for 2002-2012. In addition, a limited set of model tests of the path-diagram based SEM and LCA-SEM has been run using the 2002-2012 data, so that the two-level SEM results can be compared like-for-like against the path-diagram based SEM. Not all the tests of the path-diagram based SEM and LCA-SEM have been re-run using the 2011 and 2012 data for lack of time. However, the tests done so far indicate that incorporating the 2011 and 2012 data is very unlikely to alter the results obtained on the 2002-2010 data.

As stated in Chapter 1, this dissertation has focussed on employed adults (i.e. 16 to 64 years old individuals who are economically active) as the most mobile sector of population which has also experienced the momentous changes in its composition in recent years (Handel, 2012). The potential influence of apparent shift in labour composition on travel patterns as well as the growing requirement to address accessibility and mobility inequality within this group has made the analysis of their travel behaviour of particular interest to both policy makers and transport modellers.

NTS data for 2002-2012 forms a consistent time series of eleven years. There are in total 1,137,259 trips and 9.9 million passenger miles travelled for commuting, shopping and other journeys by employed adults. For the same group of people for 2002-2010 there are in total 933,296 trips and 8.2 million passenger miles travelled.

The NTS data is organised in nested related tables (Morris et al., 2014) as shown in Table 1. In this dissertation, we used the first five tables, up to the journey level which are linked through associated journey, individual, household, and PSU (Primary Sampling Unit) identification codes. For instance, **journeys** refer to those for each **individual** within each **household** settled in specific **PSUs** for each survey year reported in the **Day** table.

Table 1 Main sets of data in NTS

Data table	Description / contents
Day	Travel survey year and day (1 to 7)
PSU.	Primary Sampling Unit - Variables specific to the post code sector unit in which household is located (e.g. area type and population density)
Household	Household related variables – e.g. numbers of resident adults, income etc
Individual	Individual related variables – e.g. gender, age etc
Journey	Variables specific to each journey made – e.g. purpose from, time started
Stage	Journey stage
Vehicle	Information on vehicles available to households surveyed.
Ldj	The long distance survey: entries cover a longer interval than the survey week

Source: NTS technical report (Morris et al., 2014)

Table 2 has listed the main sets of attributes for households, individuals and their trip-making which are used for all developed models. These variables are selected based on literature review, our previous work in analysing NTS data and large number of experiments for this study.

Table 2 NTS data: Definitions of variables selected for SEM analysis

Data for	Classifications of respective variables			
Households (from Households table)	Household size: 1 Adult; >1 adult.	Annual income: <£25,000; £25,000-49,999; ≥£50,000.	Household head occupation: Manual; Skilled manual; White collar clerical; Professional.	Car ownership: No access to car; Own or access to one or more than one car.
Employed adults (from individual table)	Gender: Male; Female.	Work status: Full time (FT); Part time (PT).		
Journeys (from journeys table)	Journey purpose (for outbound purpose): Home-based commuting(HBW); All shopping (Sh); All other purposes (Oth).	Trip frequency: Trips/week	Journey distance: Miles/trip	Journey time: Minutes/trip
Access to transport services at household location (from household table)	Maximum frequency of local bus services: Level 1: <1 bus per day; Level 2: at least 1 bus/ day; Level 3: at least 1/hour; Level 4: at least 1/half hr; Level 5: at least one every quarter hour;	Walk to bus stop (min): Level 1: < 6 ; Level 2: 7-13; Level 3: 14-26 ; Level 4: 27-43 ; Level 5: >44.	Bus time to rail station (minutes): Level 1:no bus/quicker to walk; Level 2: <6 ; Level 3: 1-13; Level 4: 14-26 ; Level 5: 27-43 ; Level 6: >44 .	Rail station type: Level 1: frequent service all day; Level 2: frequent service rush hour only; Level 3: less frequent service.
Built form characteristics at household location (from post code unit level- PSU-table)	Area type: Rural areas; Urban areas <25,000 population; Urban areas 25,000-250,000 population; Urban areas >250,000 population; Metropolitan areas outside London; London.	Population density (persons/hectare): Level 1: <10; Level 2: 10-14.99; Level 3: 10-14.99; Level 4: 15-19.99; Level 5: 20-24.99; Level 6: 30-34.99; Level 7: 35-39.99; Level 8: 40-49.99; Level 9: 50-59.99; Level 10: ≥60.		

4.3. Data preparation

In this Section, first we provide a brief explanation of weighting strategy within NTS. We would then explain the detailed procedure for linking the tables and the use of data expansion weights.

4.3.1. Weights in NTS

The NTS makes use of weighting factors to help to offset differential response rates to the survey. A weighting strategy for the NTS was developed following a recommendation in the 2000 National Statistics Quality Review of the NTS. The NTS results for 2005 were based on weighted data for the first time. The weighting methodology was then applied to data back to 1995 so that all NTS figures for 1995 onwards which have recently been released are now based on weighted data. As well as adjusting for non-response through a household weight, the weighting strategy for the NTS also uses a trip weight to adjust:

a. For the drop-off in the number of trips recorded by respondents during the course of the travel week;
b. For uneven recording of short walks by day of the week; and
c. For the short-fall in reporting of long distance trips.

In order to benefit from the NTS weighting system, this study uses household weights to weigh regressions in SEM. This will enable me to use a more precise approach for estimating travel indicators in which the trips are weighted only by trip weights and the household weights are used instead to provide input weights to the weighted regression models.

4.3.2. Assembling the data

The initial data includes all eight tables and were received in SPSS format. The SPSS dataset contained both the data and the definitions (attributes) of the variables. NTS team within Department for Transport (DfT) have kindly provided some further information such as more detailed population density to be used for this study. We use the package "foreign" within "R" statistical software to convert the data and its associated attributes to csv and text files respectively.

The data were then imported into a local database both for checking and assembling the required data for analysis. The checking includes comparing data with the published totals reported in NTS technical document to ensure completeness of the data received. We then design and run a set of queries to assemble the required data for analysis. The main steps are listed below:

a) PSU, households, and individual tables are linked to construct a full table of all individuals. An identification code is defined from the combination of survey year, PSU id, household id, and individual id. In this step, all required attributes from aforementioned tables are selected including the household weights.
b) Three sets of tables are constructed from "Journeys" which contains weekly trip information: table of "Home Based Work" which includes the trips from "home" or "visiting friends" with the main purpose of "work"; table of "all shopping trips" which contains trips from all locations for the main purpose of shopping; and table of "all

other trips" which includes all the trips not in the first two tables. It also excludes all return trips. Number of trips, travel distance and travel time are weighted with trip weights at this stage.

c) At the final stage, trip tables aggregated at individual level from step (b) are linked with individuals' attribute records from step (a) based on the constructed unique individual identification code. The query is built in a way to include all records from individual table to ensure the inclusion of those who have not reported any trip for a particular purpose in a particular week (i.e. those with 0 trip). At the end, the individual records with associated PSU, household, individual and trip attributes are exported.

The average of travel indicators by market segments and individual attributes are then calculated to be used as a benchmark for analysis of findings. These details for 2002 to 2010 are provided in Section 4.4 below.

4.4. Descriptive statistics

Table 3 presents the headline averages of travel distance, travel time and trip frequency per week for all employed adults and for the reference traveller segment- i.e. those served as reference categories in all SEM regression equations. The averages serve as benchmarks for analysing the findings and will be discussed further in Chapter 5. Further descriptive analyses of the data by variables of interest, e.g. by market segment are provided in Table 4. Extensive descriptive analyses have been reported in NTS publications[18].

[18] See https://www.gov.uk/government/collections/national-travel-survey-statistics, accessed 14 May 2015

Table 3 Average travel time, distance and frequency per person per week: employed adults

Period	Home-based commuting			Home- and Non-home-based shopping			All other purposes			All Purposes		
	Distance (miles)	Time (minutes)	Frequency (trips)	Distance (miles)	Time (minutes)	Frequency (trips)	Distance (miles)	Time (minutes)	Frequency (trips)	Distance (miles)	Time (minutes)	Frequency (trips)
All employed adults												
2002- 2010	30.3	88.5	3.3	11.3	39.3	2.3	72.9	183.5	7.6	114.4	311.3	13.1
2002-2006	30.9	89.5	3.4	11.7	40.6	2.3	75.0	186.9	7.8	117.6	317.0	13.5
2008-2010	29.2	86.8	3.1	10.6	37.2	2.1	69.5	177.8	7.2	109.3	301.8	12.5
Difference 08-10 vs 02-06	-1.7	-2.7	-0.3	-1.0	-3.4	-0.2	-5.5	-9.1	-0.6	-8.3	-15.2	-1.1
% Difference 08-10 vs 02-06	-5.6%	-3.0%	-7.7%	-9.0%	-8.4%	-8.9%	-7.4%	-4.9%	-7.6%	-7.1%	-4.8%	-7.8%
Reference segment: female, part-time, clerical workers in households of >1 adult and income £25,000-49,999 living in urban areas of <25,000 population												
2002- 2010	9.8	39.9	2.3	12.4	44.0	2.9	61.2	188.3	10.1	83.4	272.3	15.4
2002-2006	9.8	39.8	2.4	13.1	45.5	3.0	58.7	184.9	10.4	81.6	270.2	15.7
2008-2010	9.9	40.2	2.3	11.2	41.5	2.8	65.4	194.0	9.7	86.5	275.8	14.9
Difference 08-10 vs 02-06	0.1	0.5	0.0	-1.9	-3.9	-0.2	6.7	9.0	-0.6	5.0	5.6	-0.8
% Difference 08-10 vs 02-06	1%	1%	-1%	-14%	-9%	-6%	11%	5%	-6%	6%	2%	-5%

Note: the data represents outbound travel by employed adults in an average 7-day week. It excludes any return trips since the return trips cannot be classified as precisely by travel purposes.

As it can be observed from Table 3, at the aggregate level, all travel indicators have declined in magnitude across all trip purposes when comparing the later period (2008-2010) against the earlier one (2002-2006). However, the level of change is not the same across population segments. For instance, the reference group have experienced an increase in their travel distance and time across all purposes except for shopping. This shows the requirement for more detailed analysis at disaggregate level. Only through disaggregate analysis with careful control for potential endogeneities, one can identify whether the aggregate level of changes is due to change in travel patterns at individual level or the shift in labour composition. This will be the subject of the next Chapter.

Table 4 Average travel time, distance and frequency per person per week: employed adults by segments

4a – segmented by Individual characteristics

		Home-based commuting				Home- and Non-home-based shopping				All other purposes				All Purposes		
Period		Distance (miles)	Time (minutes)	Frequency (trips)		Distance (miles)	Time (minutes)	Frequency (trips)		Distance (miles)	Time (minutes)	Frequency (trips)		Distance (miles)	Time (minutes)	Frequency (trips)
2002- 2010	Female	21	74	3		13	46	3		65	180	9		98	299	14
2002-2006	Female	21	74	3		13	48	3		65	180	9		99	302	14
2008-2010	Female	21	73	3		12	43	2		64	179	8		97	296	13
Difference 08-10 vs 02-06	Female	0	0	0		-1	-6	0		-1	-1	-1		-2	-6	-1
% Difference 08-10 vs 02-06	Female	0	0	0		-8	-13	0		-2	-1	-11		-2	-2	-7
2002- 2010	Male	39	102	4		10	33	2		80	185	7		128	320	13
2002-2006	Male	40	104	4		10	34	2		84	192	7		134	329	13
2008-2010	Male	37	99	3		9	32	2		74	177	7		120	308	12
Difference 08-10 vs 02-06	Male	-3	-4	0		-1	-2	0		-10	-15	0		-14	-21	-1
% Difference 08-10 vs 02-06	Male	-8	-4	0		-10	-6	0		-12	-8	0		-10	-6	-8
2002- 2010	FT	36	102	4		11	37	2		76	181	7		123	320	13
2002-2006	FT	37	104	4		11	38	2		79	185	7		128	327	13
2008-2010	FT	35	100	3		10	35	2		72	176	7		117	311	13
Difference 08-10 vs 02-06	FT	-2	-4	0		-1	-3	0		-7	-9	0		-11	-16	0
% Difference 08-10 vs 02-06	FT	-5	-4	0		-9	-8	0		-9	-5	0		-9	-5	0
2002- 2010	PT	11	46	2		13	46	3		62	187	10		86	280	14
2002-2006	PT	11	46	2		13	47	3		63	191	10		87	284	15
2008-2010	PT	11	47	2		12	44	2		61	183	9		85	275	13
Difference 08-10 vs 02-06	PT	0	1	0		0	-3	0		-2	-8	-1		-2	-9	-2
% Difference 08-10 vs 02-06	PT	0	2	0		0	-6	0		-3	-4	-10		-2	-3	-13

4b – segmented by household characteristics

		Home-based commuting				Home- and Non-home-based shopping				All other purposes				All Purposes		
Period		Distance (miles)	Time (minutes)	Frequency (trips)		Distance (miles)	Time (minutes)	Frequency (trips)		Distance (miles)	Time (minutes)	Frequency (trips)		Distance (miles)	Time (minutes)	Frequency (trips)
2002- 2010	1 adult	27	86	3		10	40	2		71	190	8		108	315	14
2002-2006	1 adult	27	87	3		10	40	2		74	194	8		111	321	14
2008-2010	1 adult	27	85	3		10	39	2		68	184	7		104	308	13
Difference 08-10 vs 02-06	1 adult	-1	-2	0		0	-1	0		-6	-10	-1		-7	-13	-1
% Difference 08-10 vs 02-06	1 adult	-4	-2	0		0	-3	0		-8	-5	-13		-6	-4	-7
2002- 2010	2+ adult	31	89	3		12	39	2		73	182	8		115	309	14
2002-2006	2+ adult	31	90	3		12	41	2		76	186	8		119	316	14
2008-2010	2+ adult	30	87	3		11	37	2		69	177	7		109	301	13
Difference 08-10 vs 02-06	2+ adult	-2	-3	0		-1	-4	0		-7	-9	-1		-10	-15	-1
% Difference 08-10 vs 02-06	2+ adult	-6	-3	0		-8	-10	0		-9	-5	-13		-8	-5	-7
2002- 2010	clerical	30	93	3		12	41	2		76	196	9		117	329	14
2002-2006	clerical	31	95	3		12	43	2		78	198	9		121	336	14
2008-2010	clerical	29	90	3		11	39	2		73	193	8		112	321	13
Difference 08-10 vs 02-06	clerical	-2	-5	0		-2	-4	0		-5	-5	-1		-9	-15	-1
% Difference 08-10 vs 02-06	clerical	-6	-5	0		-17	-9	0		-6	-3	-11		-7	-4	-7
2002- 2010	manual	24	85	4		10	38	2		47	139	6		81	263	12
2002-2006	manual	24	84	4		10	39	2		49	142	6		83	266	12
2008-2010	manual	24	87	4		9	36	2		45	135	5		78	259	11
Difference 08-10 vs 02-06	manual	0	3	0		-1	-3	0		-4	-7	-1		-5	-7	-1
% Difference 08-10 vs 02-06	manual	0	4	0		-10	-8	0		-8	-5	-17		-6	-3	-8
2002- 2010	Prof	36	94	3		12	39	2		97	220	9		145	354	14
2002-2006	Prof	37	96	3		13	41	2		101	228	9		151	365	15
2008-2010	Prof	35	92	3		11	37	2		91	211	8		138	340	13
Difference 08-10 vs 02-06	Prof	-2	-3	0		-1	-5	0		-10	-17	-1		-13	-25	-2
% Difference 08-10 vs 02-06	Prof	-5	-3	0		-8	-12	0		-10	-7	-11		-9	-7	-13
2002- 2010	skilled manual	27	77	3		11	37	2		56	148	7		93	262	12
2002-2006	skilled manual	28	79	3		11	37	2		59	154	7		97	270	13
2008-2010	skilled manual	25	74	3		11	36	2		52	141	6		88	252	11
Difference 08-10 vs 02-06	skilled manual	-2	-5	0		0	-1	0		-7	-13	-1		-9	-18	-2
% Difference 08-10 vs 02-06	skilled manual	-7	-6	0		0	-3	0		-12	-8	-14		-9	-7	-15
2002- 2010	25k to 49.9k	30	86	3		12	39	2		69	175	8		110	300	14

		Home-based commuting				Home- and Non-home-based shopping				All other purposes				All Purposes		
Period		Distance (miles)	Time (minutes)	Frequency (trips)		Distance (miles)	Time (minutes)	Frequency (trips)		Distance (miles)	Time (minutes)	Frequency (trips)		Distance (miles)	Time (minutes)	Frequency (trips)
2002-2006	25k to 49.9k	31	89	3		12	40	2		74	183	8		117	312	14
2008-2010	25k to 49.9k	28	83	3		11	37	2		62	166	7		101	285	13
Difference 08-10 vs 02-06	25k to 49.9k	-4	-6	0		-1	-4	0		-12	-17	-1		-16	-27	-1
% Difference 08-10 vs 02-06	25k to 49.9k	-13	-7	0		-8	-10	0		-16	-9	-13		-14	-9	-7
2002- 2010	50k and over	39	103	3		12	40	2		97	221	9		148	364	14
2002-2006	50k and over	40	105	3		13	43	2		103	230	9		156	378	15
2008-2010	50k and over	37	100	3		11	37	2		89	209	8		137	346	13
Difference 08-10 vs 02-06	50k and over	-4	-5	0		-2	-6	0		-14	-21	-1		-19	-32	-2
% Difference 08-10 vs 02-06	50k and over	-10	-5	0		-15	-14	0		-14	-9	-11		-12	-8	-13
2002- 2010	Less than 25k	21	75	3		10	39	2		52	151	7		82	266	13
2002-2006	Less than 25k	22	77	3		10	39	2		53	154	7		84	270	13
2008-2010	Less than 25k	20	73	3		10	39	2		50	148	6		80	260	12
Difference 08-10 vs 02-06	Less than 25k	-2	-4	0		0	0	0		-3	-6	-1		-4	-10	-1
% Difference 08-10 vs 02-06	Less than 25k	-9	-5	0		0	0	0		-6	-4	-14		-5	-4	-8
2002- 2010	1 car	27	86	3		11	39	2		65	173	8		101	298	14
2002-2006	1 car	27	87	3		11	40	2		67	177	8		105	304	14
2008-2010	1 car	26	85	3		10	38	2		62	168	7		97	291	13
Difference 08-10 vs 02-06	1 car	-1	-1	0		-1	-3	0		-5	-9	-1		-8	-13	-1
% Difference 08-10 vs 02-06	1 car	-4	-1	0		-9	-8	0		-7	-5	-13		-8	-4	-7
2002- 2010	2+ Car	34	83	3		13	39	2		85	198	9		132	319	14
2002-2006	2+ Car	35	85	3		13	41	2		89	203	9		137	329	14
2008-2010	2+ Car	33	80	3		12	36	2		81	191	8		125	307	13
Difference 08-10 vs 02-06	2+ Car	-3	-5	0		-1	-5	0		-8	-12	-1		-12	-22	-1
% Difference 08-10 vs 02-06	2+ Car	-9	-6	0		-8	-12	0		-9	-6	-11		-9	-7	-7
2002- 2010	No Car	22	128	4		6	41	2		35	142	5		63	312	12
2002-2006	No Car	22	126	4		6	41	2		35	140	5		64	308	12
2008-2010	No Car	23	130	4		5	42	2		35	144	5		62	317	11
Difference 08-10 vs 02-06	No Car	0	4	0		0	1	0		0	4	0		-2	9	-1
% Difference 08-10 vs 02-06	No Car	0	3	0		0	2	0		0	3	0		-3	3	-8

4c – Segmented by area types

Period		Home-based commuting				Home- and Non-home-based shopping				All other purposes				All Purposes		
		Distance (miles)	Time (minutes)	Frequency (trips)		Distance (miles)	Time (minutes)	Frequency (trips)		Distance (miles)	Time (minutes)	Frequency (trips)		Distance (miles)	Time (minutes)	Frequency (trips)
2002- 2010	Lon	31	147	4		7	39	2		58	197	7		95	384	12
2002-2006	Lon	33	151	4		8	43	2		62	207	7		103	401	13
2008-2010	Lon	29	143	4		6	35	2		52	185	6		86	363	11
Difference 08-10 vs 02-06	Lon	-5	-8	0		-2	-8	0		-10	-22	-1		-17	-38	-2
% Difference 08-10 vs 02-06	Lon	-15	-5	0		-25	-19	0		-16	-11	-14		-17	-9	-15
2002- 2010	Metropolitan	26	90	4		10	40	2		66	182	8		101	312	13
2002-2006	Metropolitan	26	91	4		10	41	2		69	188	8		105	320	13
2008-2010	Metropolitan	26	89	3		9	38	2		62	174	8		97	301	14
Difference 08-10 vs 02-06	Metropolitan	-1	-3	0		-1	-3	0		-7	-14	0		-8	-19	1
% Difference 08-10 vs 02-06	Metropolitan	-4	-3	0		-10	-7	0		-10	-7	0		-8	-6	8
2002- 2010	Rural	32	70	3		15	39	2		85	184	7		132	292	12
2002-2006	Rural	33	71	3		15	40	2		86	187	7		134	298	12
2008-2010	Rural	31	68	3		15	37	2		84	181	7		130	285	11
Difference 08-10 vs 02-06	Rural	-1	-3	0		-1	-3	0		-2	-6	0		-4	-13	-1
% Difference 08-10 vs 02-06	Rural	-3	-4	0		-7	-8	0		-2	-3	0		-3	-4	-8
2002- 2010	small urban 10k-25k	30	73	3		13	38	2		72	171	8		115	282	13
2002-2006	small urban 10k-25k	30	74	3		13	39	2		74	177	8		118	290	14
2008-2010	small urban 10k-25k	30	71	3		12	37	2		70	164	7		112	272	12
Difference 08-10 vs 02-06	small urban 10k-25k	0	-3	0		-1	-2	0		-4	-13	-1		-6	-18	-2
% Difference 08-10 vs 02-06	small urban 10k-25k	0	-4	0		-8	-5	0		-5	-7	-13		-5	-6	-14
2002- 2010	Urban over 250K	30	91	4		10	41	3		74	192	8		114	324	14
2002-2006	Urban over 250K	31	91	4		11	43	3		78	198	8		119	332	14
2008-2010	Urban over 250K	29	92	3		9	39	2		70	185	8		108	315	14
Difference 08-10 vs 02-06	Urban over 250K	-2	0	0		-2	-4	0		-8	-13	0		-11	-17	0
% Difference 08-10 vs 02-06	Urban over 250K	-6	0	0		-18	-9	0		-10	-7	0		-9	-5	0
2002- 2010	Urban over 25K to 250K	30	81	4		11	38	2		71	175	7		113	294	13
2002-2006	Urban over 25K to 250K	31	83	4		11	39	2		73	176	7		116	298	13
2008-2010	Urban over 25K to 250K	29	79	3		11	37	2		69	174	8		109	290	13
Difference 08-10 vs 02-06	Urban over 25K to 250K	-2	-4	0		-1	-2	0		-4	-2	1		-7	-8	0
% Difference 08-10 vs 02-06	Urban over 25K to 250K	-6	-5	0		-9	-5	0		-5	-1	14		-6	-3	0

The average values of travel outcomes by type of individuals, households and built form characteristics that are provided in Table 4 will form the basis for comparison with the results of SEM analysis. We will refer back to this table in Chapter 5.

5. Analyses of SEM model results

Chapter 3 and 4 present respectively the method of study including the conceptual framework, and the variables of interest within NTS dataset. This chapter provides the findings of analyses. The outline of this Chapter is as follow: Section 5.1 tests the effect of structural ambiguity by looking into the influence of changing structural framework on built form influences. Section 5.2 presents the findings from the path-diagram based SEM where the influences on travel distance, travel time and number of trips are evaluated after controlling for potential self-selection, the endogeneity of car ownership, and the interactions among travel purposes; Section 5.3 reports the findings from the LCA-SEM which models built form as a categorical latent variable - this allows the identification of tangible built form categories and the evaluation of the variations in travel behaviour across them. Section 5.4 highlights the findings from a two-level SEM- this allows model intercepts vary across the built form classes to present a more precise approach for measuring the extent of residual self-selection effects.

For a relatively short time series, we examine the changes in influences over time; first the variations in influences from 2002 to 2010[19]; then those in two equally divided periods: before and after 2007, 2007 being the year when the world financial crisis started to set in. For path-diagram based SEM, we employ multi-group analysis within SEM framework to compare the estimated influences from 2002 to 2006 with that from 2008 to 2010. Wald test is used to examine whether the variations in influences is statistically significant. We used longer period of NTS data, which includes the years 2011 and 2012, as they become available for random intercept model to examine how the influences within- and between-built form clusters have evolved over time. Unfortunately the data was not available at the time we were developing path-diagram based SEM and LCA-SEM.

Because the NTS is a very large dataset, in reporting the results we consider the coefficients to be statistically significant only when the estimated coefficients have a ≥99% confidence interval (i.e. the respective p-values are ≤1%). The exception is between-level model of the two-level SEMs where the degree of freedom is limited to the number of built form clusters.

[19] The examination of 2002 to 2010 data (refer to Section 5.2.4) suggests that the unstable influences which shows systematic changes over time mainly experience shift in the trend after 2006. Consequently, we decided to compare the influences between two model chunks: 2002 to 2006 and 2008 to 2010/12. Grouping years increase the sample size and helps overcoming potential biases in analysing trend of changes over a limited time period.

5.1. Examining the influence of structural ambiguity and reverse causality

As explained in Chapter 2 and 3, the major SEM drawback is its structural ambiguity. SEM structural path overcomes issues such as multicollinearity and endogeneity in regressions but requires assumptions on the dependencies of variables. These assumptions are mainly made based on the theories from literatures but sometimes are required to be tested specifically when the reverse causality is suspected.

In our case, the major structural ambiguity rests in the direction of the dependency between car ownership and travel distances; whether it is car ownership defining distance of travel as one of the travel outcomes (refer to Figure 2) or travel distance as a measure of accessibility, in conjunction with built form characteristics, affecting car ownership (refer to Figure 5 below). In this dissertation we adopted the former structure as in theory travel distance is mainly the function of people decision on where to live, work, do shopping, visiting friends, etc; these decisions are more fundamental and less sensitive to car ownership than the other way around.

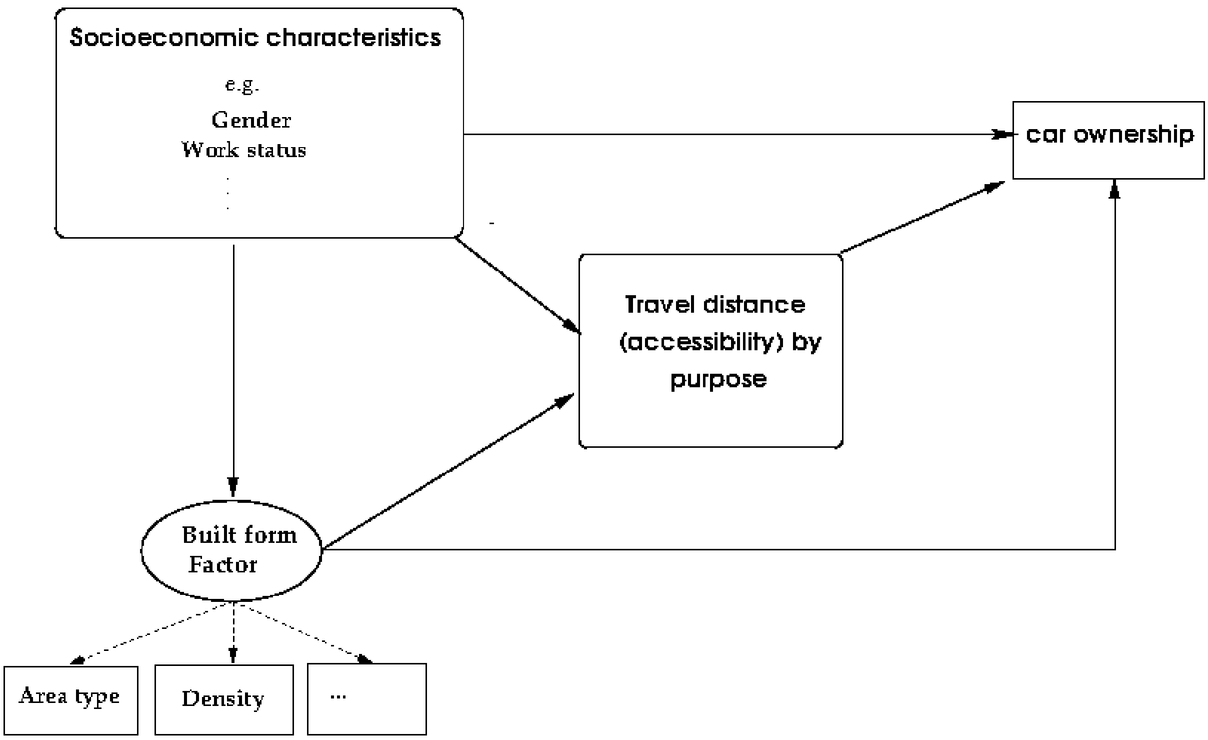

Figure 5 The alternative structure to test reverse causality

The comparison of the two model structures could provide practical evidence on the stronger direction of the effect. Moreover, it is beneficial to compare the influences of built form in the two alternative model structures. Considering the main goal of this study in determining built form influences, this helps evaluating the potential effect of structural ambiguity on our findings.

Both structures are modelled by using path-diagram based SEM. More detailed discussion on the specification of built form latent variable, and the significance and inference of the influences from path-diagram based SEM will be provided in Section 5.2. Here, we use the same covariates determined as important in Section 5.2 to compare the alternative model structures. Table 5 below provides the comparison of model coefficients which are called ST1 and ST2 respectively. The numbers in parentheses are the P-Values.

Table 5 The comparison of model structures

Direct Effect		ST1- Car Ownership affects Travel distance (Figure 2)	ST2 - Travel distance influences car ownership (Figure 5)
Built form latent variable Measured BY			
	Area Type	1.979 (0.000)	1.97 (0.000)
	Bus Frequency	1.095 (0.000)	1.094 (0.000)
	Population Density	1.76 (0.000)	1.757 (0.000)
Built form latent variable regressed ON			
	Male	-0.013 (0.076)	-0.013 (0.087)
	Full time working	0.102 (0.000)	0.104 (0.000)
	1 adult households	0.202 (0.000)	0.203 (0.000)
	Manual workers	-0.05 (0.010)	-0.048 (0.014)
	Skilled manual workers	-0.199 (0.000)	-0.197 (0.000)
	Professionals	-0.183 (0.000)	-0.182 (0.000)
	Household income less £25k	0.028 (0.059)	0.03 (0.046)
	Household income more than £50k	0.067 (0.000)	0.069 (0.000)
No car in household regressed ON			
	Male	-0.012 (0.436)	-0.012 (0.459)
	Full time working	-0.006 (0.797)	-0.007 (0.755)
	1 adult households	0.533 (0.000)	0.585 (0.000)
	Manual workers	0.413 (0.000)	0.356 (0.000)
	Skilled manual workers	-0.284 (0.000)	-0.329 (0.000)
	Professionals	-0.298 (0.000)	-0.277 (0.000)
	Household income less £25k	0.541 (0.000)	0.505 (0.000)
	Household income more than £50k	-0.235 (0.000)	-0.201 (0.000)

Direct Effect		ST1- Car Ownership affects Travel distance (Figure 2)	ST2 - Travel distance influences car ownership (Figure 5)
	Built form latent variable	0.609 (0.000)	0.556 (0.000)
	Commuting travel distance	N/A	-0.021 (0.000)
	Shopping travel distance	N/A	-0.138 (0.000)
	Other travel distance	N/A	-0.026 (0.000)
	Threshold	1.689 (0.000)	1.392 (0.000)
Commuting travel distance regressed ON			
	Male	10.65 (0.000)	10.66 (0.000)
	Full time working	16.87 (0.000)	16.9 (0.000)
	1 adult households	2.9 (0.000)	2.41 (0.000)
	Manual workers	-3.21 (0.000)	-3.54 (0.000)
	Skilled manual workers	-4.45 (0.000)	-4.29 (0.000)
	Professionals	2.6 (0.000)	2.72 (0.000)
	Household income less £25k	-4.37 (0.000)	-4.75 (0.000)
	Household income more than £50k	4.52 (0.000)	4.58 (0.000)
	Built form latent variable	-3.44 (0.000)	-3.74 (0.000)
	No car in household	-3.91 (0.000)	N/A
	Intercepts	10.5 (0.000)	10.3 (0.000)
Shopping travel distance regressed ON			
	Male	-3.13 (0.000)	-3.12 (0.000)
	Full time working	-0.91 (0.000)	-0.89 (0.000)
	1 adult households	0.71 (0.001)	0.33 (0.112)
	Manual workers	-1.46 (0.000)	-1.72 (0.000)
	Skilled manual workers	-1.17 (0.000)	-1.04 (0.000)
	Professionals	-0.09 (0.688)	0.01 (0.978)
	Household income less £25k	-0.63 (0.001)	-0.92 (0.000)
	Household income more than £50k	0.18 (0.392)	0.22 (0.293)
	Built form latent variable	-3.42 (0.000)	-3.67 (0.000)
	No car in household	-3.13 (0.000)	N/A
	Commuting travel distance	0.001 (0.691)	0.001 (0.497)
	intercepts	13.8 (0.000)	13.6 (0.000)
Other travel distance regressed ON			
	Male	14.95 (0.000)	15.03 (0.000)
	Full time working	2.53 (0.002)	2.66 (0.001)
	1 adult households	20.37 (0.000)	17.34 (0.000)
	Manual workers	-19.92 (0.000)	-21.89 (0.000)
	Skilled manual workers	-20.14 (0.000)	-19.1 (0.000)

Direct Effect		ST1- Car Ownership affects Travel distance (Figure 2)	ST2 - Travel distance influences car ownership (Figure 5)
	Professionals	13.56 (0.000)	14.31 (0.000)
	Household income less £25k	-10.3 (0.000)	-12.6 (0.000)
	Household income more than £50k	17.03 (0.000)	17.33 (0.000)
	Built form latent variable	-12.06 (0.000)	-13.86 (0.000)
	No car in household	-24.33 (0.000)	N/A
	Commuting travel distance	-0.138 (0.000)	-0.134 (0.000)
	Shopping travel distance	0.334 (0.000)	0.353 (0.000)
	intercepts	60.78 (0.000)	58.64 (0.000)

It is reassuring to observe that the direction and absolute values of almost all significant influences across the two models, specifically that of built form, are very similar. This can be due to the fact that built form latent variable is a good representative of the accessibility influences. This in essence diminishes the effects from travel distances reverse causality on car ownership. This suggests that the potential simultaneity biases from SEM structural ambiguity is minimal.

Comparing the influences of car ownership on travel distance (from ST1) with those in the reverse direction (from ST2) also reveals some interesting results. Table 6 shows the percentage change in travel distance when the reference group[20] forgo household car (second column)[21]; it compares that with the percentage change in the odds of having no access to household car for the equivalent change in the travel distances (third column). The latter is estimated by the ratio of the probability[22] of not having a household car for the reference group with mean Built form factor score and travel distances set to the values of intercepts[23] with the

[20] The coefficients of the reference variables are by definition zero. As shown in the right most column of Table 8, the reference segment of employed adults consists of part-time female workers of white collar clerical occupation living in middle income (£25-50,000), car owning households with more than one adults.
[21] This is calculated for each travel purposes by dividing the coefficient value of "No car in Household" (i.e. the marginal effect of having no car) by the intercepts. The intercept is the total travel distance by the reference group and the coefficient of "No car in Household" represents the difference between the travel distance of the reference group with them when they have no access to the household car (i.e. marginal effect). The land use latent variable is assumed to be at its mean (i.e. 0.018-refer to appendix B)
[22] As we have used probit regression to model car ownership, the probability is estimated by calculating the standard normal cumulative density of the sum of the descriptive variables multiplied by their respective coefficients minus 'No car in Household' threshold (i.e. $F(\sum \beta X - \tau)$).
[23] For estimating the Travel distance of the reference group in ST1 and car ownership probability for reference group in ST2 we assumed that all dummy variables have the coefficients of zero. We also assume that the Land

otherwise same group who additionally experience changes in their travel distances equivalent to the coefficient of the influences of "No Car in Household" on travel distances from ST1.

Table 6 The comparison of model structures

	Car Ownership influence on travel distance (i.e. absolute change in travel distance) - from ST1)	%age Changes in travel distance of the reference group when they have no access to household car (from ST1)	%age change in the probability of having no car of the reference group after the change in travel distance equivalent to those reported in column 1
Commuting	-3.91	-37%	1.8%
Shopping	-3.13	-23%	9.6%
Other Purposes	-24.33	-38%	14.4%

The comparison clearly shows that the effect of car ownership on travel distances is substantially stronger when compares to the reverse directional influence. For instance, non-car owner commute 3.91 miles or 37% shorter than the car owners (from ST1). However, having shorter travel distance to work by 3.91 miles would shows only 1.8% increase in the probability of having no car. This would justify the choice of ST1 as the basis for this research study.

5.2. Findings from path-diagram based SEM[24]

We have run a large number of SEM estimations using both Weighted Least Square (WLS) and Maximum Likelihood (ML) algorithms. We find that WLS and ML generally produce results of the same sign, magnitude and statistical significance, but the coefficient values do vary. Since WLS is more convenient to run, we tend to use it as a precursor for identifying significant variable interactions. ML tests are then carried out for more precise quantification of the effects. The results reported below are all ML results. The comparison with WLS results are provided in Appendix B where a graphical representation of results is also presented.

Section 5.2.1 provides a brief explanation of SEM specification and the explanatory factor analysis for constructing built form latent variable. An SEM test is characterised by its extensive range of outputs given the multiple interdependencies. We summarise the findings as (1) direct influences, (2) indirect influences and (3) results by year groups to examine the influences over time. These are reported in Sections 5.2.2, 5.2.3, and 5.2.4 respectively

Use latent variable has the value equivalent to its mean (see appendix B) and the continuous variables of travel distance to work, shopping and other purposes is at its intercept value estimated in the model.
[24] Based on Jahanshahi et al (2015)

5.2.1. SEM model specifications

The first step of the model specifications is to define the built form latent variable through an explanatory factor analysis (Albright and Park, 2009). From our previous NTS analysis we are aware of six NTS variables that are closely associated with built form in the UK: area type, population density, frequency of local buses, walk times to bus stop, bus times to rail station and rail station type. Moreover, the rate of missing values in some other collected accessibility variables such as walk time and distance to nearest doctor, post office, chemist, food store, and shopping centre is between 35% to 65% which makes them unfit for our analyses purposes. The high levels of inter-correlations among some of aforementioned built form variables make it impossible to treat them as independent explanatory variables. In addition, constructing fewer factors out of these variables makes it more feasible to control for built form endogeneity (i.e. self-selection and spatial sorting effects). The explanatory factor analysis (EFA) turns the inter-correlations into an advantage by investigating which variables are closely associated with one another, and therefore can contribute to a cluster to support a composite, latent variable that is better capable of representing the pattern of influences than any of the constituents. In other words, the EFA reveals to what extent the variabilities among input variables (i.e. factor indicators) are due to common factors.

The variables in EFA are modelled as ordered categorical (Table 2). Six factor indicators provide sufficient degree of freedom to fit maximum of two latent variables. Table 7 shows the EFA outputs; for the first latent variable, three indicators (area type, density and bus frequency) have a correlation coefficient (i.e. the varimax rotated factor loadings) greater than 0.7 which is a clear indication that they make a material contribution to the latent variable; for an alternative, second latent variable, only one factor loading (rail station type) reaches above 0.7. This suggests that “Built form latent variable” is best supported by area type, density and bus frequency, which is in line with our expectations. The distribution of the built form latent variable and its descriptive statistics is provided in Figure 12 in appendix B.

Table 7 Varimax rotated factor loadings for built form latent variable definition

	Factor 1	Factor 2
Area type	-0.879	-0.141
Population density	-0.825	-0.081
Frequency of local buses	-0.720	-0.143
Walk time to bus stop	0.263	0.063
Bus time to rail station	0.232	-0.155
Rail station type	0.190	0.766

Having specified the NTS variables, we use the conceptual model in Section 3.3 as a guide to experiment with the interaction links among variables in pilot SEM tests. We start by assuming extensive links and gradually thinning out the statistically insignificant ones.

Figure 6 presents the eventual SEM path diagram that we have settled upon. The path structures are the same for all three travel outcomes: on top left there are the demographic and socioeconomic characteristics of households and individuals; on the bottom left, a latent variable for built form as defined by the EFA. To the right there are three dependent variables by trip purpose, where the amount of commuting influences that of shopping, and both commuting and shopping influence other travel. In the middle is household car ownership as a mediating, endogenous variable. The arrows indicate the direction of the influences.

Similar to other regression models, for each categorical variable, one category is left out so that coefficient estimation can treat it as the reference category. In Figure 6 the reference categories are shown in parenthesis. For instance, for gender the reference category is 'Female'.

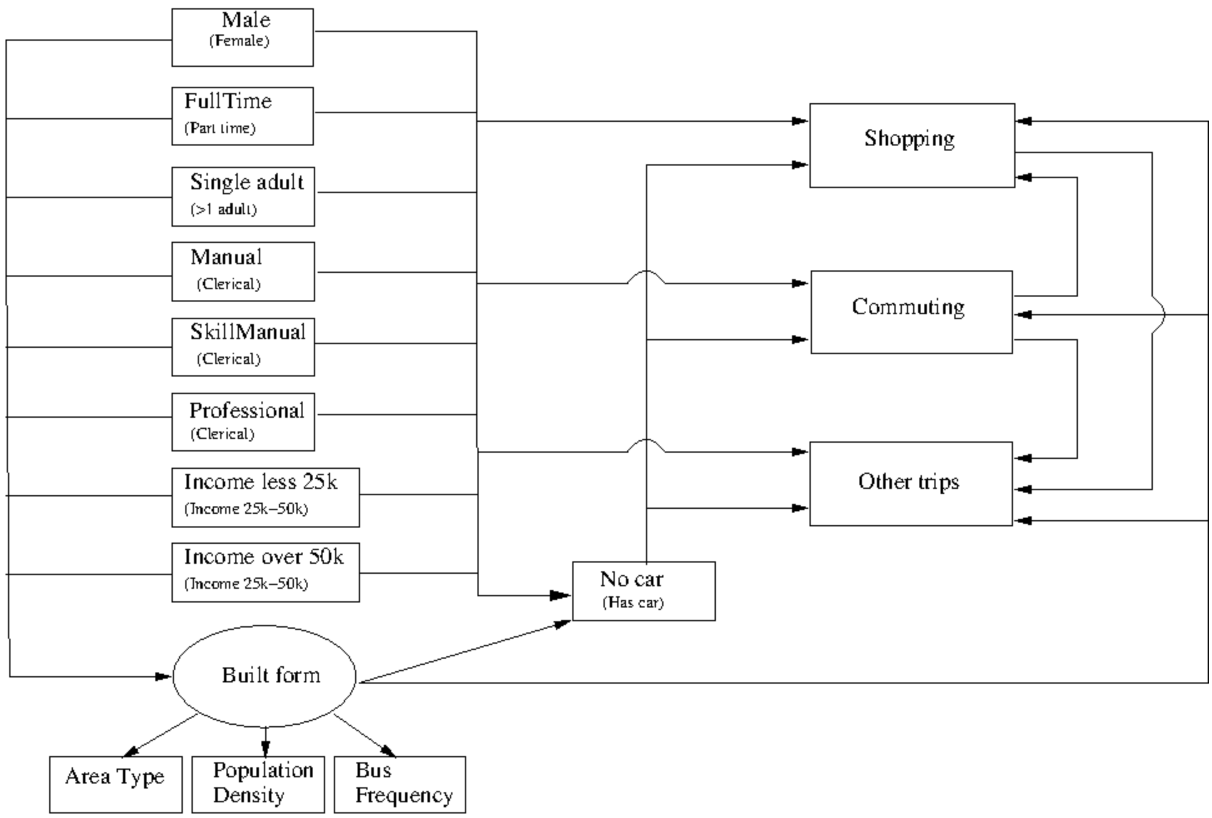

Figure 6 The main SEM path diagram adopted for the NTS data

5.2.2. Direct influences

Table 8 shows the direct influences of residents' socioeconomic profiles on two variables that condition travel choices, i.e. the built form latent variable and household car ownership. Reassuringly the direct influences of socioeconomic profiles are very similar across the models for travel distance, time and frequency.

More specifically, Panel 8a shows the influence of residents' socioeconomic profiles on built form characteristics of their residential location. The coefficients are all estimated relative to the reference variable, shown in the right most column. A coefficient that is negative indicates that the influence is for their residential location to rank lower than that of the reference segment. It shows that whilst the influence of gender is tiny, skilled manual workers and professionals tend to reside in considerably less dense and more rural areas.

Panel 8b shows significant influences of socioeconomic profiles upon car ownership. A large positive coefficient indicates a strong influence for not owning or having regular access to a car. In particular, after controlling for the modelled interdependencies, not only does the influence of built form on car ownership remain highly significant, it is also the strongest influence among all direct influences with the highest coefficients 0.605-0.607[25].

The coefficients presented in Table 8 are unit-free and we will return to them when quantifying indirect influences in Section 5.2.3.

Table 8 Direct influences on the built form latent variable and car ownership status

Direct effect	Travel time model	Travel distance model	Trip frequency model	Reference variable
Panel 8a. Influence of socioeconomic profile on the built form latent variable				
Male	-0.015**	-0.014*	-0.015**	Female
Full time working	0.097***	0.098***	0.097***	Part time working
1 adult households	0.2***	0.2***	0.199***	>1 adult households
Semi- or unskilled manual workers	-0.054***	-0.053***	-0.054***	White collar clerical
Skilled manual workers	-0.203***	-0.202***	-0.202***	White collar clerical
Professionals	-0.186***	-0.186***	-0.186***	White collar clerical
Household income less £25k	0.026*	0.027*	0.026*	Income 25-50k
Household income more than £50k	0.063***	0.063***	0.063***	Income 25-50k
Panel 8b. Influence of socioeconomic profile and built form on car ownership				
Male	-0.012	-0.012	-0.012	Female
Full time working	-0.002	-0.003	-0.002	Part time working
1 adult households	0.532***	0.531***	0.532***	>1 adult households
Manual workers	0.411***	0.412***	0.411***	White collar clerical
Skilled manual workers	-0.284***	-0.282***	-0.284***	White collar clerical
Professionals	-0.296***	-0.297***	-0.296***	White collar clerical
Household income less £25k	0.545***	0.546***	0.546***	Income 25-50k
Household income more than £50k	-0.241***	-0.242***	-0.242***	Income 25-50k
Built form latent variable	0.605***	0.607***	0.606***	Not applicable

*** significant with 99% confidence interval, ** 95%, * 90%.

Table 9 presents direct influences of socioeconomic profiles, the built form latent variable and car ownership upon travel distances, times and frequencies for all three trip purposes. It is not surprising that lower income occupations and lower household incomes travel less. For commuting (Panel 9a), the most striking difference is between full- and part-time workers. Full-time workers spent 41.4 more minutes, travel 17 more miles and make 50% more trips per week[26] than part-time workers. Residing in denser urban areas implies 3 mile less commuting

[25] We have done WLS-based tests as well which confirm this ranking through standardised coefficients (See Appendix B).

[26] Since negative binomial regression is used for trip frequency, the intercept and coefficients in Table 9 needs to be interpreted. For example, the coefficient for FT workers is 0.41. This means that the influence of full-time working compared with the reference segment is equal to exp (0.41)=1.5 times the reference trips. We have reported the trip frequency coefficients in exponential form because it can be readily used to add to indirect

distance, and this influence remains highly significant after controlling for self-selection, spatial sorting and car ownership effects. Further, having no access to a car in the household implies 4 miles less commuting distance, but 25.4 minutes more travel time and 12% (coefficient 0.116) more trips per week.

For shopping (Panel 9b), dense urban built form and not having a car implies shorter travel distances (3 miles in each case). Dense urban areas also implies 2.4 minutes shorter travel time, and 4% (coefficient -0.039) fewer trips. Males on average spent 11.8 minutes, travel 3 miles, and make 24% (coefficient -0.270) trips less than females. Working full time implies less shopping travel, although the influence is well less than half that of gender.

For other travel (Panel 9c), dense urban built form and not having a car imply shorter travel distance, less travel time and fewer trips, although the effects are far less prominent when compared with the weekly totals.

influences (see Section 5.2) or to estimate trip frequency of specific individuals. For ease of understanding we have converted the coefficients in to elasticity of trips when commenting on them.

Table 9 Direct influences on travel time, distance and trips arising from traveller profiles

Direct influence	Travel Distance (miles)	Travel Time (minutes)	Trip Frequency (trips in exponential unit)	Reference variable
		Panel 9a. Direct influences on commuting		
Male	11***	12.1***	0.005	Female
Full time working	17***	41.4***	0.410***	Part time working
1 adult households	3***	-1.3	-0.068***	>1 adult households
Semi- or unskilled manual workers	-3***	-4.5***	0.126***	White collar clerical
Skilled manual workers	-4***	-11.0***	0.012	White collar clerical
Professionals	3***	1.2	-0.009	White collar clerical
Household income less £25k	-4***	-6.4***	0.006	Income 25-50k
Household income more than £50k	5***	8.0***	-0.072***	Income 25-50k
Built form latent variable	-3***	10.4***	0.023***	(Not applicable)
No car in household	-4***	25.4***	0.116***	With car in household
Intercepts	11***	44.6***	0.782***	
All group averages for comparison	30.3 miles	88.5 min	3.3 trips	(from Table 2)
		Panel 9b. Direct influences on shopping		
Male	-3***	-11.8***	-0.270***	Female
Full time working	-1***	-5.0***	-0.148***	Part time working
1 adult households	1***	2.4***	0.119***	>1 adult households
Semi- or unskilled manual workers	-1***	-3.8***	-0.091***	White collar clerical
Skilled manual workers	-1***	-3.7***	-0.127***	White collar clerical
Professionals	0	-0.9	-0.031**	White collar clerical
Household income less £25k	-1***	0.1	0.011	Income 25-50k
Household income more than £50k	0	0.3	-0.028**	Income 25-50k
Built form latent variable	-3***	-2.4***	-0.039***	(Not applicable)
No car in household	-3***	0.8	-0.183***	With car in household
Intercepts	13***	46***	1.002***	
All group averages for comparison	11.3 miles	39.3 min	2.3 trips	(from Table 2)
		Panel 9c. Direct influences on other purposes combined		
Male	15***	16.3***	-0.063***	Female
Full time working	2***	-14.7***	-0.163***	Part time working
1 adult households	21***	40.8***	0.163***	>1 adult households
Semi- or unskilled manual workers	-20***	-42.9***	-0.181***	White collar clerical
Skilled manual workers	-20***	-44.1***	-0.183***	White collar clerical
Professionals	13***	18.6***	0.059***	White collar clerical
Household income less £25k	-11***	-17.6***	-0.052***	Income 25-50k
Household income more than £50k	18***	27.3***	0.025**	Income 25-50k
Built form latent variable	-12***	-4.4***	-0.047***	(Not applicable)
No car in household	-24***	-33.0***	-0.409***	With car in household
Intercepts	61***	190.5***	2.159***	
All group averages for comparison	72.9 miles	183.5 min	7.6 trips	(from Table 2)

*** significant with 99% confidence interval, ** 95%, * 90%.

Some significant interactions exist between different purposes of travel. Table 10 Panel 10a shows that each one marginal minute a worker spends on commuting would imply a reduction of 0.22 minute on shopping, and each one marginal trip for commuting a reduction of 3.2% shopping trip. Increasing commuting distance appears to have little effect on shopping distance. Similarly, Panel 10b indicates that a marginal unit increase in commuting imply slightly less travel for other purposes: 0.139 fewer miles, 0.299 fewer minutes and

7.5% fewer trips respectively. The influence from shopping, however, is rather different: those who spend one minute more on shopping travel tend to spend on average 0.336 minutes more for other purposes, e.g. leisure and visiting friends; this positive influence also exists for travel distance and trip frequency.

Table 10 Direct influences on travel time, distance and trips arising from trip purpose interactions

Direct influence from	Travel distance	Travel time	Trip frequency
Panel 10a. Direct influences on shopping			
Commuting	0.00	-0.220***	-0.032***
Panel 10b. Direct influences on all other travel purposes			
Commuting	-0.139***	-0.299***	-0.073***
Shopping	0.328***	0.336***	0.071***

*** significant with 99% confidence interval, ** 95%, * 90%.

5.2.3. Indirect influences

The greatest added value of the models is their quantification of indirect influences. The indirect influences are quantified by multiplying the coefficients along the SEM paths (Figure 6). Below we include the direct impacts for comparison and for computing the combined influences.

Table 11 presents the indirect influences of socioeconomic attributes on car ownership via built form. It confirms the strong influence of self-selection and spatial sorting on car ownership. For instance, the first three data rows for ‘Full time -> No car’ shows that once income, occupation and households size are controlled for, working full time has little direct influence over car ownership, as indicated by the near-zero coefficients of -0.002/-0.003. However, because full time workers tend to live in denser and larger urban areas, their car ownership is actually lower (the positive coefficients of 0.058/0.060 indicates a lower level of car ownership than the reference segment). Similarly, the next three data rows for ‘Income over 50K->No car’ show that although the relatively high income implies higher car ownership, the fact that such households tend to live in denser and larger urban areas means that their household car ownership levels tend to be slightly offset. Most indirect influences reinforce the direct ones: the indirect influence of living in dense urban areas depresses car ownership of single adult households by -0.120/-0.121, or a fifth of the direct coefficient. Similarly, the skilled manual and professional workers tend to live in less dense and more rural areas, which raise their car ownership levels.

Table 11 Direct and indirect influences on household car ownership

Direct influence	Indirect influence	Travel distance	Travel time	Travel frequency
Full time->No car		*-0.003*	*-0.002*	*-0.002*
	Full time->LU->NoCar	0.060	0.058	0.059
Combined		**0.057**	**0.056**	**0.057**
Income over 50K->No car		*-0.242*	*-0.241*	*-0.242*
	Income over 50K->LU->NoCar	0.039	0.038	0.038
Combined		**-0.203**	**-0.203**	**-0.204**
1 Adult->No car		*0.531*	*0.532*	*0.532*
	1 Adult->LU->NoCar	0.120	0.121	0.121
Combined		**0.651**	**0.653**	**0.653**
Skilled Manual ->No car		*-0.282*	*-0.284*	*-0.283*
	Skilled Manual->LU->NoCar	-0.120	-0.123	-0.123
		-0.402	**-0.407**	**-0.406**
Prof->No car		*-0.297*	*-0.296*	*-0.296*
	Prof->LU->NoCar	-0.110	-0.113	-0.113
Combined		**-0.407**	**-0.409**	**-0.409**

Note: Insignificant effects are not reported. A combined effect includes only its significant components.

Results from Table 12 to Table 14 show that built form and the majority of the socioeconomic attributes have significant indirect influences on car ownership and the extents of travel. The results form a rich tapestry of reinforcing effects in some and counteracting ones in others. We consider the SEM results statistically more robust than existing quantifications. The findings on the combined influences on travel distances tend to confirm those from recent literature e.g. (Cao et al., 2007b) that the influence of built form characteristics on travel distances is larger than those of socio-demographic profiles and that denser urban areas with frequent bus services contribute to shorter travel distances. Our model results also provide lesser-known insights into travel time and trip frequency.

For commuting, Table 12 highlights considerable negative effects for workers from single adult and economically disadvantaged households: although the direct effects suggest that workers from single adult households tend to commute 3 miles longer with little differences in travel time or trip frequency. However, after combining the indirect effects, they actually commute 0.2 miles less and 18.7 more minutes, which is 32% slower than the reference segment (cf Table 3, lower panel). Further down the results suggests a similar pattern for workers in manual occupations and with household incomes less than £25,000 per year: their commuting distances are respectively 4.3 and 6.1 miles shorter, and at the same time their commuting times are 4.5 and 7.5 minutes longer than white-collar clerical workers; by contrast, professional workers commute 5 miles longer and 12.3 minutes less. Towards the bottom of Table 12 it is clear that the dense urban areas tend to imply shorter commuting distance and longer time (by 3 miles and 10.4 minutes respectively), but the combined influence of not having a car and living in dense urban areas have a much larger effect - the travel distance is -2.3-4=-6.3 miles and travel time is 15.4+25.4=40.8 minutes more, implying a speed that is 77% slower than the reference segment.

Table 12 Direct and indirect influences on home-based commuting (HBW)

Direct influence	Indirect influence	Travel distance (miles)	Travel time (minutes)	Travel frequency (trips)
FT->HBW		*17.0*	*41.4*	*0.410*
	FT->LU->HBW	-0.3	1.0	0.002
	FT->LU->NoCar->HBW	-0.2	1.5	0.007
Combined		16.5	43.9	0.419
1adult->HBW		*3.0*	*not significant*	*-0.068*
	1adult->LU->HBW	-0.7	2.1	0.005
	1adult->LU->NoCar->HBW	-0.5	3.1	0.014
	1adult->NoCar->HBW	-2.0	13.5	0.062
Combined		-0.2	18.7	0.013
Manual->HBW		*-3.0*	*-4.5*	*0.126*
	Manual->LU->HBW	0.19	-0.6	*not significant*
	Manual->LU->NoCar->HBW	0.13	-0.8	*not significant*
	Manual->NoCar->HBW	-1.6	10.4	0.048
Combined		-4.28	4.5	0.174
SkillManual->HBW		*-4.0*	*-11.0*	*not significant*
	SM->LU->HBW	0.7	-2.1	-0.005
	SM->LU->NoCar->HBW	0.5	-3.1	-0.014
	SM->NoCar->HBW	1.0	-7.2	-0.033
Combined		-1.8	-23.4	-0.052
Prof->HBW		*3.0*	*not significant*	*not significant*
	Prof->LU->HBW	0.6	-1.9	-0.004
	Prof->LU->NoCar->HBW	0.4	-2.9	-0.013
	Prof->NoCar->HBW	1.1	-7.5	-0.034
Combined		5.1	-12.3	-0.052
IncomeLess25k->HBW		*-4.0*	*-6.4*	*not significant*
	IncomeLess25k->NoCar->HBW	-2.1	13.9	0.063
Combined		-6.1	7.5	0.063
IncomeOver50K->HBW		*5.0*	*8.0*	*not significant*
	IncomeOver50K->LU->HBW	-0.2	0.6	0.001
	IncomeOver50K->LU->NoCar->HBW	-0.14	1.0	0.004
	IncomeOver50K->NoCar->HBW	0.9	-6.1	-0.028
Combined		5.54	3.5	-0.023
LU->HBW		*-3.0*	*10.4*	*0.023*
	LU->NoCar->HBW	-2.3	15.4	0.07
Combined		-5.3	25.8	0.093
No Car -> HBW		-4.0	25.4	0.116

Note: Insignificant effects are not reported. A combined effect includes all significant components.

Table 13 presents the direct, indirect and combined results for shopping travel. Gender and full time working continue to be the biggest influences on travel distance and time, after accounting for indirect influences. Interestingly, the indirect influences are minor. In particular, the indirect effect through commuting time is tiny and only accounts for 0.3 minutes per week of the difference in shopping travel time between males and females.

Table 13 Direct and indirect influences on shopping travel (Sh)

Direct influence	Indirect influence	Travel distance (miles)	Travel time (minutes)	Travel frequency (trips)
Male->Sh		*-3.0*	*-11.8*	*-0.270*
	Male->HBW->Sh	not significant	-0.3	not significant
Combined		-3	-12.1	-0.270
FT->Sh		*-1.0*	*-5.0*	*-0.148*
	FT->LU->Sh	-0.3	-0.2	-0.004
	FT->LU->HBW->Sh	not significant	0.0	not significant
	FT->LU->NoCar->HBW->Sh	not significant	1.5	not significant
	FT->HBW->Sh	not significant	-0.9	-0.013
Combined		-1.3	-4.6	-0.165
1adult->Sh		*1.0*	*2.4*	*0.119*
	1adult->LU->Sh	-0.7	-0.5	-0.008
	1adult->LU->HBW->Sh	not significant	0.0	not significant
	1adult->LU->NoCar->HBW->Sh	not significant	3.1	not significant
	1adult->NoCar->HBW->Sh	not significant	-0.3	-0.002
Combined		0.3	4.7	0.109
Manual->Sh		*-1.0*	*-3.8*	*-0.091*
	Manual->HBW->Sh	not significant	-0.3	not significant
	Manual->NoCar->HBW->Sh	not significant	-0.2	-0.002
Combined		-1	-4.3	-0.093
SkillManual->Sh		*-1.0*	*-3.7*	*-0.127*
	SM->LU->Sh	0.7	0.5	0.008
	SM->LU->HBW->Sh	not significant	0.1	not significant
	SM->LU->NoCar->HBW->Sh	not significant	0.07	not significant
	SM->HBW->Sh	not significant	0.2	not significant
	SM->NoCar->HBW->Sh	not significant	0.2	0.001
Combined		-0.3	-2.6	-0.118
Prof->Sh		*not significant*	*not significant*	*not significant*
	Prof->LU->Sh	0.6	0.4	0.007
	Prof->LU->HBW->Sh	not significant	0.07	not significant
	Prof->LU->NoCar->HBW->Sh	not significant	0.06	not significant
	Prof->NoCar->HBW->Sh	not significant	0.2	0.001
Combined		0.6	0.7	0.008
IncomeLess25k->Sh		*-1.0*	*not significant*	*not significant*
	IncomeLess25k->HBW->Sh	not significant	0.1	not significant
	IncomeLess25k->NoCar->HBW->Sh	not significant	-0.3	-0.002
Combined		-1	-0.2	-0.002
IncomeOver50K->Sh		*not significant*	*not significant*	*not significant*
	IncomeOver50K->LU->Sh	-0.2	-0.1	-0.002
	IncomeOver50K->HBW->Sh	not significant	-0.2	0.002
	IncomeOver50K->NoCar->HBW->Sh	not significant	0.1	0.001
Combined		-0.2	-0.2	0.001
LU->Sh		*-3.0*	*-2.4*	*-0.039*
	LU->HBW->Sh	not significant	-0.2	-0.001
	LU->NoCar->HBW->Sh	not significant	-0.3	-0.002
Combined		-3	-2.9	-0.069
NoCar->Sh		*-3.0*	*not significant*	*-0.183*
	NoCar->HBW->Sh	not significant	-0.6	-0.004
Combined		*-3.0*	-0.6	*-0.187*

Note: Insignificant effects are not reported. A combined effect includes all significant components.

There is a long list of significant but minor indirect influences for combined other travel purposes, we have selected the largest effects to report in Table 14. Similar to commuting and shopping, the indirect influences through car ownership are the largest ones. For instance, those who live in dense urbanized areas tend to make fewer trips and travel for shorter time and distances. This is where urban, mixed use and high population density are effective in improving accessibility without an adverse effect on travel mobility.

Table 14 Direct and indirect influences on other travel (Oth)

Direct influence	Indirect influence	Travel distance	Travel time	Travel
1adult->Oth		*21*	*40.8*	*0.163*
	1adult->NoCar>Oth	*-12.8*	*-17.5*	*-0.218*
Combined		*8.2*	*23.3*	*-0.055*
IncomeLess25k->Oth		*-11*	*-17.6*	*-0.052*
	IncomeLess25k->NoCar>Oth	-13.1	*-18.0*	-0.223
Combined		-24.1	*-35.6*	-0.275
LU->Oth		*-12*	*-4.4*	*-0.047*
	LU->NoCar>Oth	-14.6	*-19.9*	-0.248
Combined		-26.6	*-24.3*	-0.295
FT->Oth		*2*	*-14.7*	*-0.163*
	FT->HBW>Oth	-2.3	*-12.4*	-0.03
Combined		-0.3	*-27.1*	-0.166

In order to further confirm the importance of the indirect effects, we test an alternative model that treats car ownership as an exogenous variable – that model is otherwise identical to our main model as shown in Figure 6. Table 15 compares the model results for commuting. In the alternative model, the direct effect of not owning a car on commuting distance is to travel 4 miles less on average, which is identical to the results from the SEM model. The direct influence from built form is comparable. The difference in total effects is clearly attributed to the indirect influences of interactions between living in a denser area and resulting lower propensity of car ownership. The alternative model predicts an overall influence of -7.0 miles, compared with -9.0 miles from the SEM, which points to an underestimation of the impacts by 29%. This is also the case for commuting time and frequency with respectively an overall underestimation of the impacts respectively by 36% and 50% by the alternative model.

Table 15 Comparison of results from the SEM (with an endogenous car ownership variable) and the alternative model with an exogenous car ownership variable for commuting

	Direct influence of built form	Indirect influence of built form via car ownership	Combined influence of built form	Direct influence of car ownership	Overall influence
	on commuting distance (miles/week)				
SEM model	-3.0	-2.0	-5.0	-4.0	-9.0
Exogenous car ownership model	-3.0	Excluded	-3.0	-4.0	-7.0
	on commuting time (minutes/week)				
SEM model	10.5	15.4	25.9	25.3	51.2
Exogenous car ownership model	10.4	Excluded	10.4	27.1	37.5
	on number of commuting trips (trips/week)				
SEM model	0.028	0.070	0.098	0.116	0.213
Exogenous car ownership model	0.022	Excluded	0.022	0.120	0.142

5.2.4. Variations over year

We further extend the path-diagram based SEM analyses through subdividing the NTS data into nine subsets by year (i.e. 2002 to 2010). The purpose is to examine any systematic variations in influences over time. For comparison, we also set up a benchmark model in which the coefficients are not allowed to vary. A comparison of the models' goodness-of-fit indicates the one performs better, and the coefficients estimates reveal any significant changes.

Table 16 compares the goodness-of-fit using three measures: the Akaike Information Criterion (AIC), the Bayesian Information Criterion (BIC) and the Sample-Size Adjusted BIC (ABIC). Whilst the AIC aims to select the model that most adequately describes an unknown, high dimensional reality, the BIC family of measures are developed for comparison between known, candidate models.

Table 16 Goodness of fit statistics: Constrained Model vs Grouped Model

	AIC	BIC	ABIC
		Travel distance	
Constrained model	249,123	250,286	249,885
Grouped model	249,249	254,323	252,575
		Travel time	
Constrained model	1,632,658	1,633,820	1,633,420
Grouped model	1,632,578	1,637,652	1,635,904
		Trip frequency	
Constrained model	2,221,466	2,222,628	2,222,228
Grouped model	2,221,482	2,226,556	2,224,808

The goodness of fit indicators suggests that there is no strong evidence in support of the model which allows variations over time. With the exception of AIC for the Travel Time model which is marginally lower in the grouped model, for all other cases the goodness of fit indicators of the benchmark model is smaller. This indicates that grouping the model into 9 years is not improving the overall performance. This, however, does not rule out the potentials for systematic variations in some influences over years.

Figure 19 to Figure 28 in Appendix B illustrate the trend of changes in all influences. Figure 7 to Figure 9 below extracted those with systematic trend of changes over time for each of the three models (i.e. travel distance, time and number of trips). In order to make comparison across years, the coefficients for the year 2002 is set to 100 as the benchmark index and the rest are shown relative to those. The actual coefficients and the P-values are shown in Table 17.

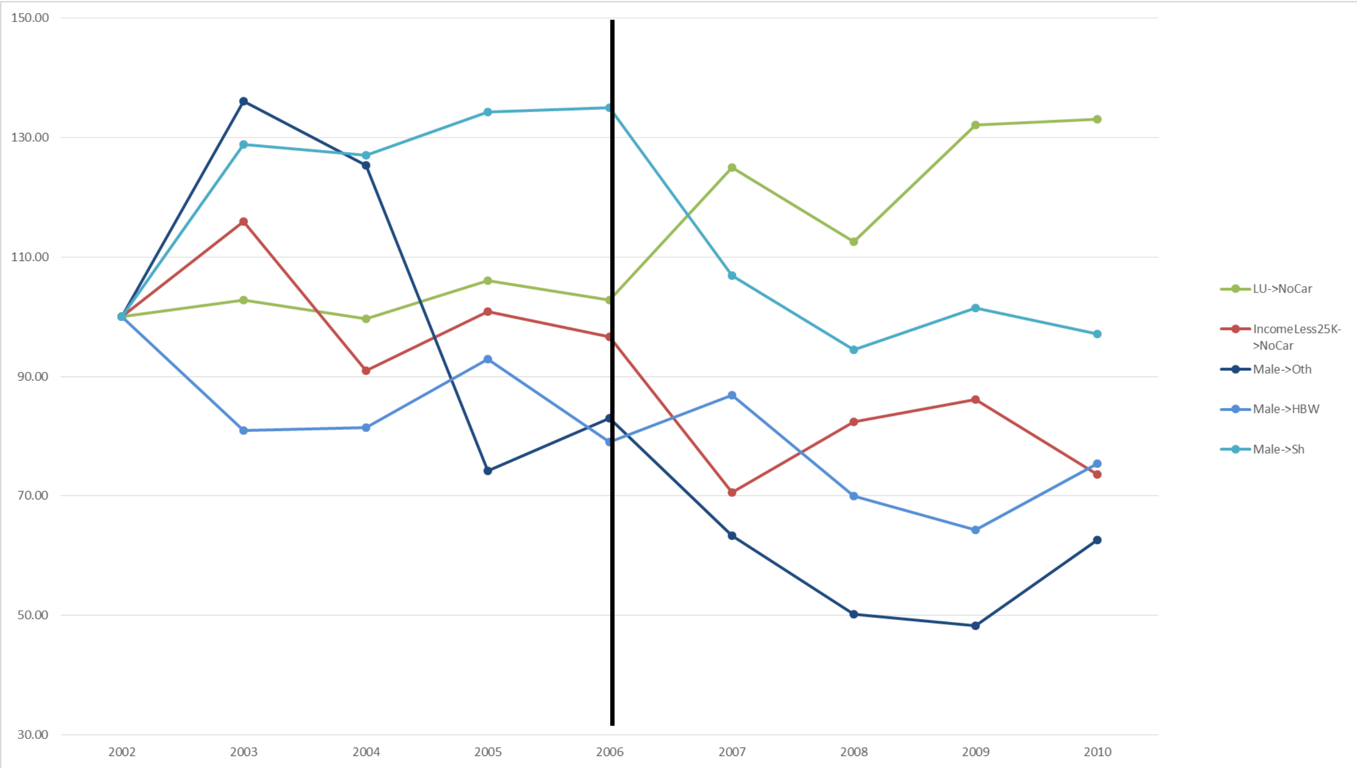

Figure 7 influences with significant trend of changes over time (travel distance model)

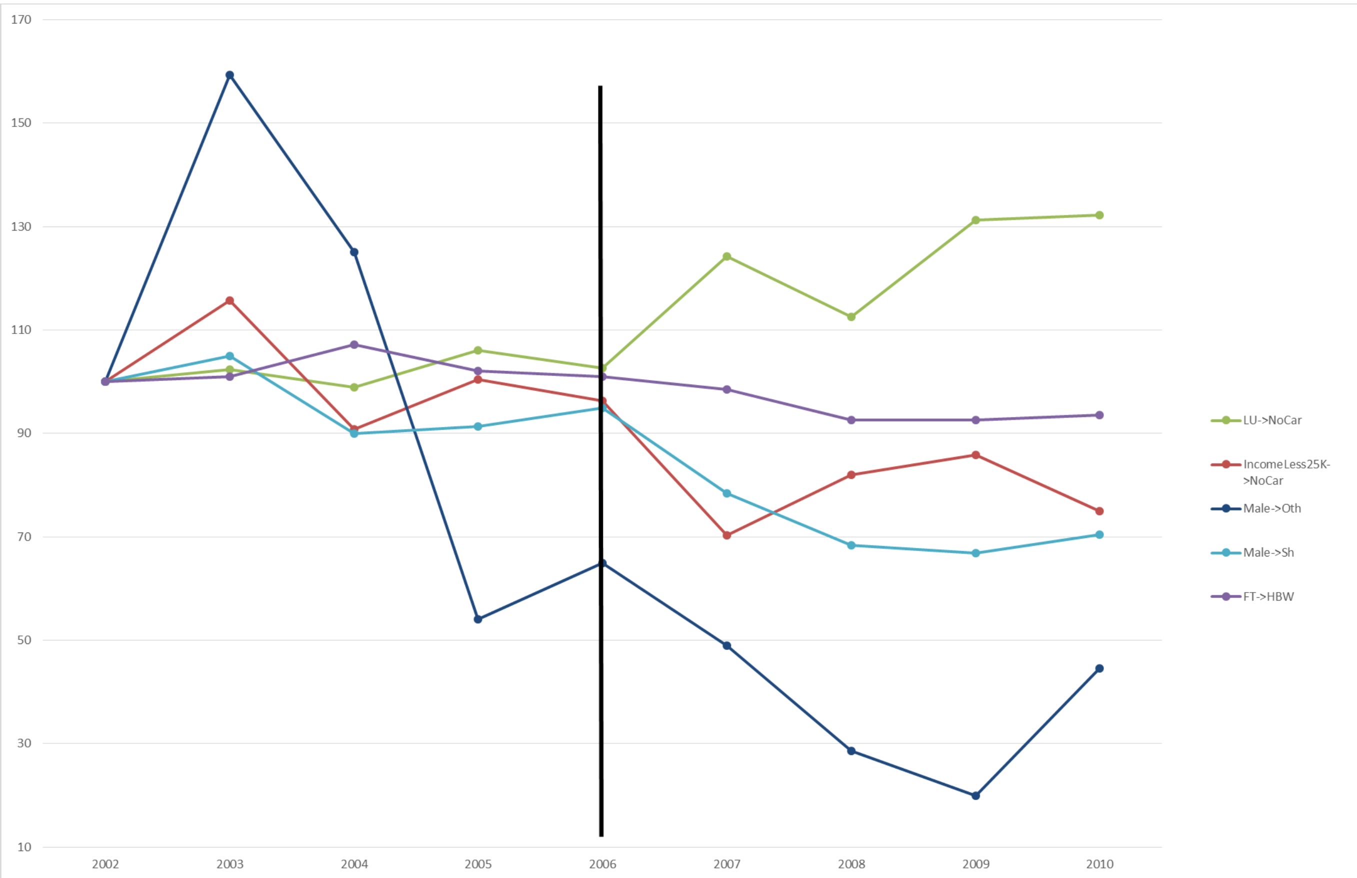

Figure 8 influences with significant trend of changes over time (travel time model)

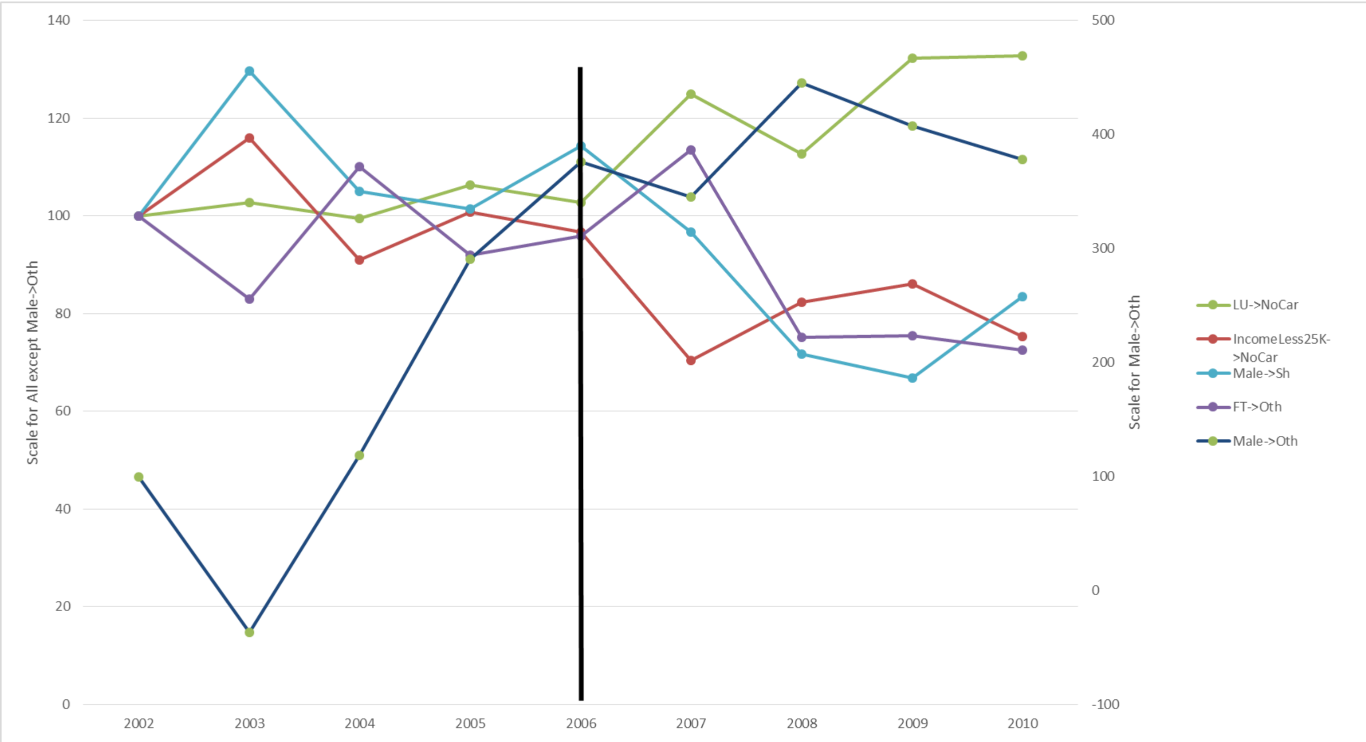

Figure 9 influences with significant trend of changes over time (trip frequency model) [27]

[27] Note that the chart has primary and secondary axis. This is due to rapid rise in the influences of male on other trip frequency from around 100 in 2002 to 2004 to around 400 after 2006.

Table 17 coefficients and P-values of influences with systematic trend of changes over time

Year	LU->NoCar	IncomeLess 25k->NoCar	Male->HBW	Male->Sh	Male->Oth	FT->HBW	FT->Oth
Travel Distance Model							
2002	0.541 (0.000)	0.601 (0.000)	13.1 (0.000)	-2.7 (0.000)	18.2 (0.000)	No systematic trend over time	No systematic trend over time
2003	0.556 (0.000)	0.697 (0.000)	10.6 (0.000)	-3.5 (0.000)	24.8 (0.000)		
2004	0.539 (0.000)	0.547 (0.000)	10.7 (0.000)	-3.5 (0.000)	22.8 (0.000)		
2005	0.574 (0.000)	0.606 (0.000)	12.2 (0.000)	-3.7 (0.000)	13.5 (0.000)		
2006	0.556 (0.000)	0.581 (0.000)	10.4 (0.000)	-3.7 (0.000)	15.1 (0.000)		
2007	0.676 (0.000)	0.424 (0.000)	11.4 (0.000)	-2.9 (0.000)	11.5 (0.000)		
2008	0.609 (0.000)	0.495 (0.000)	9.2 (0.000)	-2.6 (0.000)	9.1 (0.000)		
2009	0.715 (0.000)	0.518 (0.000)	8.4 (0.000)	-2. 8 (0.000)	8.8 (0.000)		
2010	0.72 (0.000)	0.442 (0.000)	9.9 (0.000)	-2.7 (0.000)	11.4 (0.000)		
Travel Time Model							
2002	0.54 (0.000)	0.602 (0.000)	No systematic trend over time	-13.9 (0.000)	23.1 (0.000)	41.9 (0.000)	No systematic trend over time
2003	0.553 (0.000)	0.697 (0.000)		-14.6 (0.000)	36.8 (0.000)	42.3 (0.000)	
2004	0.534 (0.000)	0.547 (0.000)		-12.5 (0.000)	28.9 (0.000)	44.9 (0.000)	
2005	0.573 (0.000)	0.605 (0.000)		-12.7 (0.000)	12.5 (0.002)	42.8 (0.000)	
2006	0.554 (0.000)	0.58 (0.000)		-13.2 (0.000)	15.0 (0.000)	42.3 (0.000)	
2007	0.671 (0.000)	0.423 (0.000)		-10.9 (0.000)	11.3 (0.005)	41.3 (0.000)	
2008	0.608 (0.000)	0.494 (0.000)		-9.5 (0.000)	6.6 (0.060)	38.8 (0.000)	
2009	0.709 (0.000)	0.517 (0.000)		-9.3 (0.000)	4.6 (0.225)	38.8 (0.000)	
2010	0.714 (0.000)	0.451 (0.000)		-9.8 (0.000)	10.3 (0.007)	39.2 (0.000)	
Number of Trips Model							
2002	0.54 (0.000)	0.343 (0.000)	No systematic trend over time	-0.578 (0.000)	-0.169 (0.316)	No systematic trend over time	-1.34 (0.000)
2003	0.555 (0.000)	0.332 (0.000)		-0.75 (0.000)	0.062 (0.677)		-1.11 (0.000)
2004	0.537 (0.000)	0.455 (0.000)		-0.607 (0.000)	-0.2 (0.154)		-1.47 (0.000)
2005	0.574 (0.000)	0.36 (0.000)		-0.586 (0.000)	-0.491 (0.001)		-1.23 (0.000)
2006	0.555 (0.000)	0.465 (0.000)		-0.661 (0.000)	-0.636 (0.000)		-1.29 (0.000)
2007	0.675 (0.000)	0.475 (0.000)		-0.559 (0.000)	-0.584 (0.000)		-1.52 (0.000)
2008	0.609 (0.000)	0.367 (0.000)		-0.415 (0.000)	-0.753 (0.000)		-1.01 (0.000)
2009	0.714 (0.000)	0.448 (0.000)		-0.386 (0.000)	-0.689 (0.000)		-1.01 (0.000)
2010	0.717 (0.000)	0.427 (0.000)		-0.483 (0.000)	-0.639 (0.000)		-0.97 (0.000)

The examination of Figure 7 to Figure 9 reveals that most influences show noticeable shift after the year 2006. For instance, the speed of rise in average built form influence and drop in average income effect on car ownership has been increased considerably after 2006. Also the gender gap in shopping travel distance, time and frequency is narrowing sharper after 2006.

Some of the substantial change in the trend of changes after 2006 can be associated to the recent economic downturn. For instance, one can argue that the decline in car ownership could have reduced family expenditure specifically in denser areas where alternative modes of transport is more accessible and the cost of keeping car is higher. To make a better understanding of this trend, we need to explicitly compare the changes before and after the year 2006. Considering the limited number of years of data which is available, grouping 2002 to 2006 and 2007 to 2010 can also help making more rigorous conclusion by limiting the model complexity. The comparison of influences pre- and post-2006 is reported and discussed in details in section 5.2.5.[28] Here we have highlighted the main learnings from analysis of the overall trends:

a) The gender gap in travel is declining. The differences between female and male in their travel distance to work, shopping and all other purposes is reduced. Also, the travel time spent for shopping and other purposes, and the number of trips made for shopping is fallen. The only change in reverse is males' trip frequency; when compared to that of female, males are making more and more trips for other purposes over time. However, even this sharp trend is stabilizing after 2006.
b) The influence of built form characteristics on car ownership is increasing with the trend of that become sharper after 2006. Travellers are progressively prepared to forgo cars when they live in denser more urbanized areas.
c) The car ownership of lower income group is getting closer to medium and higher income bands. This is not due to increase in the purchasing power of the lowest income band, but is mainly due to drop in car ownership of the higher income groups specifically after 2006[29].

[28] As explained in section 5.2.5, we have omitted the year 2007 from pre- post-2006 (7) analysis as the financial crisis might have affected some sectors but not others in 2007.
[29] This can also be due to the economic downturn and associated reduction in families' purchasing power. However, it is difficult to make firm conclusion without having access to more years of data. A repeat of analysis with more years of data is recommended as a follow-up study

5.2.5. Direct and indirect influences pre- and post-2007

We further extend the path-diagram based SEM analyses through subdividing the NTS data into two subsets: 2002-2006 and 2008-2010. We exclude the year 2007 because the financial crisis had already crept in for some sectors but not others in the UK. The purpose is to evaluate the extent and significant level of the shifts observed in the influences post-2007 in section 5.2.4 above. The model estimation is carried out through a multi-group model where the influences are allowed to vary between the two groups of years.

Table 18 compares the goodness-of-fit using three measures: the Akaike Information Criterion (AIC), the Bayesian Information Criterion (BIC) and the Sample-Size Adjusted BIC (ABIC). For all three travel characteristics, the AIC suggests that the grouped model is better performing whereas the BIC and ABIC prefer the benchmark model. Although this contradictory signal is a caveat, we consider the AIC results important because it indicates that there had been real changes in the underlying patterns of the influences. Three-group model shows more improvement in comparison to the benchmark than the nine-group model reported in section 5.1.4.

Table 18 Goodness of fit statistics: Constrained Model vs Grouped Model

	AIC	BIC	ABIC
		Travel distance	
Constrained model	61,333	62,163	61,877
Grouped model	61,254	63,062	62,440
		Travel time	
Constrained model	1,445,032	1,445,863	1,445,577
Grouped model	1,444,897	1,446,705	1,446,082
		Trip frequency	
Constrained model	1,910,941	1,911,771	1,911,485
Grouped model	1,910,850	1,912,658	1,912,035

As we are employing simulation algorithms to estimate Maximum Likelihood estimator, only relative goodness of fit statistics such as AIC and BIC can be reported. Table 38 in Appendix B reports the absolute goodness of fit statistics of WLS model of travel time which confirms the good fit to the observed data.

WLS and ML algorithm confirm that there are five types of statistically significant coefficient changes (Table 19). First, the gender gap appears to be slightly narrowing after 2007 for travel, with the differences between females and males in shopping reduced by 14% in distance, 23%

in time and -7% in trip frequency. Similarly the gap in commuting distance narrowed by 18%, and in other travel distance by 47% and time by 70%. The only significant change in reverse is the frequency of the males' other trips, which widened slightly (by around 7%). Our analyses of the NTS data show that for shopping trips, it is the females who have reduced their travel distance; for commuting and other trips, it is the males' reduced travel that narrowed the gap (cf Table 4).

Secondly, there is an increased influence of built form on car ownership – built form is already the strongest influence; post 2007, living in a larger, denser urban area is a 21-24% stronger influence on forgoing car ownership.

Thirdly, full time working post 2007 appears to have had an influence in slightly reducing shopping travel distance.

Fourthly, in line with the trends above, the gap in travel distance between the low and the middle income group had narrowed from 13 to 7 miles. This is because the rate of drop in travel distance for middle income group has been higher vis-à-vis that for the low income.

Fifthly, the positive association between the frequencies of shopping and other travel appears to have marginally strengthened (by 0.7%).

Furthermore, it is important to note that the majority of the influences remain remarkably stable over time. For instance, the large differences between full- and part-time working in terms of commuting distance and time had not changed, in spite of the rapid rise in part-time and free-lancing work, and in the spread of ICT usage.

Table 19 Summary of significantly changed influences pre- and post-2007

Direct Effects	Coefficients and (p-value) pre 2007	Coefficient and (p-value) post 2007	% change in coefficient values
	Travel Distance		
Male->HBW	11.0 (0.00)	9.0 (0.00)	-18%
Male->Sh	-3.5 (0.00)	-3.0 (0.00)	-14%
Male->Oth	19 (0.00)	10 (0.00)	-47%
LU->NoCar	0.55 (0.00)	0.68 (0.00)	24%
FT->Sh	0.00 (0.084)	-2.0 (0.00)	--
IncomeLess25k->Oth	-13.0 (0.00)	-7.0 (0.00)	-46%
	Travel Time		
Male->Sh	-13 (0.00)	-10 (0.00)	-23%
Male->Oth	23 (0.00)	7 (0.00)	-70%
LU->NoCar	0.55 (0.00)	0.67 (0.00)	22%
	Number of Trips		
Male->Sh	-0.29 (0.00)	-0.22 (0.00)	-7%
Male->Oth	-0.034 (0.00)	-0.1 (0.00)	7%
Sh->Oth	0.068 (0.00)	0.075 (0.00)	0.7%
LU->NoCar	0.56 (0.00)	0.68 (0.00)	21%

5.3. Findings from the LCA-SEM

As discussed in Chapter 3, LCA-SEM follow the same framework as a path-diagram based SEM except for the built form factor which is now modelled as a categorical rather than continuous latent variable. While modelling built form as one continuous latent variable provides a first indication of the web of direct and indirect influences on travel, LCA-SEM allows the identification of distinct built form clusters through which we can measure potential nonlinear built form influences[30]. We summarise the main findings of LCA-SEM in three steps.

First, we present the latent built form classes, their definition and unconditional and conditional probabilities for individuals to be in each class. Second, we compare the socioeconomic characteristics of residents within the built form latent classes. Finally, within each built form class, we explore influences on travel distance and time by journey purpose after controlling for interactions among journey purposes as well as endogeneities arising from self-selection, spatial sorting and car ownership.

[30] Path-diagram based SEM provides average influences on and influences of land use latent variable. LCA-SEM, however, allow comparing influences across land use clusters which might not be easy to capture by one average coefficient from path-diagram based SEM.

5.3.1. Latent classes of the built form in the UK

The basic approach to categorisation of latent classes of the built form is to run the LCA using NTS variables that describe the relevant characteristics of the areas the respondents live in. We have developed an extended, conditional LCA model, in which we include the demographic and socioeconomic characteristics as covariates. This involves a simultaneous estimation of the influence of the residents' demographic and socioeconomic profiles so that the effects arising from spatial sorting are accounted for.

The LCA is built on the EFA for continuous latent variable analysis explained in Section 5.2.1. In the EFA five built form attributes, namely "area type", "population density", "frequency of local buses" , "walk time to bus stop", and "walk time to rail station" are found to have large loading factors, sufficient to be considered as the defining characteristics of the built form[31]. The LCA which defines built form as discrete categorical classes (as opposed to defining a continuous latent variable for the built form in EFA) has similarly found those five attributes to have large and significant loading factors. The availability of five attributes can allow us to define up to 3 distinct built form classes with the sufficient degree of freedom for model estimation.

Our conditional LCA identifies three latent built form classes with an entropy of 0.832[32]. This suggests that the latent classes are very well defined. A cross-tabulation of the most likely latent class membership (row) by latent class (column) in Table 20 corroborates the high entropy value.

---

[31] In path-diagram based SEM, we only consider the three main built form indicators with the loading factor of above 0.7. This is because three indicators would give sufficient degree of freedom to define one continuous factor. Here, however, the more indicators give freedom to test more number of clusters so five indicators with biggest and still significant loading factors are selected

[32] Entropy is measured on a zero to one scale with the value of one indicating the individuals are perfectly classified into latent classes, and a value that is greater than 0.8 indicates a well-defined categorisation (Wang and Wang, 2012).

Table 20 Average latent class probabilities for residents' most likely latent class membership (row) by latent class of the built form (column)

	Class 1	Class 2	Class 3
Class 1 membership	0.917	0.083	0
Class 2 membership	0.045	0.919	0.036
Class 3 membership	0	0.061	0.939

Panel 21a of Table 21 shows the unconditional and conditional probabilities of individuals in each latent class. Based on the estimated model, classes 1 to 3 contain respectively 18%, 54%, and 27% of all working adults.

Conditional probabilities further reveal the patterns of the latent classes benchmarked by the specific characteristics of the built form (Panel 21b of Table 21). For example, residents in Latent Class 1 consists of respectively those from the medium urban, big urban, metropolitan and London area types (of respectively 2.2%, 15.8%, 16.2 and 65.8%), with no one from rural or small urban (see Panel 21b-1). The members of this class also reside in the densest areas (see Panel 21b-2) and benefit from the most frequent buses and highest level of accessibility to public transport (see Panel 21b-3 to 21b-5). These parameters in this latent class prompt us to label it 'Dense urban' in terms of travel behaviour patterns. Similarly, the dominance of medium urban in latent Class 2 (of 46.8% of the residents in this class) and the dominance of rural in latent Class 3 (of 72% of residents) give rise to the labels 'Medium urban' and 'Rural' respectively. The individuals in Class 3 reside in the least dense area with the least convenient access to public transport. Those in Class2 are located between Class 1 and Class 3 in terms of population density, bus frequency, and public transport access.

A comparison across the three columns of latent classes gives us an insight into the distribution of residents within a NTS area type across the latent classes. For instance, for the London area type, 93.7% of the residents there belong to Latent Class 1[33]. This composition by NTS area type is presented in Figure 10.

[33] i.e. (0.658*13853)/(0.658*13853+0.015*40874+0.00*20301) using data in Panel 4b-1 of Table 4.

Table 21 Unconditional and conditional probabilities for the three-class built form LCA model

Indicators		Latent class		
		1-Dense urban (N = 13853)	2-Medium urban (N = 40874)	3-Rural areas (N = 20301)
Panel 21a: Unconditional probabilities				
		0.18	0.54	0.27
Panel 21b: Conditional probabilities				
21b-1: Area type				
	Rural	0	0.003	0.720
	Small Urban	0	0.080	0.179
	Medium Urban	0.022	0.468	0.078
	Big Urban	0.158	0.231	0.022
	Metropolitan	0.162	0.201	0.001
	London	0.658	0.015	0
21b-2: Population Density (person/hectare)				
	Under 10	0.003	0.200	0.949
	10-14.99	0.021	0.125	0.027
	15-19.99	0.019	0.134	0.019
	20-24.99	0.020	0.119	0.005
	25-29.99	0.039	0.122	0
	30-34.99	0.048	0.089	0
	35-39.99	0.053	0.080	0
	40-49.99	0.164	0.096	0
	50-59.99	0.168	0.021	0
	over 60	0.465	0.013	0
21b-3: Bus Frequency				
	Less than once a day	0	0.008	0.206
	At least once a day	0	0	0.027
	At least once every hour	0.005	0.128	0.432
	At least once every 30 minutes	0.131	0.462	0.283
	At least once every 15 minutes	0.864	0.401	0.051
21b-4: Walk time to bus stops				
	44 minutes and more	0	0	0.021
	27-43 minutes	0	0.001	0.021
	14 to 26 minutes	0.007	0.013	0.057
	7 to 13 minutes	0.072	0.078	0.108
	6 minutes or less	0.921	0.908	0.793
21b-5: Walk time to rail station				
	44 minutes and more	0.093	0.336	0.665
	27-43 minutes	0.176	0.207	0.103
	14 to 26 minutes	0.355	0.292	0.129
	7 to 13 minutes	0.224	0.105	0.058
	6 minutes or less	0.150	0.060	0.044

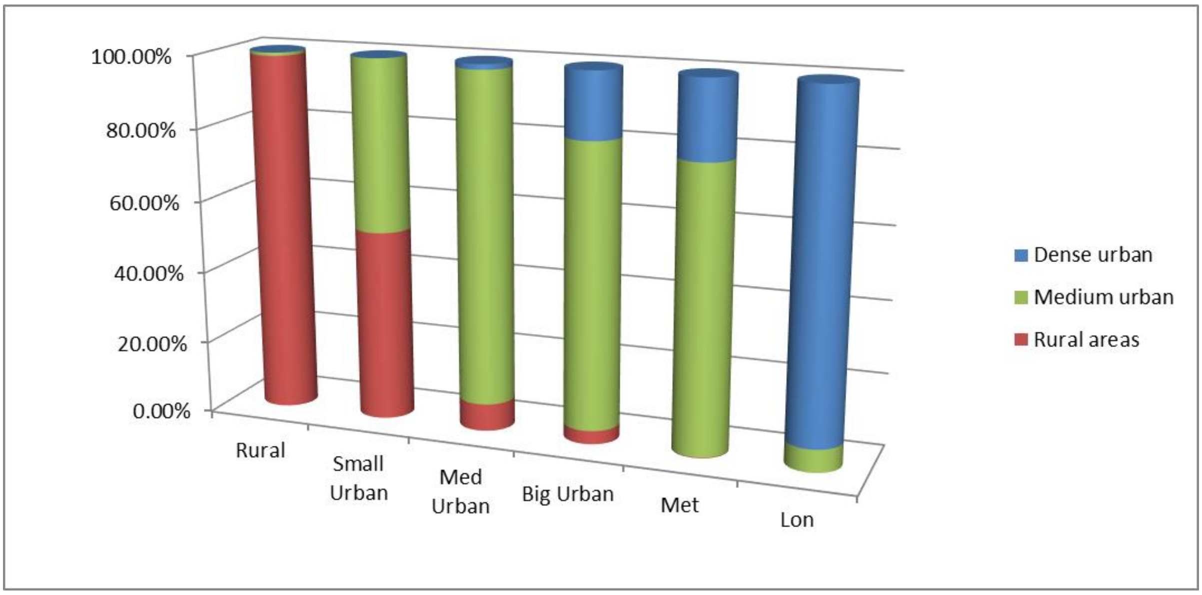

Figure 10 Composition of built form latent classes by NTS area type

5.3.2. Spatial sorting of residents among latent built form classes

The second step of the analysis is to understand how the latent built form class membership interacts with the demographic and socioeconomic profiles of the residents – self-selection and spatial sorting of the residents of different demographic and socioeconomic profiles often has a material bearing on where they live. This is carried out through the estimation of the covariates in the LCA.

The results of this analysis of the covariates are reported in terms of odds ratios with one of the latent classes designated as a reference class. This is shown in Table 22 where Latent Class 2 (Medium urban) is chosen as the reference class. For residents of a particular demographic or socioeconomic characteristic, an odds ratio for a given class of built form that is higher than 1 indicates that those residents are more likely to live in that class of built form than in the reference class areas. Similarly, an odds ratio less than 1 implies the reverse. For instance the odds ratio for being male is 1.077 for the 'Dense urban' class, and this means that male workers are 7.7% more likely to live in the 'Dense urban' areas than the 'Medium urban' areas[34]. The magnitudes of the odds ratios indicate the strength of that difference. For instance, further down in Table 22 the odds ratio of skilled manual workers suggest that they are 15.8% more

[34] This result is different to that produced in section 5.2 where built form is modelled as a continuous latent variable – results from path-diagram based SEM indicate that male workers tend to commute from less dense and more rural locations with less frequent bus services, which is counter-intuitive. This highlights the benefits of modelling built form as a categorical as opposed to a continuous latent variable.

likely to live in ‘Rural areas’ and 43.1% less likely to live in the ‘Dense urban’ areas than in the ‘Medium urban’ areas.

Table 22 Odds ratios of demographic and socioeconomic covariates

Covariates	Built form latent classes		
	1-Dense urban	2-Medium urban	3-Rural areas
Male	1.077***	Used as a reference latent class	1.077***
Full time working	1.115***		0.87***
1 adult households	1.61***		0.866***
Semi- or unskilled manual workers	0.807***		0.978
Skilled manual workers	0.569***		1.158***
Professionals	0.797***		1.294***
Household income less £25k	1.055		0.969
Household income more than £50k	1.565***		1.176***

Base or reference group is class 2 (medium urban class)
*** significant within 99% CI, ** significant within 95% CI, * significant within 90% CI

Not surprisingly, the results in Table 22 suggest that relative to the Medium urban class, working adults who reside in the ‘Dense urban’ areas are more likely to be male, coming from 1 adult households, and with full time working patterns; professionals and skilled manual workers are more likely to be found in the ‘Rural areas’ class. As for household income profiles, the ‘Dense urban’ class has 56.5% more high income households (with income greater than 50k per year) than the ‘Medium urban’; the ‘Rural areas’ by contrast has 17.6% more high income households than in ‘Medium urban’.

These results reconfirm those from path-diagram based SEM (cf Table 8) which shows that full timers and 1 adult households tend to live in more dense urbanized area while professionals and skilled manuals prefer living in less dense more rural areas. However, the findings from LCA-SEM provides more precise interpretation by classifying the built form into clusters and linking those with built form characteristics. For instance, while the linear interpretation from path-diagram based SEM suggest that in average higher income groups tend to live in denser, more urbanized urban area, LCA-SEM shows that they mainly reside in London and Metropolitan areas. When it comes to comparing ‘Medium urban’ with ‘Rural area’, high income groups prefer residing in the latter. This is also the case for manual workers where the nonlinear modelling of built form in LCA-SEM shows they mainly live outside ‘Dense urban’ areas but there is no evidence for specific preferences from the model results between ‘Medium urban' and 'Rural areas'. These nonlinear patterns cannot be captured from the path-diagram based SEM.

5.3.3. Influences on car ownership, distance and time travelled

Table 23 to Table 25 report the results of the influences upon car ownership, travel distance, and travel time across built form latent classes respectively. The incorporation of the LCA provides a unique opportunity to decompose more precisely the influences both for each of the demographic and socioeconomic variables and across the different built form classes. Furthermore, to identify the additional insights of incorporating a categorical built form variable in the SEM model, we compare results from LCA-SEM with those from a constrained SEM where the model parameters do not vary across the built form classes. This constrained SEM is typical of the existing models that do not account for the specific influences of the built form characteristics.

The model intercepts and coefficients can help to quantify the levels of influences of the demographic and socioeconomic variables in the context of the built form latent classes. Whilst an intercept represents the average level of car ownership, travel distance or travel time of the Reference Group (To aid intuitive interpretation of the model outputs, in all tables we define a Reference Group of residents who are female, part time working in white collar clerical occupations from a car-owning household with more than one adults and a household income of 25-50k per year), the coefficients indicate how much influence a change in the demographic and socioeconomic profiles has. The rest of the model results provide opportunities to compare the car ownership, travel distances or travel times both within each column (i.e. holding the built form class constant and decompose the influences of demographic, socioeconomic and car ownership characteristics) and across the columns for each row (i.e. to identify the influence of the built form given a particular demographic, socioeconomic and car ownership profile). Wald test p-value shows whether there is statistically significant differences between the two extreme built form classes (i.e. 'Dense urban' and Rural areas) for each socioeconomic covariate. Given the standard errors, the Wald test compares the coefficients of Dense urban with those of Rural areas for each case. Here we assume that significant difference exists (i.e. we can reject the hypothesis of 0 difference) when the p-value is less than 0.01.

Note that the values for the demographic, socioeconomic and car ownership variable rows are additive within each column for travel distance and travel time. This allows the readers to work out the specific distances or time travelled for an arbitrary type of resident. For car ownership which is estimated by probit regression, the interpretation is not as straightforward. The increase in probability attributed to one-unit increase in a given predictor is also dependent on the reference value of that predictor. This is because the link function for the probit model follows a nonlinear distribution function of the standard normal.

Table 23 shows the influences on car ownership across built form latent classes. Reassuringly, the results from constrained model are very similar in both magnitude and directional effect of influences to the direct effects of socioeconomic characteristics reported from path-diagram based SEM (cf .Panel 6b of Table 8). In particular, 1 adult households, manual workers, and lowest income groups are more likely to have no car when compared with the reference group, whilst skilled manual, professionals and high income groups are more likely to have a car in their household.

The first line of the model outputs reports the model intercept values which show significant differences in the level of car ownership across built form classes for the reference group.

The main benefit of LCA-SEM is the insights we get from comparison across built form clusters which demonstrate significant variations for some socioeconomic groups[35]. 1 adult households are more likely to have no access to car. The Wald test suggests that for household size, the difference in the influences on car ownership between the first and the third built form class is statistically significant with a bigger gap in the level of car ownership in more rural areas. There is also evidence of significant variations across built form clusters for low income households and manual and skilled manual workers. The difference in car ownership between manual workers and white collar clericals (reference group) is bigger in 'Rural areas' than that in 'Dense urban' area; for skilled manuals, however, this difference is bigger in 'Dense urban' area when compared to that in 'Rural areas'. This suggest that this is the white collar clericals who are more prepared to forgo their cars by living in more dense urbanized areas. The difference in car ownership between low, high and medium income households is also significantly larger in 'Rural areas' when compared to the more urbanized areas again due to

[35] Although we take into account the self-selection influence by looking into variations across built form clusters for a particular socioeconomic group, we still cannot claim we are measuring causality effects. It is hard to say whether that particular socioeconomic class is influenced by built form characteristics of their residential area or that they have chosen to live in a particular area due to their preferences for particular travel patterns. The two-level SEM that I discuss in section 5.4 should be able to shed more lights on this issue.

the willingness of the higher income households to have a lower level of car ownership when living in denser areas.

Table 23 Influence of socioeconomic profile on car ownership

Direct effect	Constrained model	1-Dense urban	2-Medium Urban	3-Rural areas	Class 1 vs Class 3 wald test p-value
Model threshold for the reference group, which is represented by a female, part time working white collar clerical worker from a car-owning household with more than one adults and a household income of 25-50k per year		0.618***	1,673***	2.348***	
Male	-0.021	0.022	-0.038	-0.069	0.107
Full Time working	0.009	0.017	0.014	-0.026	0.534
1 adult households	0.524***	0.439***	0.518***	0.714***	0.000
Manual workers	0.416***	0.341***	0.418***	0.579***	0.016
Skilled manual workers	-0.268***	-0.397***	-0.266***	0.050	0.000
Professionals	-0.305***	-0.298***	-0.343***	-0.175**	0.255
Household income less £25k	0.537***	0.317***	0.635***	0.573***	0.003
Household income more than £50k	-0.233***	-0.209***	-0.314***	-0.441***	0.078

*** significant within 99% CI, ** significant within 95% CI, * significant within 90% CI

Table 24 shows the influence on distance travelled for different purposes across the latent built form classes. The first line of the model outputs in Panel 24a reports how this group differ in their average weekly commuting distances among the three built form classes through the model intercept values: those live in the 'Dense urban' areas travel 10.4 miles per week, in 'Medium urban' 9.6 miles, and in 'Rural areas' 13.59 miles. Similarly, the first lines under Panel 24b and 24c in Table 24 show that for shopping and other travel purposes, the more rural the area, the longer the distances travelled which is intuitive. As expected, the reference group residents commute well below the working adult average of 30.3 miles per week for all classes of areas, but for shopping and other travel (for which the average weekly distances travelled are respectively 11.3 and 72.9 miles) they travel shorter than the average in more urban areas and longer in the rest (cf Table 3).

The general patterns of small coefficients for the 'Dense urban' class (i.e. relative to its model intercept), and the large ones for the other two built form latent classes indicates that the influence of the built form on travel is relatively strong in the 'Dense urban' class; this influence is much weaker in areas of the other two classes relative to that of demographic and socioeconomic profiles.

For instance, the coefficient for high income households (Households with income more than £50k) in the 'Dense urban' class is 2.1, which shows that by virtue of the higher income, such commuters travel 2.1 km more relative to the Reference groups' intercept of 10.39km, or 20.2% more. By contrast, commuters from high income households in medium urban and rural areas travel respectively 54.2% (coefficient 5.2 divided by intercept 9.6) and 34.7% (4.71/13.59) more. This pattern is mirrored by the commuting distances for commuters from households with less than 25k income per year. Similarly, households with no cars in London travel only 23.7% less (-2.46/10.39), whilst those in 'Medium urban' and 'Rural areas' respectively 60.3% (-5.79/9.6) and 68.1% (-9.25/13.59) less.

The results are intuitively correct and they provide a substantially more robust set of quantifications of the influences upon distance travelled by working adults. For instance, existing models suggest that those households with no cars tend to travel much shorter distances than those with cars. However, when we take account of the latent built form classes, then we see considerable variability than suggested by the existing models: In the 'Dense urban' areas, those with cars only commute slightly more (2.46 miles per week or 8% of the national average) than those without cars. In 'Rural areas' the corresponding value is 3.7 times higher, or 9.25 miles more per employed person per week.

Table 24 Direct influences on travel distance (in miles) arising from traveller profiles

Direct influence	**Constrained model**	**1-Dense urban**	**2-Medium Urban**	**3-Rural areas**	**Class 1 vs Class 3 Wald Test p-value**
Panel 24a. Direct influences on commuting					
Model intercept for the reference group, which is represented by a female, part time working white collar clerical worker from a car-owning household with more than one adults and a household income of 25-50k per year		10.39***	9.60***	13.59***	
Male	10.66***	6.31***	11.84***	10.84***	0.000
Full time working	16.8***	12.83***	15.96***	20.54***	0.000
1 adult households	2.88***	-0.08	3.64***	4.87***	0.004

Direct influence	Constrained model	1-Dense urban	2-Medium Urban	3-Rural areas	Class 1 vs Class 3 Wald Test p-value
Semi- or unskilled manual workers	-3.13***	-0.35	-3.11***	-5.33***	0.001
Skilled manual workers	-4.4***	0.01	-3.87***	-7.73***	0.000
Professionals	2.68***	3.1***	2.3***	2.71**	0.787
Household income less £25k	-4.32***	-2.32***	-4.53***	-5.18***	0.023
Household income more than £50k	4.45***	2.1***	5.2***	4.71***	0.043
No car in household	-4.6***	-2.46***	-5.79***	-9.25***	0.000
Panel 24b. Direct influences on shopping					
Model intercept for the reference group, which is represented by a female, part time working white collar clerical worker from a car-owning household with more than one adults and a household income of 25-50k per year		7.75***	12.41***	20.36***	
Male	-3.13***	-1.79***	-2.7***	-4.99***	0.000
Full time working	-0.98***	-0.58***	-0.7	-1.5***	0.074
1 adult households	0.69***	0.79***	0.84***	0.43	0.570
Semi- or unskilled manual workers	-1.37***	-0.42	-1.54***	-1.47**	0.176
Skilled manual workers	-1.12***	0.02	-1.26***	-1.43**	0.028
Professionals	-0.02	0.16***	0	-0.25	0.511
Household income less £25k	-0.56***	-0.28***	-0.64***	-0.28	0.989
Household income more than £50k	0.07	-0.28**	0.01	0.47	0.207
No car in household	-3.83***	-2.58***	-4.41***	-7.48***	0.000
Panel 24c. Direct influences on other purposes combined					
Model intercept for the reference group, which is represented by a female, part time working white collar clerical worker from a car-owning household with more than one adults and a household income of 25-50k per year		44.37***	55.99***	79.40***	
Male	15.03***	7.12***	15.55***	19.03***	0.000
Full time working	2.25**	0.85	1.6	4.31***	0.1881
1 adult households	20.33***	18.67***	20.59***	23.74***	0.224
Semi- or unskilled manual workers	-19.72***	-14.11***	-16.67***	-29.05***	0.000
Skilled manual workers	-20.04***	-15.64***	-17.15***	-27.55***	0.000
Professionals	13.82***	6.43**	15.08***	15.14***	0.031
Household income less £25k	-10.13***	-7.78***	-9.18***	-12.14***	0.143
Household income more than £50k	16.88***	14.23***	15.44***	21.45***	0.043
No car in household	-26.06***	-16.06***	-32.13***	-47.42***	0.000

*** significant within 99% CI, ** significant within 95% CI, * significant within 90% CI

Table 25 shows the influence on travel time for different purposes across the latent built form classes. Like previous tables, the first lines in Panel 25a to 25c report how the reference groups' average weekly time for commuting, shopping, and other purposes differ across the three built form classes through the model intercept values. Unlike travel distance which is longer for those residing in ‘Rural areas’, the intercept values in Table 25 suggest that, with the exception of shopping trips, residing in ‘Medium urban’ areas is associated with longer travel in time when compared with the other two latent classes. For instance, those live in the ‘Dense urban’ areas spend 59 minutes per week, in ‘Medium urban’ 81.7 minutes, and in ‘Rural areas’ 40.6 minutes for commuting. This result is in line with those concluded from path-diagram based SEM (cf panel 9a of Table 9) where built form latent variable shows positive influence on commuting time but negative effect on commuting distance. However, through categorizing built forms, we show that the influence is not linear; the commuting time is longer for ‘Medium urban’ areas than that in ‘Dense urban’ areas.

For other purposes, the first look might suggest that the results from Table 25 is in contradiction with those from panel 9c of Table 9 where the negative influence of built form latent variable on travel time indicates longer travel time for more rural residents. Further investigation, however, reveals that this is due to the nonlinearity in the effects. The first row of Panel 25c shows that those in ‘Medium urban’ area spend the longest for travelling (i.e. in average 312.50 minutes per week). However, compared with ‘Dense urban’ area, those in ‘Rural area’ experience longer travel time. This in average result in the negative sign we have observed from Panel 9c. This is another evidence for the benefits in applying LCA in combination with SEM.

The p-values from the Wald tests suggest that the built form influence is larger for commuting trips when compared to shopping and other travel purposes. Among the most significant influences, part–time workers spend more time for commuting in ‘Dense urban’ than ‘Rural areas’. This is also the case for manual and skilled manual workers who spend around 12 minutes (-8.1-3.9) and 9 minutes (-14.9+6) per week more on commuting in ‘Dense urban’ vis-à-vis ‘Rural areas’ respectively. The most striking difference is for car ownership with those with no access to car tend to travel 18.1 minutes and 12.4 minutes longer when residing in ‘Dense urban’ and ‘Medium urban’ areas respectively when compared with their counterparts in ‘Rural areas’.

At the first glance these results might seem unexpected. However, it can be better absorbed by comparing against influences on travel distance reported in Table 24. While typical main stream full time workers tend to live closer to their workplace in denser urban areas, skilled and unskilled manual workers and those with no access to car tend to make shorter commuting distances. This can be explained by the latter groups' reliance on public transport specifically when they reside in denser urban areas[36]. Their significant longer commuting time can then suggest that the alternative travel modes to private cars do not well respond to the requirements of those in more disadvantage situation.

Table 25 Direct influences on travel time (in minutes) arising from traveller profiles

Direct influence	Constrained model	1-Dense Urban	2-Medium Urban	3-Rural areas	Class 1 vs Class 3 Wald Test p-value
Panel 25a. Direct influences on commuting					
Model intercept for the reference group, which is represented by a female, part time working white collar clerical worker from a car-owning household with more than one adults and a household income of 25-50k per year		59.0***	81.70***	40.60***	
Male	11.8***	11.4***	13.6***	9.2***	0.447
Full time working	41.6***	58.6***	36.9***	41.2***	0.000
1 adult households	-1.8	-12.5***	0.7	4*	0.000
Semi- or unskilled manual workers	-3.7***	3.9	-4.1***	-8.1***	0.016
Skilled manual workers	-10.2***	-6.0	-9.2***	-14.9***	0.065
Professionals	1.0	2.7	0.5	0.6	0.605
Household income less £25k	-5.9***	-7.3**	-5.6***	-5.4***	0.631
Household income more than £50k	6.3***	7.3**	6.1***	5.1***	0.575
No car in household	22.1***	26.5***	20.8***	8.4**	0.000
Panel 25b. Direct influences on shopping					
Model intercept for the reference group, which is represented by a female, part time working white collar clerical worker from a car-owning household with more than one adults and a household income of 25-50k per year		45.70***	23.10***	55.20***	
Male	-11.8***	-11.7***	-11.1***	-13***	0.286
Full time working	-4.9***	-3.7***	-4.8***	-5.9***	0.200
1 adult households	2.3***	4.6***	1.2	2.2	0.240
Semi- or unskilled manual workers	-3.8***	-4.0**	-3.6***	-4***	0.980
Skilled manual workers	-3.7***	-4.1***	-2.7***	-5.5***	0.477
Professionals	-0.8	-0.1	-1.3*	-0.5	0.854
Household income less £25k	0.4	1.0	0.4	-0.2	0.507
Household income more than £50k	-0.2	-1.3	0.1	0.1	0.409

[36] House prices are not included in this analysis but they tend to be higher in CBDs of metropolitan and medium urban areas, making it less affordable for disadvantage groups to live close to their workplace when living in denser urban areas.

Direct influence	Constrained model	1-Dense Urban	2-Medium Urban	3-Rural areas	Class 1 vs Class 3 Wald Test p-value
No car in household	0.2	-2.0	2.4*	-1.9	0.993
Panel 25c. Direct influences on other purposes combined					
Model intercept for the reference group, which is represented by a female, part time working white collar clerical worker from a car-owning household with more than one adults and a household income of 25-50k per year		193.30***	312.50***	208.60***	
Male	16.1***	5.8	18.2***	18.6***	0.004
Full time working	-14.3***	-15.5***	-15.6***	-11***	0.459
1 adult households	40***	43.6***	40.5***	39.3***	0.589
Semi- or unskilled manual workers	-42.3***	-42.6***	-36.1***	-52.9***	0.213
Skilled manual workers	-43.3***	-48.2***	-38***	-51.2***	0.664
Professionals	18.9***	10.3*	20.6***	20.1***	0.179
Household income less £25k	-16.6***	-11.7**	-15.1***	-19.4***	0.233
Household income more than £50k	24.9***	23.9***	23***	29.9***	0.368
No car in household	-37.5***	-19.6***	-51.9***	-65.9***	0.000

*** significant within 99% CI, ** significant within 95% CI, * significant within 90% CI

Overall, these results reconfirm the broad thrust of the findings from the path-diagram based SEM that there is significant inequality in mobility, with the fastest growing sector of the employed population being in more disadvantaged situations. This is paradoxically more true in denser urban areas with better provisions of public transport. With LCA-SEM, we show that this phenomena is specifically true for those belonging to the 'Medium urban' group which forms 54% of the population (cf Table 21)

5.4. Findings from the Two-level SEM

The two-level SEM is an expansion of the path-diagram based SEM where the model intercepts are allowed to vary across pre-defined built form clusters. Replacing the built form latent variable with the random intercepts helps identify more precisely the extents of self-selection. We use the 14 population density levels and 8 area types as defined in NTS to form 98 built form cluster categories, the categories containing variously 24 to over 16000 households in the NTS sample. The smallest group is an exceptional cluster in Inner London with a population density of 1 to 5 persons per hectare – such low density is rare in Inner London and it is not surprising that only 24 individuals belong to this cluster. The largest one is comprised by rural areas with a population density of less than 1 individual per hectare, and it has over 16000 individuals. Naturally there are a few density/area type combinations (such as 'high density rural areas') which are omitted from the analysis.

As explained in Section3.5, we first run an SEM that incorporates random intercepts with unrestricted variance (**Model A**). We then run a second SEM where the random intercepts are modelled as a function of household socioeconomic profiles; this is essentially a random intercept model with second level determinants (**Model B**). Figure 11 presents the structure of Model A and Model B. For comparison with the latest SEM model using the NTS data, we run an updated version of the path-diagram based SEM, using the same dataset for 2002-2012 – this single level SEM serves as a benchmark and we name it **Model Benchmark**.

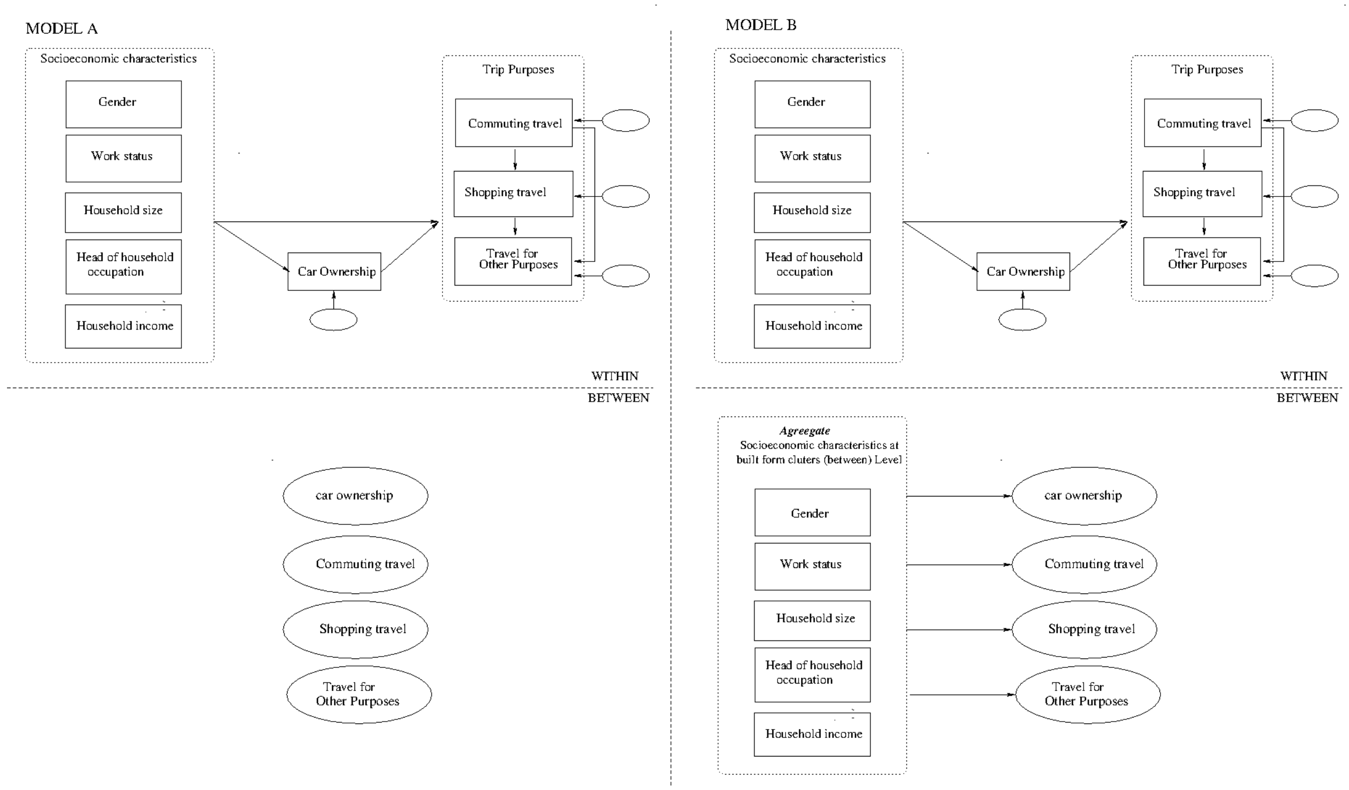

Figure 11 Two-Level SEM with unrestricted (Model A) and restricted (Model B) random intercepts

We examine the model results in three steps. First, we examine the goodness of fit of the models. Second, we explore effects of introducing random intercepts for car ownership, travel distance and travel time through Model A. Third, we examine the random intercepts as a function of socioeconomic profiles of the households and differentiate the influences of self-selection and spatial sorting from those specifically associated with the built form categories. Finally, we examine any effects over time by comparing the data in two periods – before and after 2007 – to see how the influences have evolved over time.

5.4.1. Model goodness of fit

Table 26 compares the goodness of fit statistics of three alternative models. It shows that Models A and B fit much better to the observed data – the sharp reductions of AIC, BIC and ABIC are a little unexpected but a vindication of the introduction of the random intercepts. Modelling the random intercepts as a function of household socioeconomic variables continues to improve the fit, although the majority of gains are achieved when the random intercepts are introduced under Model A.

Table 26 Goodness of fit: Models A and B vs Model Benchmark

	AIC	BIC	ABIC
		Travel distance model	
Model Benchmark	2,472,702	2,473,785	2,473,419
Model A	1,624,785	1,625,511	1,625,266
Model B	1,624,510	1,625,338	1,625,059
		Travel time model	
Model Benchmark	1,657,683	1,658,766	1,658,401
Model A	803,521	804,255	804,007
Model B	803,361	804,189	803,910

5.4.2. The influences of household socioeconomic profiles

First, we examine the influences of household socioeconomic profiles as estimated by the random intercept model relative to Model Benchmark.

Table 27 and Table 28 show the influences of household socioeconomic profiles upon car ownership after controlling for the built form characteristics of the residential areas – the model coefficients (including the nonsignificant coefficients) which are reported from the travel distance model (i.e. Table 27) , and the equivalent set of coefficients from the travel time model (i.e. Table 28) are reassuringly similar . It can be observed from these tables that the coefficients from Model A are in fact quite similar to the Model Benchmark. As with regression models for categorical variables, one category per set is left out so that the coefficient estimation can treat it as the reference category (which is reported in the right most column of Table 27 and Table 28)

Table 27 Influence of household socioeconomic profile on car ownership after controlling for built form categories (from travel distance model)

Household socioeconomic variables	Model A coefficient	Model Benchmark coefficients	Reference variable for model
Male	0.017	0.026	Female
Full time working	0.022	0.018	Part time working
1 adult households	0.478***	0.484***	>1 adult households
Manual workers	0.366***	0.373***	White collar clerical
Skilled manual workers	-0.030	-0.035	White collar clerical
Professionals	-0.150***	-0.135***	White collar clerical
Household income less £10k	0.545***	0.540***	Income 25-30k
Household income £10k to £15k	0.469***	0.476***	Income 25-30k
Household income £15k to £20k	0.340***	0.341***	Income 25-30k
Household income £20k to £25k	0.128***	0.134***	Income 25-30k
Household income £30k to £35k	-0.062	-0.065	Income 25-30k
Household income £35k to £40k	-0.111***	-0.102***	Income 25-30k
Household income £40k to £50k	-0.152***	-0.130***	Income 25-30k
Household income £50k to £60k	-0.238***	-0.225***	Income 25-30k
Household income more than £60k	-0.319***	-0.264***	Income 25-30k

*** significant with 99% confidence interval; as a rule we only report model coefficients that are significant with 99% confidence interval, because we have more than 90,000 individuals in the dataset and thus a large degree of freedom for within-level model. The same principle is applied throughout this paper.

Table 28 Influence of household socioeconomic profile on car ownership after controlling for built form categories (from travel time model)

Household socioeconomic variables	Model A coefficient	Model Benchmark coefficients	Reference variable for model
Male	0.017	0.026	Female
Full time working	0.022	0.018	Part time working
1 adult households	0.479***	0.484***	>1 adult households
Manual workers	0.367***	0.373***	White collar clerical
Skilled manual workers	-0.032	-0.035	White collar clerical
Professionals	-0.149***	-0.135***	White collar clerical
Household income less £10k	0.545***	0.540***	Income 25-30k
Household income £10k to £15k	0.471***	0.476***	Income 25-30k
Household income £15k to £20k	0.340***	0.341***	Income 25-30k
Household income £20k to £25k	0.129***	0.134***	Income 25-30k
Household income £30k to £35k	-0.063	-0.065	Income 25-30k
Household income £35k to £40k	-0.111***	-0.102***	Income 25-30k
Household income £40k to £50k	-0.152***	-0.130***	Income 25-30k
Household income £50k to £60k	-0.238***	-0.225***	Income 25-30k
Household income more than £60k	-0.320***	-0.264***	Income 25-30k

*** significant with 99% confidence interval; as a rule we only report model coefficients that are significant with 99% confidence interval, because we have more than 90,000 individuals in the dataset and thus a large degree of freedom for within-level model. The same principle is applied throughout this paper.

Table 29 presents the influences of household socioeconomic profiles and car ownership upon travel distances and travel times for each of the three trip purposes from Model A. Again the Model A coefficients are generally very similar to those from the Model Benchmark. In Table 29 we present the coefficients in miles and minutes – e.g. a coefficient of 10.8 for the travel distance of variable 'Male' implies that all being equal a male worker commutes 10.8 miles more per week relative to females. Looking across the variables, it is not surprising that workers from lower income occupations and lower household incomes travel less (both in distance and time). For commuting (Panel 29a), the most striking difference is between full- and part-time workers – full time workers commute 16.6 miles and 41.9 minutes longer than part timers..

For shopping (Panel 29b), not having a car implies shorter travel distances (by 3.4 miles). Males on average travel 3.1 miles and spend 11.7 minutes less than females. Working full time implies less shopping travel, although the influence is well less than half the influence of gender.

For other travel (Panel 29c), not having a car imply shorter travel distance and less travel time. Males on average travel 14.6 miles and spend 15.4 minutes more than females.

Table 29 Fixed influences on travel distance and times arising from traveller profiles

Direct influence	*Model A - Travel Distance (miles)*	*Model A - Travel Time (minutes)*	*Model Benchmark - travel distance (miles)*	*Model Benchmark - travel time (minutes)*	*Reference variable*
			Panel 29a. Direct influences on commuting		
Male	10.8***	12.4***	10.7***	12.7***	Female
Full time working	16.6***	41.9***	16.5***	41.7***	Part time working
1 adult households	3.2***	-1.6	3.2***	not significant	>1 adult households
Semi- or unskilled manual workers	-3.3***	-3.6**	-3.2***	-3.6***	White collar clerical
Skilled manual workers	-4.8***	-13.5***	-4.7***	-13.2***	White collar clerical
Professionals	2.4***	0.00	2.3***	not significant	White collar clerical
Household income less £10k	-1.7	-3.7	not significant	not significant	Income 25-30k
Household income £10k to £15k	-4.7***	-6.4***	-4.7***	-6.5***	Income 25-30k
Household income £15k to £20k	-2.2***	-0.7	-2.2***	not significant	Income 25-30k
Household income £20k to £25k	-0.6	-1.3	not significant	not significant	Income 25-30k
Household income £30k to £35k	0.8	0.6	not significant/	not significant	Income 25-30k
Household income £35k to £40k	2.0***	2.2	2.0***	not significant	Income 25-30k
Household income £40k to £50k	4.4***	5.4***	4.5***	6.3***	Income 25-30k
Household income £50k to £60k	5.0***	5.4***	5.1***	7.1***	Income 25-30k
Household income more than £60k	7.5***	9***	7.6***	11.8***	Income 25-30k
No car in household	-4.7***	19.6***	-4.4***	20.3***	With car in household
			Panel 29b. Direct influences on shopping		
Male	-3.1***	-11.7***	-3.0***	-11.6***	Female
Full time working	-1.1***	-5.2***	-0.9***	-5.1***	Part time working
1 adult households	0.8***	2.8***	0.8***	2.8***	>1 adult households
Semi- or unskilled manual workers	-1.3***	-3.4***	-1.4***	-3.5***	White collar clerical
Skilled manual workers	-1.0***	-3.8***	-1.1***	-3.8***	White collar clerical
Professionals	0	-1	not significant	not significant	White collar clerical
Household income less £10k	-0.98***	0.1	-0.9***	not significant	Income 25-30k
Household income £10k to £15k	-0.91	0.4	-1.0***	not significant	Income 25-30k
Household income £15k to £20k	-0.78	-0.1	-0.8***	not significant	Income 25-30k
Household income £20k to £25k	0.25	1.9	not significant	not significant	Income 25-30k
Household income £30k to £35k	0.34	0.8	not significant/	not significant	Income 25-30k
Household income £35k to £40k	0.13	0.3	not significant/	not significant	Income 25-30k
Household income £40k to £50k	0.57	1	not significant/	not significant	Income 25-30k
Household income £50k to £60k	0.49	0.9	not significant/	not significant	Income 25-30k
Household income more than £60k	0.4	-0.2	not significant/	not significant	Income 25-30k
No car in household	-3.4***	-0.4	-2.7***	not significant	With car in household
			Panel 29c. Direct influences on other purposes combined		
Male	14.6***	15.4***	14.3***	15.2***	Female
Full time working	2.1	-14.3***	2.4***	-14.5***	Part time working
1 adult households	20.8***	43.5***	21.0***	44.1***	>1 adult households
Semi- or unskilled manual workers	-19.7***	-43.8***	-19.7***	-44.6***	White collar clerical
Skilled manual workers	-18.8***	-44.1***	-19.2***	-44.9***	White collar clerical
Professionals	14.0***	20.0***	13.5***	19.5***	White collar clerical
Household income less £10k	-9.44***	-14.9***	-9.1***	-14.4***	Income 25-30k
Household income £10k to £15k	-10.87***	-17.9***	-10.9***	-18.1***	Income 25-30k
Household income £15k to £20k	-6.8***	-11.1***	-6.7***	-11.4***	Income 25-30k
Household income £20k to £25k	-3.92***	-4.4	-3.8***	not significant	Income 25-30k
Household income £30k to £35k	3.00	5.9	not significant/	not significant	Income 25-30k
Household income £35k to £40k	4.37***	9.9***	4.4***	10.0***	Income 25-30k
Household income £40k to £50k	7.60***	13.6***	7.7***	14.3***	Income 25-30k
Household income £50k to £60k	13.85***	22.3***	14.0***	22.9***	Income 25-30k
Household income more than £60k	24.0***	38.6***	24.2***	40.9***	Income 25-30k
No car in household	-22.0***	-28.1***	-19.1***	-22.7***	With car in household

******* *significant with 99% confidence interval*

The interactions between different purposes of travel are also very similar to those from Model Benchmark across the SEMs, showing significant negative influence of commuting time on shopping travel time as well as significant influence of commuting and shopping on other purposes, in the context of both travel distance and time.

The comparison of the model results with Model Benchmark shows that whilst Model A have a better fit to data after introducing the random intercepts, its actual coefficient values have only changed fairly slightly. However, Model A provides the foundation for exploring the variations across the built form clusters which we now turn to below.

5.4.3. Self-selection vs intrinsic built form effects

The simple formulation of the random intercept model under Model A outputs the mean and variance of the intercepts for car ownership, which are shown in Panel 29a of Table 29. Model B defines the random intercepts as a function of household socioeconomic profiles. This is a way to control specifically for self-selection and spatial sorting of the residents when quantifying the intrinsic influences of the built form.

Table 30 Panel 30b further reports statistically significant influences of the socioeconomic variables on the variation in intercept for car ownership, and the residual variance (i.e. the variation which cannot be explained by socioeconomic variables) at built form level. It would seem sensible in the context of the NTS dataset to attribute this residual variance to built form effects and this is what we will be doing below. It is possible that the residual variance will cover other influences that the control variables from the NTS dataset cannot distinguish – this is an issue that we will return to in Chapter 6.

The substantial overall variance of the car ownership intercept in panel 30a (0.216) confirms that the level of car ownership varies across built form categories. Further, after controlling for the influences of socioeconomic profiles, the residual variance is considerably reduced – from 0.216 for both travel distance and travel time models to 0.046 and 0.034 respectively. The ratios of the residual variances to the overall variances (0.046/0.216= 21% and 0.034/0.216=16%) are the share of influence of the built form categories that is not explained by self-selection and spatial sorting of the households- these are within the wide range of 2 to 66 percent suggested in the existing literature (Mokhtarian and van Herick, 2016). In other words, 79% (1-21%) of the variance as shown in the travel distance model and 84% (1-16%) in the travel time model is explained by household socioeconomic profiles.

Out of the range of socioeconomic variables, the most statistically significant variables are the proportions of 1 adult households and of professional and skilled manual workers (panel 30b). In other words, around 80% of the observed outcomes of the much lower proportions of car owners residing in dense urban areas could be attributed to the fact that there are more 1 adult households living there who have a considerably lower levels of car ownership, and the fact that households with skilled manual and professional workers who have high car or van ownership tend to live in less dense, rural areas. We have already suspected this from the findings in path-diagram based SEM and LCA-SEM, but through random intercept SEM, we are able to provide an unambiguous quantification of the effects.

Table 30 Between-built form variations in car ownership and the main underlying influences

Influences on intercept		Travel distance model	Travel time model
Panel 30a car ownership random intercept threshold and variance - Model A			
Intercept			
	Threshold	1.17***	1.17***
	Variance	0.216***	0.216***
Panel 30b Influences on car ownership random intercept - Model B			
Percentage of			
	1 adult households	6.00***	5.9***
	Skilled manual workers	-3.38***	-2.62***
	Professional	-1.99***	-1.55**
	Household income less £25k	Not significant	-1.20*
Residual variance		0.046***	0.034***

*** significant level within 99% ,** 95%, * 90% -

Table 31 presents similar information for commuting (panel 31a), shopping (panel 31b), and combined other travel purposes (panel 31c). Within each panel, the upper parts (i.e. panel 31a-1, 31b-1, and 31c-1) shows the means and overall variances, meanwhile the lower parts present all those statistically significant influences on the variations of the intercepts for the travel distance and travel time outcomes. The ratio of the residual variance from Model B to the overall variance from Model A implies the percentage of travel outcome variations across the built form categories that are not explained by the household socioeconomic profiles. The outputs from Table 31 show that after controlling for the household socioeconomic profiles, respectively 54%, 43% and 53% of the travel distance variations in commuting, shopping, and other travel can be attributed to intrinsic built form characteristics. Similarly, the percentages of travel time variations that can be attributed to built form characteristics are 75%, 43% and 77% respectively for commuting, shopping and other travel.

Table 31 Between-level influences on travel

Influences on intercept		Travel distance model coefficient	Travel time model coefficient
Panel 31a modelling commuting random intercept			
Panel 31a-1 commuting random intercept mean and variance _Model A			
Intercept	Mean	7.8***	50.6***
	Variance	10.3***	330.5***
Panel 31a-2 Influences on commuting random intercept- Model B			
Proportion of	1 adult households	-35.06**	18.49***
	Skilled manual workers	Not significant	-6.38**
	Full time workers	-43.70***	Not significant
Residual variance		5.6***	246.5***
Panel 31b modelling shopping random intercept			
Panel 31b-1 shopping random intercept mean and variance-Model A			
Intercept	Mean	12.45***	48.7***
	Variance	6.5***	6.6***
Panel 31b-2 Influences on shopping random intercept-Model B			
Proportion of	Full time workers	-16.7***	Not significant
	1 adult households	-14.6**	-31.1***
	Skilled manual workers	22.7***	24.0***
	Professional	17.36***	33.3***
	Household income less £25k	Not Significant	21.2**
Residual variance		2.8***	2.8***
Panel 31c modelling Other purposes random intercept			
Panel 31c-1 Other purposes random intercept mean and variance-Model A			
Intercept	Mean	55.19***	192.1***
	Variance	125.6***	133.0***
Panel 31c-2 Influences on other purposes' random intercept- Model B			
Proportion of	1 adult households	-121.4***	Not significant
	Professional	105.2***	140.9**
	Household income more than £50k	-66.5***	Not significant
Residual variance		66.7***	102.3

*** significant within 99% interval ,** 95%, * 90% .

Similar to the car ownership model, the significant household socioeconomic influences on travel outcomes are 1 adult households, skilled manual workers, professional workers, etc, which indicate substantial extents of self-selection and spatial sorting effects.

5.4.4. Within and between built form influences pre- and post-2007

This Section reports the findings from subdividing the NTS data into two subsets: 2002-2006 and 2008-2012. The model estimation is carried out through a multi-group SEM where the influences are allowed to vary between the two time periods. This roughly doubles the number of unknown parameters to be estimated. To optimise the model runs, we run the multi-period model in two stages: First, we run the model only for car ownership (cf Table 32) to identify any highly significant between-level socioeconomic variables. we then run the travel distance and travel time models for each travel purposes with the best between-level car ownership model (See Table 33 below).

The car ownership model shows no significant coefficient changes pre- and post-2007. However, the variance for car ownership intercept (between built form variation) has increased by 36% after 2007 (panel 32a). This corroborates the findings in Path-diagram based SEM and the Model Benchmark which show an increase of 22% in the effect of built form latent variable on car ownership. Controlling specifically for variations across built form categories in a multi-period model shows a stronger growth in forgoing car ownership in more dense built-up areas after 2007.

Panel 32b shows that this increased influence on car ownership in dense urban areas is not due to household socioeconomic profiles. In other words, the role of self-selection effect is minimal in explaining this increase, and the influences would appear to have come from changing built form characteristics.

Table 32 Summary of significantly changed influences pre- and post-2007 for between-level model car ownership

Influences on intercept		**Coefficients and (p-value) pre 2007**	**Coefficient and (p-value) post 2007**	**% change in coefficient values**
	Panel 32a car ownership free random intercept variance			
	Intercept variance	0.196***	0.266***	36%***
	Panel 32b Influences on car ownership random intercept			
Proportion of				
	1 adult households	5.21***	6.41***	Not significant
	Skilled manual workers	-2.87***	-3.10***	Not significant

*** significant with 99% interval ,** 95%, * 90% .

Travel distance and travel times show no significant between-level changes across the two periods. The analysis of within built form influences (cf Table 33) confirms the findings by

Path-diagram based SEM and Model Benchmark. First, the gender gap appears to be narrowing after 2007 for travel distance and time. Our analyses of the NTS data show that for shopping trips, it is the females who have reduced their travel distance; for commuting and other trips, it is the males' reduced travel that narrowed the gap.

Second, in line with the trends identified above, the gap in travel distance between the low and the middle income group is no longer significantly different from 0 after 2007. This is because the rate of drop in travel distance for middle income group has been higher vis-à-vis that for the low income.

Thirdly, the gap in travel time of manual workers to work appears to have disappeared after 2007. This is a two-sided effect involving on the one hand increases in manual workers' commuting time and decrease in commuting time of other occupation groups (cf. Table 4).

Finally, it is useful to note that very small number of significant changes (which are all reported above) indicates that the majority of the influences remain remarkably stable over time. For instance, the large differences between full- and part-time working in terms of commuting distance and time have not changed, in spite of the rapid rise in part-time and free-lancing work, and in the spread of ICT usage.

Table 33 Summary of significantly changed influences pre- and post-2007 (intra-built form level)

Influences	Coefficients and (p-value) pre 2007	Coefficient and (p-value) post 2007	% change in coefficient values
Panel 33a Travel distance analysis			
Male->HBW	11.47***	9.86***	-14%***
Male->Sh	-3.48***	-2.66***	24%***
Male-> Oth	16.49***	8.26***	-50%***
IncomeLess25k->HBW	-2.76***	-0.78	-
Skilled Manual->Sh	-1.42***	-0.48	-
Panel 33b Travel time analysis			
Manual->HBW	-4.47***	-1.61	-
Male->Sh	-14.27***	-9.61***	33%***
Male-> Oth	15.48***	1.23	-

*** significant with 99% confidence interval

6. Conclusion

This dissertation sets out to investigate the influences of the built form on multiple travel outcomes of employed adults in the UK, controlling for a comprehensive range of demographic, socioeconomic and car ownership attributes as recorded by the UK National Travel Survey (NTS). In particular, the research objectives are set for: a) quantifying the built form characteristics in explaining travel behaviour after controlling for self-selection and spatial sorting, car ownership endogeneity, and interactions among travel purposes; b) measuring the scale of changes in travel behaviour of employed adults in the context of an apparent drop in aggregate travel distance and time since 2007[37].

Apart from understanding the trend of changes in influences, objective (b) is motivated by current interest in understanding the drop in aggregate travel distance and time in recent years. For a few decades, it was a common wisdom that per person trip rates and travel time tend to be stable, whilst the per person travel distance rises over time. However, data from the UK National Travel Survey seems to suggest that there has been a major deviation from this trend in recent years. The invariance in the average trip numbers and travel time in conjunction with steady growth in the average travel distance did hold up to 2002; however, from 2002 to 2007 the average distance travelled per person in the UK has flatlined and since 2007, it has started declining (Melbourne, 2012, Jones and Le Vine, 2012). This apparent reversal of a long term trend poses an important research question that has triggered wide discussions. Exploring the cause for changes in aggregate travel patterns is outside the scope of this dissertation as it requires more data to be able to make firm conclusions. In addition, it should include the whole sampled population and separate analysis of travel modes. However, evaluating changes in influences over time specifically before and after the recent recession can underline some potential explanations which might contribute to the understanding of the shift in aggregate trends - e.g. the change in travel behaviour within specific population group or area.

Three sets of new SEM models have been developed to respond to these research questions. The model results provide a range of new insights that fill a gap in the existing literature,

[37] Refer to the NTS report published on line about drop in travel time and distance since 2007 - https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/457752/nts2014-01.pdf
The site is accessed on 16 March 2016

particularly through employing novel SEM features such as incorporating the latent categorical analysis or developing SEMs with random intercepts.

The findings from the three SEMs broadly corroborate one another, and reveal insights into the use of different methods. Moreover, each model adds its unique insights which broaden some aspects of our understanding of built form influences on travel patterns. The path-diagram based SEM highlights the importance of direct and indirect influences of the built form through a built form latent variable. It also highlights the endogeneity of car ownership and interactions among travel purposes. LCA-SEM reconfirms the findings from the path-diagram based SEM but in addition identifies the form and extent of variations in travel outcomes across some tangible built form clusters. This helps engage with transport and urban planners. Finally, the two-level SEM helps measure more precisely the extent of the built form influence through more specific controls for self-selection and spatial sorting across built form clusters.

The rest of this chapter is structured to respond to the two main objectives of this dissertation: Section 6.1 provides the response to the first research question by highlighting what we have learnt about built form influences on travel behaviour; Section 6.2 summarises the findings from comparing travel behaviour over time and also more specifically before and after 2007; Section 6.3 discusses the policy implications of the findings; and finally Section 6.4 considers the strengths and weaknesses of the findings and in that context, possible future research directions.

6.1. Built form influences

6.1.1. Built form effects after controlling for endogeneities

Conditioning on demographic, socioeconomic and car ownership characteristics of the households and individuals recorded in the NTS, all three models show statistically significant influences of built form characteristics. The path-diagram based SEM results on the direct influences corroborate the existing literature. After controlling for interactions with socioeconomic variables, not only does the influence of built form on car ownership remain highly significant, it is also the strongest influence among all direct influences upon car ownership. The effect of built form on travel distance, time, and trip frequency for different purposes of travel also remains significant after controlling for the main interactions among explanatory variables including self-selection and spatial sorting. Denser, more built-up areas are associated with lower level of car ownership and shorter travel distance and time for all travel purposes except for commuting. The striking message here however is the relatively high commuting times despite relatively short commuting distances in dense urban areas. we will return to this when we discuss the self-selection effects and policy implications in Sections 6.1.2 and 6.3 below.

The significant influence of built form characteristics are reconfirmed from LCA-SEM where built form is modelled as a categorical latent variable and from the two-level SEM where the intercepts are allowed to vary across built form clusters.

The latent categorical analysis reveals three distinct built form categories in the UK: Dense urban, Medium Urban, and Rural areas. The latent classes are defined based on a specific combination of the built form characteristics which provides the insights into their joint influences upon travel decisions. Our findings confirm that the built form characteristics remain an important influence upon car ownership, the distances travelled, and the time spent for travelling even after controlling for the endogeneities. This is evidenced by strong variations in the models' intercepts in addition to the variations in influences upon travel outcomes across built form latent classes. The advantage of LCA, however, is its capability in capturing travel behaviour variations across distinct built form categorise as opposed to estimating average influences with linearity assumption (i.e. as we did in the path-diagram based SEM). Through latent categorization we can get deeper insights on accessibility and mobility discussed above. The comparison of average travel distance and time for the Reference Group (see definition in Section 5.3) by travel purpose confirms the findings from the path-diagram based SEM that those who live in rural, sparsely populated areas tend to travel longer distances for all travel purposes. For travel time, however the results are not as linear as reported in path-diagram based SEM. Those in 'Medium urban' areas experience longer travel time per week for commuting and for other purposes combined when compared with their counterparts residing in 'Dense urban' and 'Rural areas'. The inhabitants of 'Dense urban' areas experience longer travel time for commuting but shorter travel time for other purposes when compared with those living in 'Rural areas'. These differences cannot be quantified from one single latent built form coefficient such as reported in the path-diagram based SEM.

The significant between-level residual variance in the two level SEM corroborates the influence that can be attributed to the built form differences. Additionally, its comparison with variance of total random intercept outlines a more precise picture of built form influences as opposed to self-selection effects arising from between-level socioeconomic influences. After controlling for the household self-selection and spatial sorting among the built-up areas, around 20% of the car ownership and respectively 54%, 43% and 53% of the travel distance variations in commuting, shopping, and other travel can be attributed to the intrinsic built form characteristics in the UK. The percentages of travel time variations that can be attributed to built form characteristics are respectively 75%, 43% and 77%. Not only does the two-level SEM reconfirm our earlier findings, but also it shows the more specific extents to which travel behaviour can be explained by built form characteristics. we will discuss this further in Section 6.1.2 below.

6.1.2. Built form influences by socioeconomic group and car ownership

Variations in travel behaviour across built form areas are not constant for all socioeconomic groups. This self-selection and spatial sorting effect is evidenced from significant direct influences of socioeconomic characteristics on the built form latent variable, variations of socioeconomic classes across built form categories in the LCA-SEM, and significant between-level influences in the two-level SEM. Self-selection and spatial sorting effects plus the influences arising from car ownership endogeneity exert noticeably different influences among socioeconomic influences, especially on commuting.

The results from the path-diagram based SEM highlight considerable negative effects for single adult and low paid workers after including the influences from indirect effects via built form latent variable and car ownership. Albeit commuting shorter distances, single adults, due to their tendency for living in denser areas and forgoing the car, in average travel 18.7 minutes longer, which is 21% above the average UK commuting time. This is also the case for manual workers and lower income groups who commute respectively 14% less in distance but 5% more in travel time, and 20% less in distance but 8% more in travel time relative to the UK averages. The model results also highlight striking differences between full- and part-time workers which is reinforced by the indirect effects.

The LCA-SEM adds more insights through categorizing the built forms. Full time workers are 11% and 1 adult households are 61% more likely to live in ‘Dense urban’ areas and both groups are 12% less likely to live in ‘Rural’ areas- compared to ‘Medium urban’ areas. This trend is reversed for skilled manual and professionals who prefer living in ‘Rural areas’. One interesting result is unskilled manual workers who tend to live in ‘Medium urban’ more than

‘Dense urban’. When it comes to commuting mobility and accessibility, we can observe some nonlinear effects which cannot be measured from the path-diagram based SEM. Full time workers and those with no access to car have the most striking variations across built form clusters. Full time workers commute longer distances compared to part timers. This difference is larger for those living in less dense, more rural areas. However, when it comes to commuting time, by residing in ‘Dense urban’ area, full timers spend respectively 22 and 18 minutes more than their counterparts in ‘Medium urban’ and ‘Rural areas’. Those with no access to car tend to live closer to their workplace. Their difference in their commuting distance with car owners, however, varies from 2.5 miles when they live in ‘Dense urban’ areas to around 9 miles when they reside in ‘Rural areas’. However, they do not see much benefit in terms of commuting mobility as they spend 27 and 20 minutes more for commuting by living in ‘Dense urban’ and ‘Medium urban’ area compared to car owners. This difference is only 8.4 minutes in ‘Rural areas’ where they tend to live much closer to their workplace.

The comparison of the two-level SEM with the path-diagram based SEM shows that whilst the former has better fit to data after introducing the random intercepts, its actual results for within level socioeconomic influences have only changed fairly slightly. However, the two level SEM provides the foundation for exploring the variations across the built form clusters which provides new insights on the socioeconomic groups with substantial self-selection and spatial sorting tendencies. According to the two level SEM, 80% of observed variations in car ownership across built form clusters can be explained by self-selection. The proportion of 1 adult households is associated with lower car ownership in more dense urbanized area while that of skilled manual and professionals are linked with higher rate of car owning in less dense, more rural areas. Similar to the influences on car ownership, the significant between-level household socioeconomic influences on travel outcomes are related to 1 adult households, skilled manual and professional workers and to lesser extent income and working status. However, as explained in Section 6.1.1, the influence of self-selection and spatial sorting on travel time and travel distance tend to be much lower than that of car ownership. For car ownership, the influence of the built form shows to have increased over time which we will turn to in Section 6.2 below.

The new insights we discuss above underline direct policy implications on land use planning and urban design and the need to address urgently the large mobility disadvantage among the fastest growing segments of workers. This points to the need for making public transport better suited to their needs, enhancing flexible demand-responsive services and coordination between transport and urban development, specifically in provincial cities under the ‘Medium urban’ category- this is where 54% of the employed residents live. We will discuss policy implications further in Section 6.3.

6.2. Changes travel behaviour over time

This dissertation further investigates changes in travel behaviour over time. The statistical tests suggest that there is no strong evidence in favour of the model which allows variations in the influences over time. However, a few influences have significant changes over time, mainly with the trend break from 2007.

We further look into trend breaking influences before and after 2007 through multi group SEM which subdivides the NTS data into two subsets: 2002-2006 and 2008-2010 and allows the coefficients to vary across the two time periods. Changes in travel behaviour are measured from the path-diagram based SEM and the two-level SEM. For the LCA-SEM, we decide not to include the trend breaking analyses as it would add additional dimension of complexity to an already complex model.

The findings of both models suggest that the majority of the influences remain remarkably stable over time. However, the trend breaking influences have interesting policy implications. The Path-diagram based SEM suggests that the most systematic changes are the narrowing of the gender gap in travel across all travel outcomes and the increased influence of built form patterns on car ownership post-2007. Regarding gender gaps, it is the females who have reduced their travel distance for shopping trips; for commuting and other trips, however, it is the males’ reduced travel that narrowed the gap. The influence of built form latent variable on car ownership is increased by around 23% which suggests an increasing influence of residing in the denser, more urbanized areas upon reducing car ownership.

The analysis of within-built form influences of the two level SEM confirms the findings by the Path-diagram based SEM of the narrowing gender gap. A more interesting result is that from the between-level analysis. Similar to findings from the path-diagram based SEM, the between-built form variation in car ownership has increased by 36% post 2007. Additionally, the two level SEM provides more confidence that this increased influence on car ownership arises principally from factors other than social economic profiles of the residents.

By working with an on-going survey like the NTS, these methods have the potential of producing a regular and timely update on the shifts in the influences on travel. In future work, we believe that there is a significant potential in incorporating other variables, such as the new series of social and environmental attitudes variables in the NTS, and data external to the NTS dataset like fuel prices and fares. This would be discussed further in Section 6.4.

6.3. Policy implications

The findings summarized above have two direct implications regarding land use planning initiatives as well as transportation policies in the UK.

First, the findings underline a critical and increasing importance of land use planning policies in influencing travel outcomes. They suggest that built form characteristics can often be the strongest influences upon travel demand restraint after systematically controlling for interdependencies among the main variables recorded by the NTS. The models show that much of this influence is effected through restraining car ownership in dense urban areas. The built form influences on car ownership appear to have grown more than 23% in strength post-2007 compared with the preceding five years. Although it does not capture every vignette of the social and built form changes, the extended SEM models underline, with its comprehensive coverage and systematic decomposition, the most robust evidence to date of the built form restraint on travel demand in UK cities. Given that it takes time for many urban land use planning measures to come to fruition, it is important to continue monitoring of the effects in order to inform new policy and community actions.

Secondly, the models highlight, through findings on travel time as well as travel distance, that the mobility patterns of part time, single adult, female, low paid and non-car owning workers are significantly less efficient than those of traditional full time, service-sector male commuters. This applies to all travel, but especially commuting. Of course, the low skilled workers are less specialised and as a result they travel shorter distances with a slightly greater proportion of walking, but that alone does not account for the mobility gap. Furthermore, this gap appears to remain unchanged throughout the 2000s, in spite of major initiatives in the decade to improve accessibility on public transport. For workers of lower income and lower paid occupations, this disadvantage mainly stems from the lack of access to fast and efficient means of travel. However, for part time, single adult and female workers, the model results show that the reasons are more complex. Traditional transport service provisions targeted for full time males living in predominantly suburban areas and travelling during peak hours may have left a legacy system with an embedded bias.

In spite of remarkable investment in public transport and active modes, more need to be done to address the gaps in mobility, particularly for accessing job opportunities efficiently[38] specifically in large and medium urban areas. Since part time, single adult and female workers are the fastest growing segments in the labour market in many countries including the UK, the number of workers involved is much greater than covered by previous transport access programmes such as 'wheels to work' in the UK.

Our findings point to three priority areas for policy consideration. First, following from the success of deploying land use planning measures to restrain car ownership and car use in cities specifically dense urban areas, there should now be a greater focus in reshaping the transport system to improve mobility efficiency, particularly for the disadvantaged workers to access job opportunities - this task is made all the more urgent because of the worsening shortage in affordable housing in areas of fast job growth. Secondly, since the fastest growing segments of the labour force are part time, single adult and female workers whose travel needs are quite different from the majority of traditional commuters, there may be a greater call for flexible, demand responsive services, possibly with a renewed focus on economical, paratransit systems. Finally, there should be greater coordination among transport, urban land use planning and wider policies in helping those disadvantaged workers. Given the significant influences of built form and lifestyle choices, some effective improvements in their travel mobility may well result from outside the immediate confines of the transport system.

The model developed for this dissertation is proved to be a novel tools for monitoring changes in travel behaviour and the influences of planning interventions on that. The tools can be expanded to have more detailed spatial segmentations and be used for monitoring changes in travel behaviour over time as and when more data become available. Moreover, the ability of the proposed model in determining the most important influences after controlling for complicated interrelations can provide in-depth insights for developing causal transport demand models (e.g. activity based models).

[38] Whilst I accept that slower speeds and longer trip durations may be useful and productive (Jain and Lyons, 2008), e.g. for those who have access to flexi-working arrangements and comfort in travel, this is less relevant to the disadvantaged workers, especially those needing to clock in and out to get paid.

In summary this study identifies three main policy implications: first, in terms of methodology, policy studies on the influence of built form should address more specifically the issues of inter-dependencies among the factors that influence travel; the findings show that the extent of inter-dependencies is far higher than previously assumed, and this cast a new light upon the importance of integrating transport, land use and urban design interventions. Second, the substantive findings of the SEM model point to the need to further improving public transport accessibility in medium and large urban areas and for disadvantaged groups of women, self employed and part time workers. Thirdly, in terms of the integration of the data-analysis-policy making cycle, the findings points a new way to provide more timely monitoring of possible trend breaks.

The distinct capabilities of the SEM models developed in this dissertation has already attracted the attention of policy makers and has lead us to be commissioned a Transport Technology Research Innovation Grant project by the UK Department for Transport (DfT). The aim of the project is to expand the models developed in this dissertation by including more recent years of NTS data and more detailed spatial segmentations, in order to better quantify inter-dependencies among factors influencing car ownership and travel for policy decisions. The DfT commission would allow us to assemble more detailed NTS data at user specified geographical level (subject to data disclosure conditions), to an extent which was not and still is not available for this dissertation.

6.4. Recommendations for future studies

Access to a comprehensive dataset coupled with developing some novel analysis approach provide this dissertation with ample means to determine built form influences on major travel outcomes. Studying the travel behaviour of socioeconomic groups within their spatial context also underlines some policy implications for UK cities and highlights the drawbacks in conventional transport models.

Analysing all travel outcomes through SEM, which allows modelling endogeneity of car ownership and interactions among travel purposes, provides better understanding of accessibility and mobility of different socioeconomic groups and travellers residing in different areas. Further expansion to path-diagram based SEM postulates more rigorous picture of variations across area types and more precise measure of built form influences.

However, like all studies, there are significant room for improvements. First, it is informative to repeat these analyses for non-working population to better understand their travel needs and requirements; second, it is helpful to include a longer time series of the survey as they become available within NTS – this is feasible for England only, as post-2012 the NTS data is no longer collected in the other UK countries. The continuation of monitoring changes in travel behaviour will put us in a better position to judge the cause for recent drops in aggregate travel and to monitor the influences of urban planning and design interventions. Thirdly, although these analyses are steps forward towards understanding the causalities by controlling heterogeneities, the influences cannot be interpreted in a strict sense as causality. Further steps towards finer measurements of causality can be pursued by using panel datasets where the bias from omitted variables can be more fully controlled. Finally, the study will benefit from adding further variables including attitudinal parameters and macroeconomic factors (e.g. oil or house prices). For instance, the inference of between-level residual variance from two level SEM assumes that the residual variance can be attributed to the built form characteristics. This assumption can be better justified by including the aforementioned parameters.

In addition to the above points, this study has highlighted a number of topics on which further research is beneficial. The following Sections discuss four main areas.

6.4.1. The incorporation of travel preferences and attitudes

The SEMs of this dissertation control for self-selection and spatial sorting effects through modelling the interactions between socioeconomic and built form characteristics. This relies on a strong assumption for close correlation between socioeconomic class and travel preference which allow me to consider socioeconomic characteristics as a proxy to travel preferences and attitudes.

This assumption can potentially be relaxed in future studies by including travel preferences (e.g. preferences for driving a car or riding bike or living in rural or urban areas due to beliefs and attitudes etc) alongside socioeconomic characteristics and built form patterns in the general SEM framework discussed in this dissertation.

The main issue, however, is the availability of such extra information within travel surveys. The studies which have included travel preferences and attitudes (e.g. Handy et al., 2006, 2007b, Cao et al., 2007a) have normally designed their own questionnaires and collected their own sets of data; they are therefore restricted to small or specific groups of individuals and do not have the comprehensiveness of survey datasets. Consequently, specific methodologies, e.g. data fusion, should be considered with a view to transferring what is learned from the limited tailored datasets to more comprehensive travel surveys.

6.4.2. Developing a combined model out of the three approaches

Each of the three methods presented in this dissertation has provided its unique insights through relaxing some of the assumptions imposed on conventional multivariate regressions. The Path-diagram based SEM accounts for the heterogeneity of influences; LCA-SEM further relaxes the assumption for linear influence of built form characteristics, and finally the two-level SEM expands the path-diagram based SEM by including influences at built form clusters level. There may be a scope to combine the aforementioned techniques for developing a more general model.

One specific suggestion is to develop a two level SEM combined with between level latent continuous or categorical built form variable. The two level model developed for this study can be further enhanced by allowing random influences (i.e. random slops which vary across built form clusters) in addition to random intercepts. However, this requires a much larger between-level degrees of freedom to allow smooth convergence and robust outcomes. One possible solution is to model between-level variations of influences and intercepts across NTS PSUs. This raises the between-level degrees of freedom but hampers the capability to infer on built form characteristics based on residual variances as the PSU definitions are not directly associated with built form characteristics. To account for the built form characteristics of PSUs, one approach is to construct a between-level built form as a latent continuous or categorical variable in the same way we have done for path-diagram based SEM or LCA-SEM.

6.4.3. Extending the models to travel demand forecasting

One of the main steps forward is to apply what has been learnt from this dissertation in improving travel demand models. Developing demand models which allow for endogeneity of built form characteristics and car ownership as well as the interactions among travel purposes results in substantively improved predictability for testing policy scenarios. This involves further disaggregation of population and allowing interactions (e.g. between household size and built form characteristics) within utility functions.

One proposed future research within this general context is to integrate the proposed SEM frameworks in this study with discrete choice modelling techniques. For instance, the observed utility in car ownership model can be defined as a function of direct influences of socioeconomic variables and the indirect ones through built form parameters. Furthermore, the utility of car ownership, as a latent variable, can be treated as a one of the determinants of the utility of destination choice or its components such as travel time and cost. These however require in-depth studies to develop new algorithms for calibrating such complex models.

6.4.4. Bayesian non-parametric expansion of Structural Equation Models and latent categorization

Recent developments in machine learning techniques such as neural network and Bayesian statistics are gaining momentum. However, the discussions on incorporating those into transport demand modelling is fairly new. The current discussions on machine learning techniques within transportation are very much restricted to analysing big data, and little is known with respect to their potential use for long term travel demand forecasting.

One of the main issues is the interpretability of results specifically for travel demand analysis and understanding travellers' behaviour. Expanding the models provided for this study through machine learning techniques can be a very interesting attempt to improve our understanding of travel behaviour and track its changes over time. For instance one might employ Gaussian Process to relax any particular assumption on the form of functions determining travel time and distance. Alternatively, non-parametric Bayesian approaches (e.g. the Dirichlet process) can be used to expand LCA-SEM to allow automatic estimation of the number and form of the built form clusters and to create a mechanism to track changes in that over time as new data is made available.

Finally, the comparison of outcomes from non-parametric approach with those reported in this dissertation can provide tantalising insights into the validity of structural assumptions.

References

ADITJANDRA, P. T., CAO, X. J. & 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.

ALBRIGHT, J. J. & PARK, H. M. 2009. Confirmatory factor analysis using Amos, LISREL, Mplus, and SAS/STAT CALIS. *The Trustees of Indiana University*, 1, 1-85.

ANGRIST, J.D., & Lang K., 2004. Does School Integration Generate Peer Effects? Evidence from Boston.s Metco Program,.*The American Economic Review*, 94, 1613.1634.

ANGRIST, J.D. & PISCHKE, J.S., 2008. Mostly harmless econometrics: An empiricist's companion. *Princeton university press.*

AXHAUSEN, K. W. 2003. Capturing long distance travel, Baldock, Research Studies.

BANISTER, D. 1997. Reducing the need to travel. *Environment and Planning B-Planning & Design,* 24**,** 437-449.

BANISTER, D. 2000. Sustainable urban development and transport - a Eurovision for 2020. *Transport Reviews,* 20**,** 113-130.

BANISTER, D., WATSON, S. & WOOD, C. 1997. Sustainable cities: Transport, energy, and urban form. *Environment and Planning B-Planning & Design,* 24**,** 125-143.

BEN-AKIVA, M. & BOLDUC, D. 1996. *Multinomial probit with a logit kernel and a general parametric specification of the covariance structure*, D′epartement d″economique, Universit′e laval with Department of Civil and Environmental Engineering, Massachusetts Institute of Technology.

BERKE, P.R. & CONROY, M.M., 2000. Are we planning for sustainable development? An evaluation of 30 comprehensive plans. Journal of the American planning association, 66(1), pp.21-33.

BHAT, C.R. 2002. recent methodological advances relevant to activity and travel behaviour analysis. in: in perpetual motion: travel behavior research opportunities and application challenges.

BHAT, C. R. 1997. Covariance heterogeneity in nested logit models: econometric structure and application to intercity travel. *Transportation Research Part B: Methodological,* 31**,** 11-21.

BHAT, C. R. & GOSSEN, R. 2004. A mixed multinomial logit model analysis of weekend recreational episode type choice. *Transportation Research Part B: Methodological*.

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

BOARNET, M. G. 2004. The Built Environment and Physical Activity: Empirical Methods and Data Resource. *Transport Research Board.*

BOHTE, W. & MAAT, K. 2009. Deriving and validating trip purposes and travel modes for multi-day GPS-based travel surveys: a large-scale application in the Netherlands. *Transportation Research Part C: Emerging Technologies,* 17**,** 285-297.

BOHTE, W., MAAT, K. & VAN WEE, B. 2009. Measuring Attitudes in Research on Residential Self-Selection and Travel Behaviour: A Review of Theories and Empirical Research. *Transport Reviews,* 29**,** 325-357.

BOURNE, L. S. 1992. Self-fulfilling prophecies?: Decentralization, inner city decline, and the quality of urban life. *Journal of the American Planning Association,* 58**,** 509-513.

BREHENY, M. J. 1992. *The contradictions of the compact city: a review*. *Sustainable Development and Urban Form***,** 138-159.

CAMERON, A. C. & MILLER, D. L. 2015. A practitioner's guide to cluster-robust inference. *Journal of Human Resources,* 50**,** 317-372.

CAO, X., MOKHTARIAN, P. L. & HANDY, S. L. 2007a. Cross-sectional and quasi-panel explorations of the connection between the built environment and auto ownership. *Environment and Planning A*, 39**,** 830-847.

CAO, X., MOKHTARIAN, P. L. & HANDY, S. L. 2007b. Do changes in neighborhood characteristics lead to changes in travel behavior ? A structural equations modeling approach. *Transportation,* 34**,** 535-556.

CAO, X., MOKHTARIAN, P. L. & HANDY, S. L. 2009. The relationship between the built environment and nonwork travel: A case study of Northern California. *Transportation Research Part a-Policy and Practice,* 43**,** 548-559.

CAO, X.& CHATMAN, D., 2016. How will smart growth land-use policies affect travel? A theoretical discussion on the importance of residential sorting.*Environment and Planning B: Planning and Design*, *43*(1), pp.58-73.

CERVERO, R. 1996. Jobs-housing balance revisited - Trends and impacts in the San Francisco Bay Area. *Journal of the American Planning Association,* 62**,** 492-511.

CERVERO, R. & DUNCAN, M. 2006. Which reduces vehicle travel more: Jobs-housing balance or retail-housing mixing? *Journal of the American Planning Association,* 72**,** 475-490.

CERVERO, R. & KOCKELMAN, K. 1997. Travel demand and the 3Ds: Density, diversity, and design. *Transportation Research Part D: Transport and Environment,* 2**,** 199-219.

CERVERO, R. & MURAKAMI, J. 2010. Effects of built environments on vehicle miles traveled: evidence from 370 US urbanized areas. *Environment and Planning A,* 42**,** 400-418.

CERVERO, R. & WU, K. 1997. Polycentrism, commuting, and residential location in the San Francisco Bay area. *Environment and Planning A,* 29**,** 865-86.

CHO, S.-J., PREACHER, K. J. & BOTTGE, B. A. 2015. Detecting Intervention Effects in a Cluster-Randomized Design Using Multilevel Structural Equation Modeling for Binary Responses. *Applied Psychological Measurement*.

CRANE, R. 2000. The Influence of Urban Form on Travel: An Interpretive Review. *Planning Literature,* 15**,** 3-23.

DARGAY, J. 2007. The effect of prices and income on car travel in the UK. *Transportation Research Part A: Policy and Practice,* 41**,** 949-960.

DARGAY, J. & HANLY, M. 2004. Land use and mobility. *World Conference on Transport Research.*

DARGAY, J. M. & HANLY, M. 2003. The Impact of Land Use Patterns on Travel Behaviour. *European Transport Conference.* Strasbourg, France.

DUNN, E. C., MASYN, K. E., JONES, S. M., SUBRAMANIAN, S. V. & KOENEN, K. C. 2015. Measuring psychosocial environments using individual responses: an application of multilevel factor analysis to examining students in schools. *Prev Sci,* 16**,** 718-33.

EWING, R. & CERVERO, R. 2001. Travel and the Built Environment: A Synthesis. *Transportation Research Record: Journal of the Transportation Research Board,* 1780**,** 87-114.

FARTHING, S., WINTER, J. & COOMBES, T. 1996. Travel behaviour and local accessibility to services and facilities. *The Compact City: A sustainable urban form***,** 181-189.

GAO, S., MOKHTARIAN, P. L. & JOHNSTON, R. A. 2008. Exploring the connections among job accessibility, employment, income, and auto ownership using structural equation modeling. *Annals of Regional Science,* 42**,** 341-356.

GIULIANO, G. & DARGAY, J. 2006. Car ownership, travel and land use: a comparison of the US and Great Britain. *Transportation Research Part A: Policy and Practice,* 40**,** 106–124.

GIULIANO, G. & SMALL, K. A. 1993. Is the Journey to Work Explained by Urban Structure? *Urban Studies*.

GOLOB, T. F. 2003. Structural Equation Modeling for Travel Behavior Research. *Transportation Research, B - Methodological,* 37**,** 1-25.

GORDON, P. & RICHARDSON, H. W. 1989. Gasoline Consumption and Cities: A Reply. *Journal of the American Planning Association*.

GORDON, P. & RICHARDSON, H. W. 1997. Are Compact Cities a Desirable Planning Goal? *Journal of the American Planning Association*.

HANDEL, M. J. 2012. Trends in Job Skill Demands in OECD Countries. *OECD Social, Employment and Migration Working Papers,* 143.

HANDY, S., CAO, X. Y. & MOKHTARIAN, P. 2005. Correlation or causality between the built environment and travel behavior? Evidence from Northern California. *Transportation Research Part D-Transport and Environment,* 10**,** 427-444.

HANDY, S., CAO, X. Y. & MOKHTARIAN, P. L. 2006. Self-selection in the relationship between the built environment and walking - Empirical evidence from northern California. *Journal of the American Planning Association,* 72**,** 55-74.

HAYLLAR, O., MCDONNELL, P., MOTTAU, C. & SALATHIEL, D. 2005. National Travel Survey 2003 & 2004 technical report. UK Department for Transport.

HECKMAN, J. J., & Robb, R. 1985: Alternative Methods for Evaluating the Impact of Interventions. *In Longitudinal Analaysi of Labor Market Data.*, ed. by J. Heckman and B. Singer. CUP

HESS, S., BOLDUC, D. & POLAK, J. W. 2010. Random covariance heterogeneity in discrete choice models. *Transportation,* 37**,** 391-411.

HICKMAN, R. & BANISTER, D. 2004. Reducing Travel By Design: Urban Form and the Commute to Work. *European Transport Conference.* Strasbourg, France: Association for European Transport.

HINE, J. & GRIECO, M. 2003. Scatters and clusters in time and space: implications for delivering integrated and inclusive transport. *Transport Policy,* 10**,** 299-306.

HUANG, B. 2007. The Use of Pseudo Panel Data for Forecasting Car Ownership. Munich Personal RePEc Archive.

ICHIMURA, M., 2003, January. Urbanization, urban environment and land use: challenges and opportunities. *In Asia-Pacific Forum for Environment and Development*, Expert Meeting 23 January 2003, Guilin, People's Republic of China.

JAHANSHAHI, B. 2017. Separating gender composition effects from peer effects in education. *Education Economics* 25.1: 112-126.

JAHANSHAHI, K. & JIN, Y. 2015. The built environment typologies in the UK and their influences on travel behaviour: new evidence through latent categorisation in structural equation modelling. *Transportation Planning and Technology,* 39.

JAHANSHAHI, K. & JIN, Y. 2016. Trendbreaking Influences of Built Form on Travel in UK Cities: Evidence from New Quantifications of Within- and Between-Built-Form Variations. *Transportation Research Record: Journal of the Transportation Research Board,* 2564.

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

JAHANSHAHI, K., JIN, Y. & WILLIAMS, I. N. 2013. Analyzing Sources of Variability in Travel Time Use in a Combined Framework Using Extended Structural Equation Models and UK National Travel Survey Data. *Transport Research Board Annual Meeting.* Washington DC.

JAHANSHAHI, K., WILLIAMS, I. & HAO, X. Understanding Travel Behaviour and Factors Affecting Trip Rates. European Transport Conference, 2009.

JAIN, J. & LYONS, G. 2008. The gift of travel time. *Journal of Transport Geography,* 16**,** 81-89.

JOLLIFFE, D., LANJOUW, P., CHEN, S., KRAAY, A., MEYER, C., NEGRE, M., PRYDZ, E., VAKIS, R. & WETHLI, K. 2015. A Measured Approach to Ending Poverty and Boosting Shared Prosperity : Concepts, Data, and the Twin Goals. *Development Research Group of the World Bank.*

JONES, P. & LE VINE, S. 2012. On the MoveMaking sense of car and train travel trends in Britain. London.

JOWETT, A., TAYLOR, C., HADIE, M. & KHAN, Z. 2014. An International Perspective on the UK, Labour Market Performanc. *ONS Office of the Chief Economic Advisor*.

Lawrence, D.L. and Low, S.M., 1990. *The built environment and spatial form.Annual review of anthropology*, 19, pp.453-505.

LEVINSON, D. M. & KUMAR, A. 1994. The Rational Locator: Why Travel Times Have Remained Stable. *Journal of the American Planning Association*.

LIAO, F. H., FARBER, S. & EWING, R. 2015. Compact development and preference heterogeneity in residential location choice behaviour: A latent class analysis. *Urban Studies,* 52**,** 314-337.

LUCAS, K. 2006. Providing transport for social inclusion within a framework for environmental justice in the UK. *Transportation Research Part A: Policy and Practice,* 40**,** 801-809.

LUCAS, K. 2012. Transport and social exclusion: Where are we now? *Transport policy,* 20**,** 105-113.

MAAT, K., TIMMERMANS, H. J. P. & MOLIN, E. 2004. A model of Spatial Structure, Activity Participation and Travel Behaviour. *10th World Conference on Transport Research.* Istanbul, Turkey.

MACCALLUM, R. C., BROWNE, M. W. & SUGAWARA, H. M. 1996. Power analysis and determination of sample size for covariance structure modeling. *Psychological methods,* 1**,** 130.

MALDONADO-HINAREJOS, R., SIVAKUMAR, A. & POLAK, J. W. 2014. Exploring the role of individual attitudes and perceptions in predicting the demand for cycling: a hybrid choice modelling approach. *Transportation,* 41**,** 1287-1304.

MARSH, H. W., LÜDTKE, O., ROBITZSCH, A., TRAUTWEIN, U., ASPAROUHOV, T., MUTHÉN, B. & NAGENGAST, B. 2009. Doubly-Latent Models of School Contextual Effects: Integrating Multilevel and Structural Equation Approaches to Control Measurement and Sampling Error. *Multivariate Behavioral Research,* 44**,** 764-802.

MARSH, H. W., MORIN, A. J., NAGENGAST, B. & SCALAS, L. F. 2015. Doubly latent multilevel analyses of classroom climate : an illustration. *Journal of Experimental Education,* 82**,** 143.

MARTIN, G. 2007. Global motorization, social ecology and China. *Area,* 39**,** 66-73.

MCFADDEN, D. & TRAIN, K. 2000. Mixed MNL models for discrete response. *Journal of applied Econometrics,* 15**,** 447-470.

MELBOURNE, L. 2012. Statistical release: National travel survey. *In:* TRANSPORT, D. F. (ed.).

MITCHELL, R. B. & RAPKIN, C. 1954. *Urban traffic: A function of land use,* New York: Columbia University Press.

MOKHTARIAN, P. L. & CAO, X. 2008. Examining the impacts of residential self-selection on travel behavior: A focus on methodologies. *Transportation Research Part B-Methodological,* 42**,** 204-228.

MOKHTARIAN, P. L. & VAN HERICK, D. 2016. Viewpoint: Quantifying residential self-selection effects: A review of methods and findings from applications of propensity score and sample selection approaches. *Journal of Transport and Land Use,* 9.

MORRIS, S., HUMPHREY, A., PICKERING, K., TIPPING, S., TEMPLETON, I. & HURN, J. 2014. National Travel Survey 2013 technical report. UK Department for Transport.

MUTHEN, L. K. & MUTHEN, B. O. 2007. Mplus user's guide. 5th. *Los Angeles, CA: Muthén & Muthén***,** 197-200.

MUTHÉN, B. 1984. A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika,* 49**,** 115-132.

MUTHÉN, B. & ASPAROUHOV, T. 2007. Growth mixture analysis: Models with non-Gaussian random effects. *In:* FITZMAURICE, G., DAVIDIAN, M., VERBEKE, G. AND MOLENBERGHS, G. CHAPMAN & HALL/CRC PRESS (ed.) *Advances in Longitudinal Data Analysis.*

NEWMAN, P. & KENWORTHY, J. R., 1989. *Cities and Automobile Dependence: An International Sourcebook.* Gower, Aldershot, UK

NEWMAN, P. & KENWORTHY, J. R. 1999. *Sustainability and Cities: Overcoming Automobile Dependence,* Island Press.

NÆSS, P. 2012. Urban form and travel behavior: Experience from a Nordic context. *Journal of Transport and Land Use,* 5**,** 21-45.

PALMER, N. 2014. Trends in self employment. *ONS Economic Forum*.

PAWLAK, J., ZOLFAGHARI, A. & POLAK, J. 2015. Imputing Socioeconomic Attributes for Movement Data by Analysing Patterns of Visited Places and Google Places Database: Bridging between Big Data and Behavioural Analysis.

PRATO, C. G. 2015. Latent Lifestyle and Mode Choice Decisions when Travelling Short Distances. *IATBR.* Sindsor.

PRESTON, J. & RAJÉ, F. 2007. Accessibility, mobility and transport-related social exclusion. *Journal of Transport Geography,* 15**,** 151-160.

ROBERT, C. & MURAKAMI, J. 2010. Effects of built environments on vehicle miles traveled: evidence from 370 US urbanized areas. *Environment and Planning A,* 42**,** 400-418.

SCHWANEN, T., LUCAS, K., AKYELKEN, N., SOLSONA, D. C., CARRASCO, J.-A. & NEUTENS, T. 2015. Rethinking the links between social exclusion and transport disadvantage through the lens of social capital. *Transportation Research Part A: Policy and Practice,* 74**,** 123-135.

SILVA, J. D. A. E., GOLOB, T. F. & GOULIAS, K. G. 2007. Effects of Land Use Characteristics on Residence and Employment Location and Travel Behavior of Urban Adult Workers. *Transportation Research Record: Journal of the Transportation Research Board,* 1977**,** 121-131.

SILVA, J. D. A. E., MORENCY, C. & GOULIASC, K. G. 2012. Using structural equations modeling to unravel the influence of land use patterns on travel behavior of workers in Montreal. *Transportation Research Part A: Policy and Practice,* 46**,** 1252–1264.

SPISSU, E., PINJARI, A. R., BHAT, C. R., PENDYALA, R. M. & AXHAUSEN, K. W. 2009. An analysis of weekly out-of-home discretionary activity participation and time-use behavior. *Transportation,* 36**,** 483-510.

STEAD, D. 2001. Relationships between land use, socioeconomic factors, and travel patterns in Britain. *Environment and Planning B-Planning & Design,* 28**,** 499-528.

STEAD, D. & MARSHALL, S. 2001. The relationships between urban form and travel patterns. An international review and evaluation. *European Journal of Transport and Infrastructure Research,* 1**,** 113-141.

SUN, Y., WAYGOOD, E., FUKUI, K. & KITAMURA, R. 2009. Built Environment or Household Life-Cycle Stages—Which Explains Sustainable Travel More? *Transportation Research Record: Journal of the Transportation Research Board,* 2135**,** 123-129.

SUN, Y., WAYGOOD, E. & HUANG, Z. 2012. Do Automobility Cohorts Exist in Urban Travel? *Transportation Research Record: Journal of the Transportation Research Board,* 2323**,** 18-24.

SUSILO, Y. O. 2015. The Influence of Parent's Perceptions and Residential Self-Selection to the Children's Travel Modes at Single Parent Households. *Sustainable Urban Transport (Transport and Sustainability, Volume 7) Emerald Group Publishing Limited,* 7**,** 43-64.

SUSILO, Y. O. & KITAMURA, R. 2008. Structural changes in commuters' daily travel: The case of auto and transit commuters in the Osaka metropolitan area of Japan, 1980-2000. *Transportation Research Part a-Policy and Practice,* 42**,** 95-115.

TIMMERMANS, H., VAN, D. W. P., ALVES, M., POLAK, J., ELLIS, S., HARVEY, A. S., KUROSE, S. & ZANDEE, R. 2003. Spatial context and the complexity of daily travel patterns: An international comparison. *Journal of Transport Geography,* 11**,** 37-46.

TRAIN, K. E. 2009. *Discrete choice methods with simulation*, Cambridge university press.

VAN ACKER, V., MOKHTARIAN, P. L. & WITLOX, F. 2014. Car availability explained by the structural relationships between lifestyles, residential location, and underlying residential and travel attitudes. *Transport Policy,* 35**,** 88–99.

VAN ACKER, V., WITLOX, F. & VAN WEE, B. 2007. The effects of the land use system on travel behavior: A structural equation modeling approach. *Transportation Planning and Technology,* 30**,** 331-353.

VAN WEE, B., HOLWERDA, H. & VAN BAREN, R. 2002. Preferences for modes, residential location and travel behaviour: the relevance for land-use impacts on mobility. *European Journal of Transport and Infrastructure Research,* 2**,** 305-316.

WOOLDRIDGE, J. M. 2009. *Introductory Econometrics: A Modern Approach,* Fourth international ed. Australia: South-Western. pp.838.

WALKER, J. & LI, J. 2007. Latent lifestyle preferences and household location decisions. *Journal of Geographical Systems,* 9**,** 77-101.

WANG, J. & WANG, X. 2012. *Structural Equation Modeling: Applications Using Mplus*, WILEY.

WANG, T. & CHEN, C. 2012. Attitudes, mode switching behavior, and the built environment: A longitudinal study in the Puget Sound Region. *Transportation Research Part A: Policy and Practice,* 46**,** 1594–1607.

WEIS, C. & AXHAUSEN, K. W. 2009. Induced travel demand: Evidence from a pseudo panel data based structural equations model. *Research in Transportation Economics,* 25**,** 8-18.

WSP 2009. Research into Changing Trip Rates over Time and Implications for the National Trip End Model: Final Report.

ZEGRAS, C., LEE, J. S. & BEN-JOS EPH, E. 2012. By Community or Design? Age-restricted Neighbourhoods, Physical Design and Baby Boomers' Local Travel Behaviour in Suburban Boston, US. *Urban Studies,* 49**,** 2169–2198.

Appendix A Review of other methods used in transport studies to account for intercorrelation issues

Apart from SEM, recent years have seen rapid progress in other modelling techniques. Bhat (2002) has provided a comprehensive review of advanced choice models which relax one or more IID error structure assumptions in MNL, and/or unobserved response homogeneity assumption through allowing random coefficients for attributes (e.g. Ben-Akiva and Bolduc, 1996, McFadden and Train, 2000, Train, 2009, Bhat and Gossen, 2004, Bhat and Guo, 2004, Walker and Li, 2007, Spissu et al., 2009, Hess et al., 2010, Prato, 2015, Liao et al., 2015). The latter can be specified by assuming a continuous distribution (random coefficient models) or a non-parametric discrete distribution (latent segmentation approach) for random coefficients.

As stated in McFadden and Train (2000), the generalized form of MNL (Mixed MNL- MMNL) can be used to model almost any form of discrete choice models derived from Random Utility Maximization (RUM) theory. In other words, the random parameter in MMNL models can be specified to capture the heterogeneity among individuals (i.e. taste variation) or alternatives (i.e. to relax IIA assumption) depending on the purpose of analysis. Notably, Bhat and Gossen (2004) develop a mixed logit model to analyse the type of weekend recreational activity episodes that individuals pursue. Their model allows both unobserved heterogeneity in preferences across individuals, and correlation across unobserved utility components of the alternatives. The former is accommodated through defining a random intercept to model heterogeneity across individuals, and the latter through decomposing the error terms into two components: one standard iid and one to induce correlations across unobserved utility components of the alternatives.

Bhat and Guo (2004) develop a mixed spatial correlated logit model to analyse spatial correlation in spatial location choice while unobserved taste variation is also captured. Their model is the combination of GEV with equal dissimilarity parameter across all paired nests to capture correlation in the unobserved utility components of spatial units and the normally distributed random parameters to capture potential taste variations among individuals. Their analysis highlights the potential for getting biased elasticity effects of exogenous variables should one ignores spatial correlation and unobserved response heterogeneity.

Hess et al (2010) specify a formulation for more flexible mixed model which can allow random covariance heterogeneity across respondents. They formulate that by expanding COVNL model of Bhat (1997) to allow the normally distributed structural parameter in NL model to vary across individuals. They also show the possibility for expanding ECL- expansion of MNL which allows heterogeneity across alternatives- to accommodate heterogeneity across respondents as well. The stated preference application of their model shows that considerable gains can be achieved through allowing heterogeneity across individuals in addition to allowing that across alternatives.

Spissu et al (2009) study six categories of discretionary activity participation to understand influences on the inter-personal and intra-personal variability in weekly activity engagement. They develop a mixed multiple discrete-continuous extreme value (MMDCEV) model to dismantle inter-individual and intra-individual variations. In their formulation, the intercept has three separate components: the first one is to capture the average effects across individuals on the base-line utility for a particular choice alternative (i.e. a fixed component across the individuals); the second one is to capture the heterogeneity across individuals due to unobserved individual attributes that are not correlated across alternatives; the third one is to adopt a mechanism to generate individual level correlation across unobserved utility components of the alternatives.

Maldonado-Hinarejos et al (2014) explore the role of attitudes and perception as latent variables in estimating the demand for cycling. Their hybrid choice model is estimated sequentially. First, the latent variables are specified through factor analysis and then they are used along with socio-demographic variables to estimate MNL choice model parameters. Although theoretically sequential estimation might be prone to bias, they argue that their estimation based on simulated data shows not much difference between sequential and simultaneous estimation of latent variables and choice model parameters.

One recent example of the study allowing response heterogeneity is that of Liao et al (2015) who examine the residential preferences for compact development in the State of Utah whilst controlling for heterogeneity in residential location choice arising from household socioeconomic backgrounds and attitudes. Using Latent Class Analysis (LCA) within a discrete choice framework, they classify individuals into latent categories based on their sociodemographic characteristics and attitudes toward the natural and social environments, travel mode, and environmental protection. Their results suggest strong associations between location choice and sociodemographic status and attitudes. They finally recommend the use of structural equation models as a more suitable technique to further gauge the endogenous linkages between socio-demographics, attitudes and residential preferences for future studies.

In summary, apart from the SEM which has been widely used in analysing highly correlated influences on travel, recent studies have attempted to use random intercept models and latent class analysis (mainly through GEV, mixed logit or MDCEV) to allow heterogeneity across individuals or choice alternatives. However, to the best of our knowledge, these techniques have not been applied within the SEM framework.

Appendix B Supplementary data tables, graphs and charts

B1. The characteristics of Built form continuous latent variable

Figure 12 below shows the Built form latent variable histogram. The descriptive statistics such as Mean and variance is also presented. It can be observed that in spite of some skewness, it still match well with normal distribution assumptions. This is also reiterated in QQ plot shown in Figure 13.

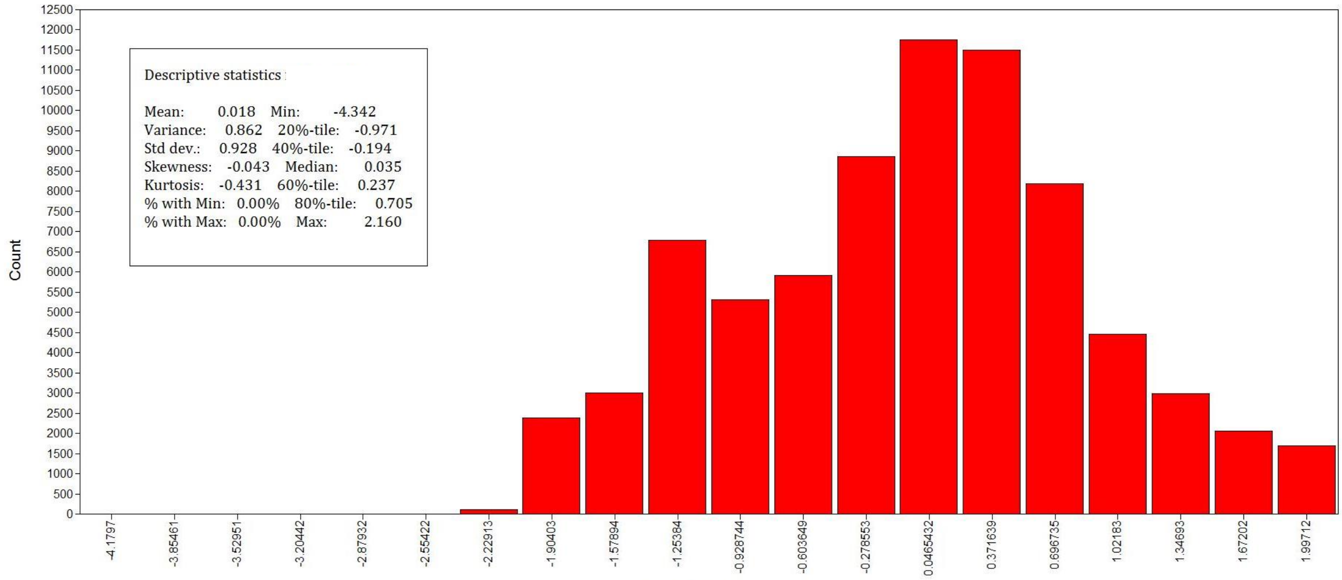

Figure 12 Built form latent variable distribution

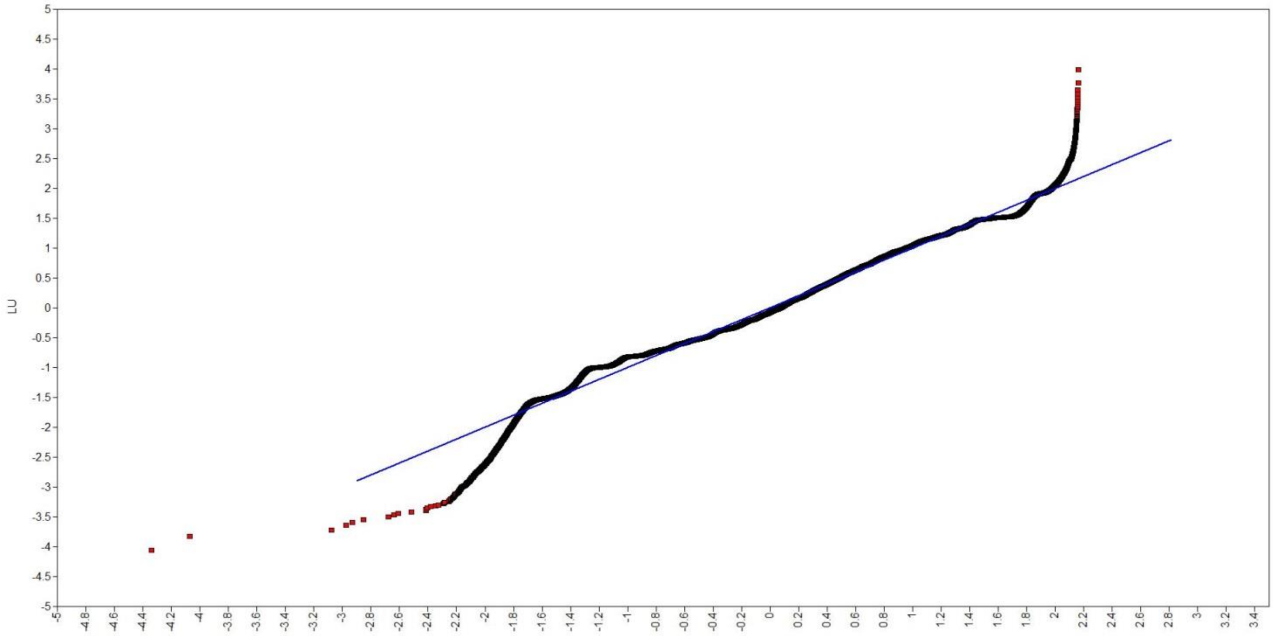

Figure 13 Built form latent variable QQ plot

B2. Graphical representation of Model estimation for travel time model (units are in 100 minutes per traveller per week)

Highlighting different components of the full model, Figure 14 to Figure 18 present the coefficients estimated by two alternative estimation methods for the travel time model: weighted least squares (WLS) and maximum likelihood (ML). In all graphs, WLS outputs are the first line and ML ones are the second line of numbers. The numbers in parentheses are the P-Values and those followed by "/" are the standardized coefficients; the unitless version of an estimated coefficient which are used to facilitate comparisons of the relative importance of influences. In Mplus, the standardized coefficients are only outputted for the WLS, being unavailable for the ML estimator. Travel time in the following figures is in the unit of 100 minutes. For instance, full time workers spend 41 minutes more than part timers on commuting. The estimated coefficient is the same based on WLS and ML estimators.

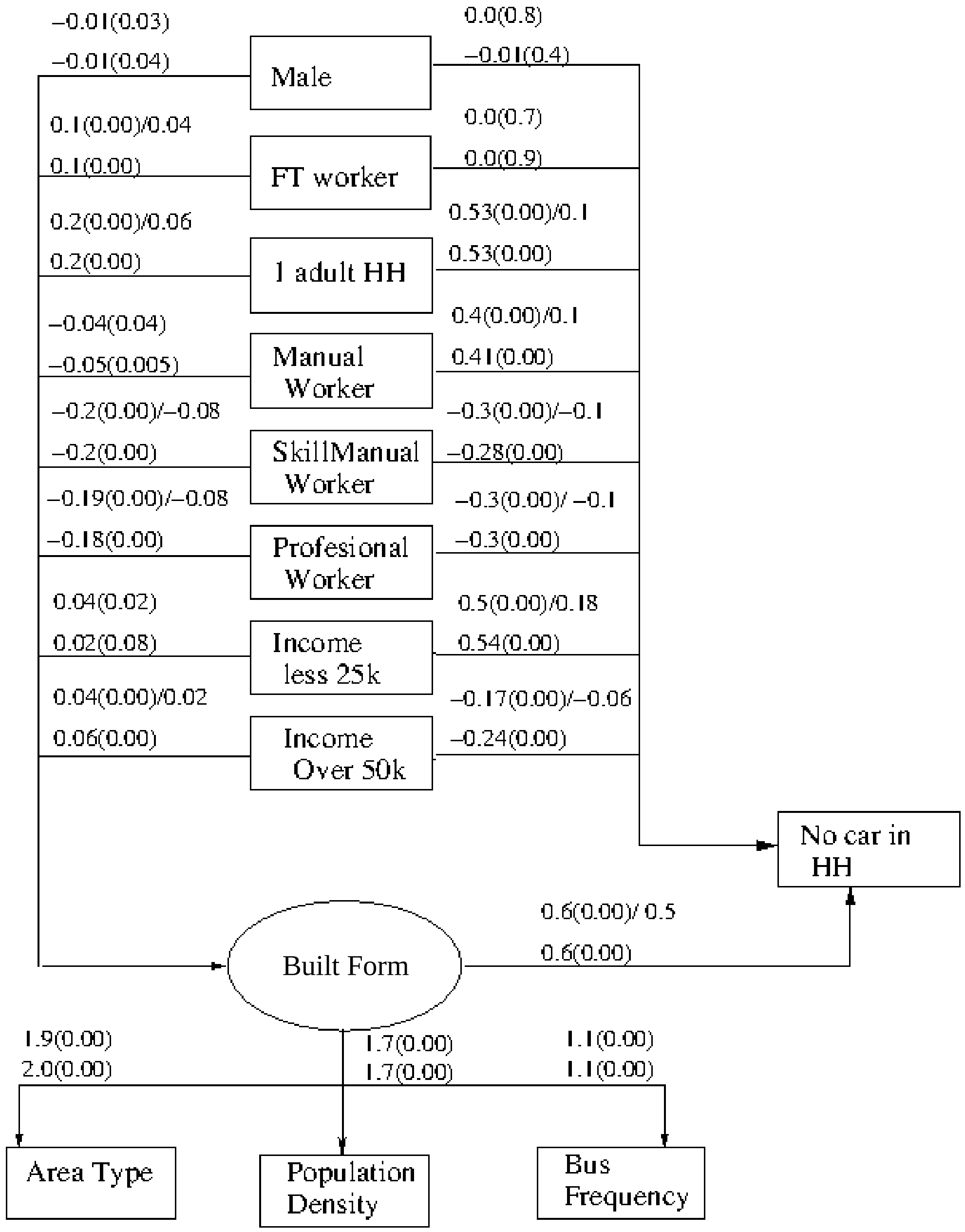

Figure 14 The effects of socioeconomic and built form variables on car ownership-WLS with standardized coefficient vs ML

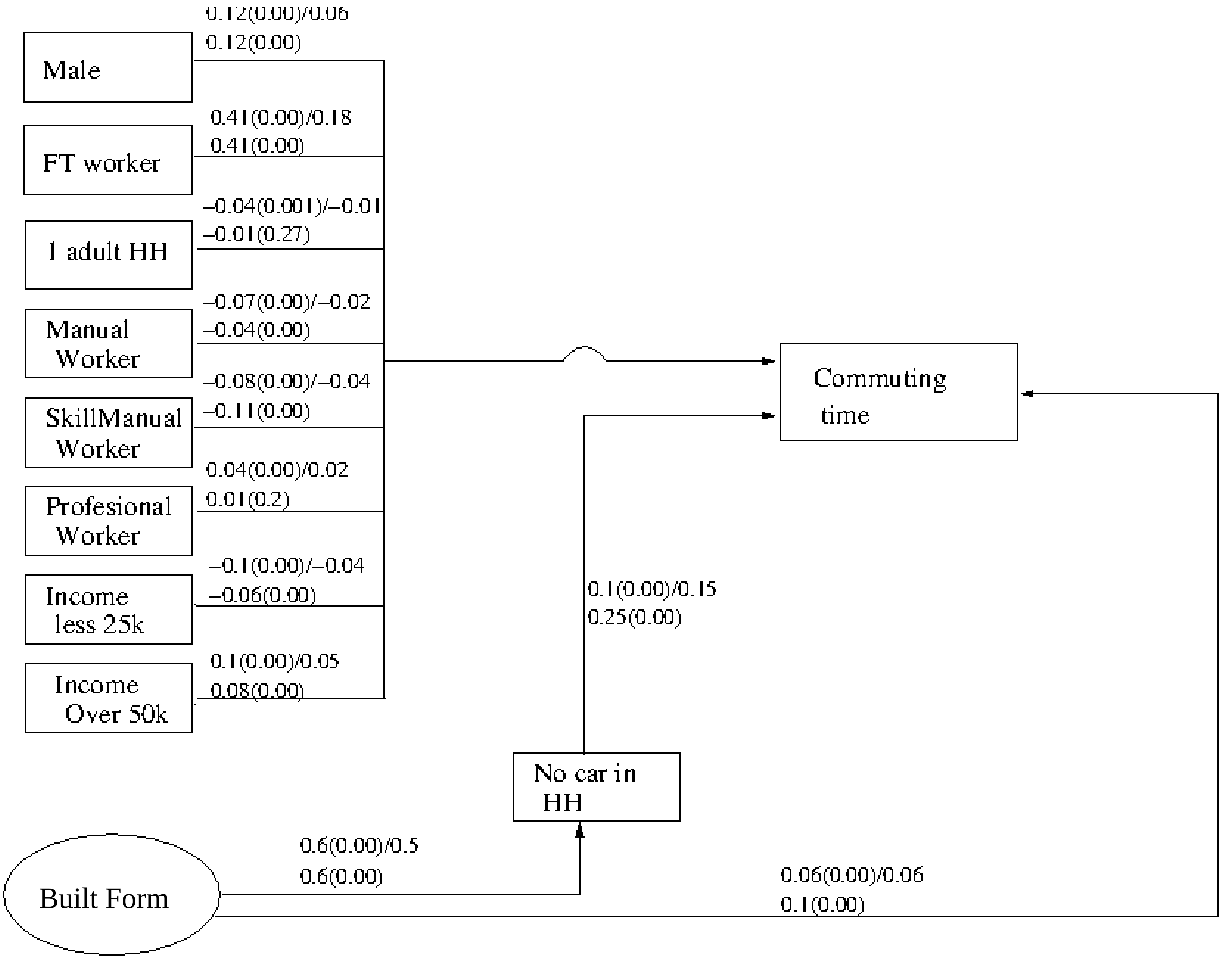

Figure 15 The effect of socioeconomic, land us and car ownership variables on commuting time-WLS with standardized coefficient vs ML

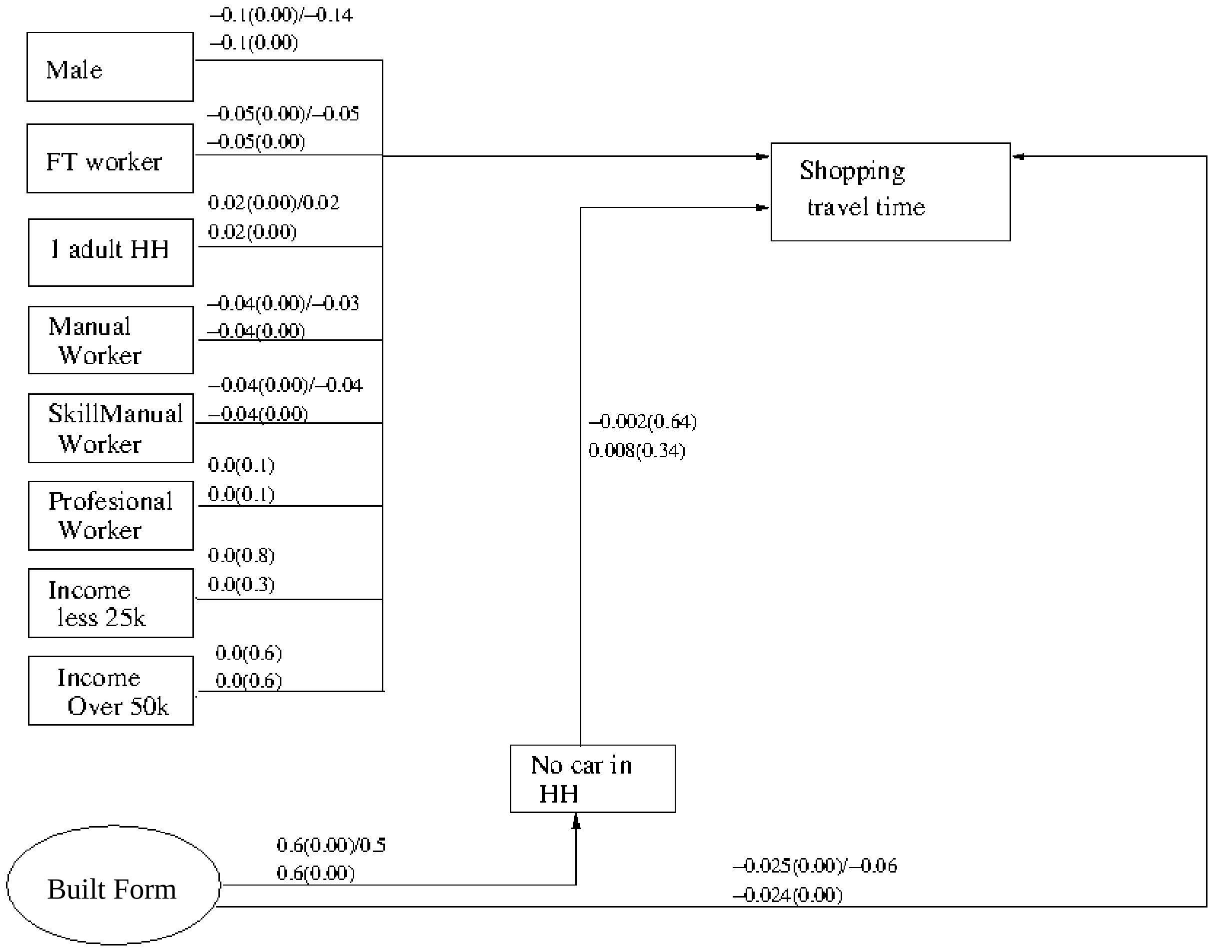

Figure 16 The effect of socioeconomic, land us and car ownership variables on shopping travel time-WLS with standardized coefficient vs ML

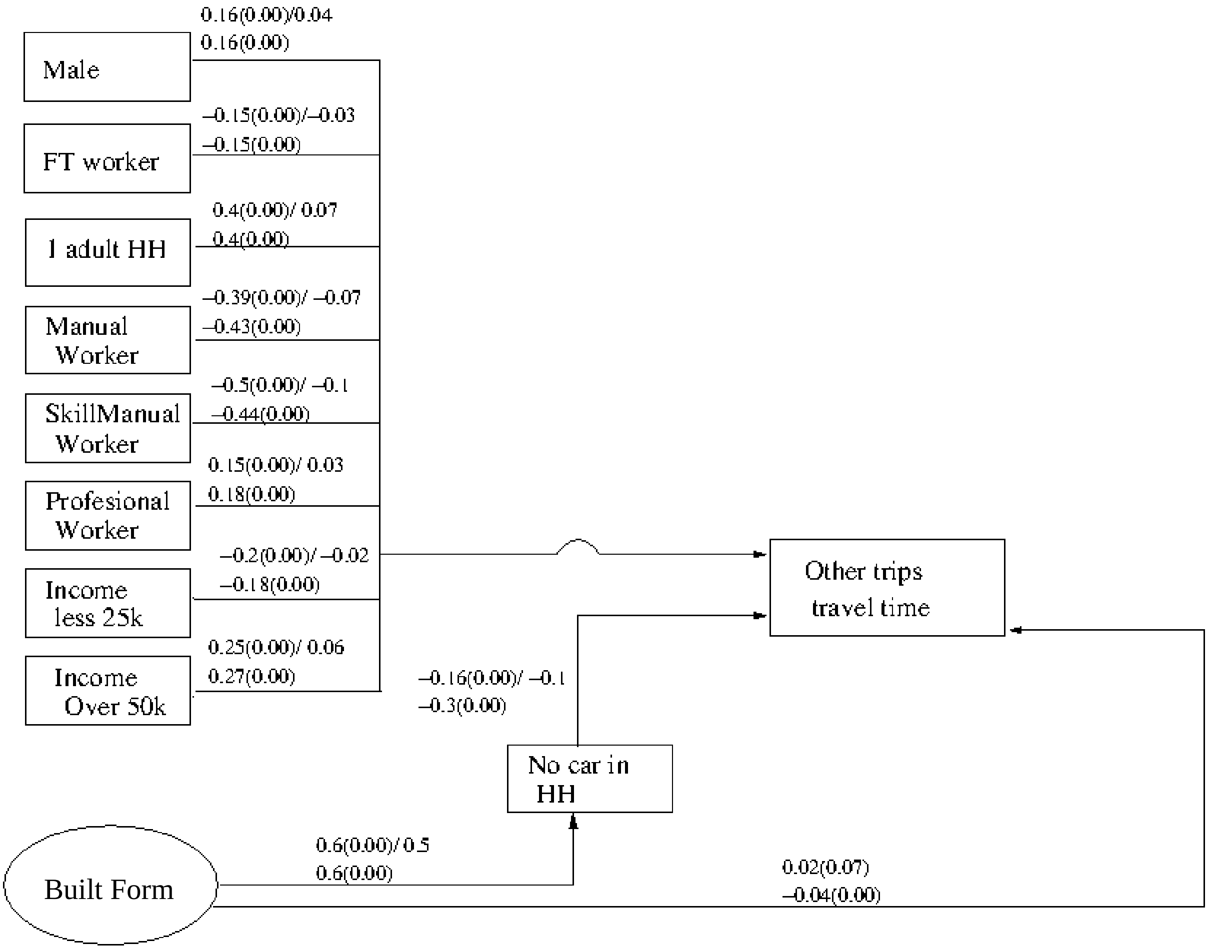

Figure 17 The effect of socioeconomic, land us and car ownership variables on other trips' travel time-WLS with standardized coefficient vs ML

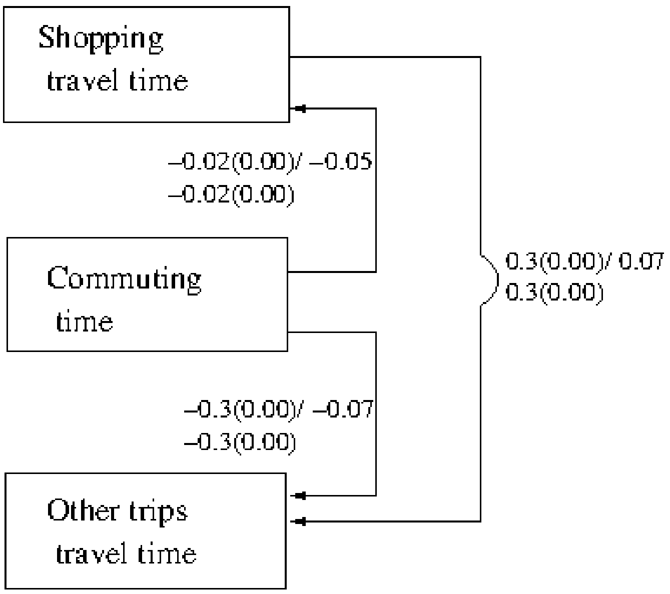

Figure 18 The interactions between travel purposes-WLS with standardized coefficient vs ML

B3. WLS and ML modelling of indirect influences

Table 34 to Table 36 list the significant indirect influences on car ownership, work and shopping travel time, travel distance and trip frequency in the WLS and ML models along with their associated direct influences. Those indirect effects with significant opposite directional influences when compared to the direct effects, and those with relatively large coefficient are each shown in Bold.

Table 34 Estimation of indirect effects on car ownership, work and shopping travel time (travel time unit is in minutes)

FT->NoCar		-0.002	not significant
	FT->LU->NoCar	0.058	0.059
1adult->Nocar		0.532	0.534
	1adult->LU->NoCar	0.121	0.116
SkillManual->NoCar		-0.284	-0.335
	SkillManual->LU->NoCar	-0.123	-0.117
Prof->NoCar		-0.296	-0.285
	Prof->LU->NoCar	-0.113	-0.111
IncomeOver50K->NoCar		-0.241	-0.175
	IncomeOver50K->LU->NoCar	0.038	0.028
FT->HBW		41.4	41.3
	FT->LU->HBW	1	0.6
	FT->LU->NoCar->HBW	1.5	0.7
1adult->HBW		not significant	-4.2
	1adult->LU->HBW	2.1	1.1
	1adult->LU->NoCar->HBW	3.1	1.3
	1adult->NoCar->HBW	13.5	6
Manual->HBW		-4.5	-6.7
	Manual->LU->HBW	-0.6	-0.2
	Manual->LU->NoCar->HBW	-0.8	-0.3
	Manual->NoCar->HBW	10.4	4.4
SkillManual->HBW		-11	-8.3
	SkillManual->LU->HBW	-2.1	-1.1
	SkillManual->LU->NoCar->HBW	-3.1	-1.3
	SkillManual->NoCar->HBW	-7.2	-3.8
Prof->HBW		not significant	3.7
	Prof->LU->HBW	-1.9	-1.1
	Prof->LU->NoCar->HBW	-2.9	-1.3
	Prof->NoCar->HBW	-7.5	-3.2
IncomeLess25k->HBW		-6.4	-10.2
	IncomeLess25k->NoCar->HBW	13.9	6.1
IncomeOver50K->HBW		8	9.8
	IncomeOver50K->LU->HBW	0.6	0.3
	IncomeOver50K->LU->NoCar->HBW	1	0.3
	IncomeOver50K->NoCar->HBW	-6.1	-2
LU->HBW		10.4	5.8
	LU->NoCar->HBW	15.4	6.7

Male->Sh		-11.8	-11.6
	Male->HBW->Sh	-0.3	-0.3
FT->Sh		-5	-5
	FT->LU->Sh	-0.2	-0.2
	FT->LU->HBW->Sh	0	0
	FT->LU->NoCar->HBW->Sh	1.5	0
	FT->HBW->Sh	-0.9	-0.9
1adult->Sh		2.4	2.5
	1adult->LU->Sh	-0.5	-0.5
	1adult->LU->HBW->Sh	0	0
	1adult->LU->NoCar->HBW->Sh	not significant	0
	1adult->NoCar->HBW->Sh	-0.3	-0.1
Manual->Sh		-3.8	-3.8
	Manual->HBW->Sh	-0.3	0.1
	Manual->NoCar->HBW->Sh	-0.2	-0.1
SkillManual->Sh		-3.7	-3.6
	SkillManual->LU->Sh	0.5	0.5
	SkillManual->LU->HBW->Sh	0.1	0
	SkillManual->LU->NoCar->HBW->Sh	-0.07	0
	SkillManual->HBW->Sh	0.2	0.2
	SkillManual->NoCar->HBW->Sh	0.2	0.1
Prof->Sh		not significant	not significant
	Prof->LU->Sh	0.4	0.5
	Prof->LU->HBW->Sh	0.1	0
	Prof->LU->NoCar->HBW->Sh	0.1	0
	Prof->NoCar->HBW->Sh	0.2	0.1
IncomeLess25k->Sh		not significant	not significant
	IncomeLess25k->HBW->Sh	0.1	0.2
	IncomeLess25k->NoCar->HBW->Sh	-0.3	-0.1
IncomeOver50K->Sh		not significant	not significant
	IncomeOver50K->LU->Sh	-0.1	-0.1
	IncomeOver50K->HBW->Sh	-0.2	-0.2
	IncomeOver50K->NoCar->HBW->Sh	0.1	0
LU->Sh		-2.4	-2.5
	LU->HBW->Sh	-0.2	-0.1
	LU->NoCar->HBW->Sh	-0.3	-0.1
NoCar->Sh		not significant	not significant
	NoCar->HBW->Sh	-0.6	-0.2
1adult->Oth		40.8	50.3
	1adult->NoCar>Oth	-17.5	-13.9
IncomeLess25k->Oth		-17.6	-6.4
	IncomeLess25k->NoCar>Oth	-18	-14
LU->Oth		-4.4	7.9
	LU->NoCar>Oth	-19.9	-15.4
FT->Oth		-14.7	-15.8
	FT->HBW>Oth	-12.4	-11.2

Table 35 Estimation of indirect effects on car ownership, work and shopping travel distance (travel distance unit is in miles)

Direct Effect	Indirect effect	ML Estimate	WLS Estimate
FT->NoCar		-0.003	Not significant
	FT->LU->NoCar	0.060	0.059
1adult->Nocar		0.531	0.535
	1adult->LU->NoCar	0.12	0.117
SkillManual->NoCar		-0.282	-0.322
	SkillManual->LU->NoCar	-0.12	-0.117
Prof->NoCar		-0.297	-0.285
	Prof->LU->NoCar	-0.11	-0.111
IncomeOver50K->NoCar		-0.242	-0.176
	IncomeOver50K->LU->NoCar	0.039	0.028
FT->HBW		17	1.7
	FT->LU->HBW	-0.3	-1.5
	FT->LU->NoCar->HBW	-0.2	-0.2
1adult->HBW		3	4.6
	1adult->LU->HBW	-0.7	-2.9
	1adult->LU->NoCar->HBW	-0.5	-4.5
	1adult->NoCar->HBW	-2	-2
Manual->HBW		-3	-1.9
	Manual->LU->HBW	0.19	not significant
	Manual->LU->NoCar->HBW	-13	not significant
	Manual->NoCar->HBW	-1.6	-1.5
SkillManual->HBW		-4	-5.6
	SkillManual->LU->HBW	0.7	0.3
	SkillManual->LU->NoCar->HBW	0.5	0.05
	SkillManual->NoCar->HBW	1	1.3
Prof->HBW		3	1.6
	Prof->LU->HBW	0.6	0.28
	Prof->LU->NoCar->HBW	0.4	0.43
	Prof->NoCar->HBW	1.1	1.1
IncomeLess25k->HBW		-4	-2.66
	IncomeLess25k->NoCar->HBW	-2.1	-2.1
IncomeOver50K->HBW		5	4.1
	IncomeOver50K->LU->HBW	-0.2	not significant
	IncomeOver50K->LU->NoCar->HBW	-0.14	-0.11
	IncomeOver50K->NoCar->HBW	0.9	0.68
LU->HBW		-3	-1.5
	LU->NoCar->HBW	-2.3	-2.3
Male->Sh		-3	-3
	Male->HBW->Sh	not significant	-0.09
FT->Sh		-1	-0.77
	FT->LU->Sh	-0.3	-0.08
	FT->LU->HBW->Sh	not significant	0
	FT->LU->NoCar->HBW->Sh	not significant	0
	FT->HBW->Sh	not significant	-0.15
1adult->Sh		1	3
	1adult->LU->Sh	-0.7	-0.16
	1adult->LU->HBW->Sh	not significant	0
	1adult->LU->NoCar->HBW->Sh	not significant	0

Direct Effect	Indirect effect	ML Estimate	WLS Estimate
	1adult->NoCar->HBW->Sh	not significant	0.02
Manual->Sh		-1	not significant
	Manual->HBW->Sh	not significant	not significant
	Manual->NoCar->HBW->Sh	not significant	0.01
SkillManual->Sh		-1	-2.64
	SkillManual->LU->Sh	0.7	0.016
	SkillManual->LU->HBW->Sh	not significant	0
	SkillManual->LU->NoCar->HBW->Sh	not significant	0
	SkillManual->HBW->Sh	not significant	0.05
	SkillManual->NoCar->HBW->Sh	not significant	-0.01
Prof->Sh		not significant	-1.4
	Prof->LU->Sh	0.6	0.15
	Prof->LU->HBW->Sh	not significant	0
	Prof->LU->NoCar->HBW->Sh	not significant	0
	Prof->NoCar->HBW->Sh	not significant	-0.01
IncomeLess25k->Sh		-1	1.8
	IncomeLess25k->HBW->Sh	not significant	0.02
	IncomeLess25k->NoCar->HBW->Sh	not significant	0.02
IncomeOver50K->Sh		not significant	not significant
	IncomeOver50K->LU->Sh	-0.2	-0.04
	IncomeOver50K->HBW->Sh	not significant	-0.04
	IncomeOver50K->NoCar->HBW->Sh	not significant	-0.01
LU->Sh		-3	-0.83
	LU->HBW->Sh	not significant	0.01
	LU->NoCar->HBW->Sh	not significant	0.02
NoCar->Sh		-3	-5
	NoCar->HBW->Sh	not significant	0.03
1adult->Oth		21	33
	1adult->NoCar>Oth	-12.8	-15.2
IncomeLess25k->Oth		-11	not significant
	IncomeLess25k->NoCar>Oth	-13.1	-15.4
LU->Oth		-12	not significant
	LU->NoCar>Oth	-14.6	-16.8
FT->Oth		2	3
	FT->HBW>Oth	-2.3	-3.2

Table 36 Estimation of indirect effects on car ownership, work and shopping trip frequency

Direct Effect	Indirect effect	ML Estimate	WLS Estimate
FT->NoCar		-0.002	Not significant
	FT->LU->NoCar	0.059	0.059
1adult->Nocar		0.532	0.531
	1adult->LU->NoCar	0.121	0.116
SkillManual->NoCar		-0.283	-0.331
	SkillManual->LU->NoCar	-0.123	-0.117
Prof->NoCar		-0.296	-0.285
	Prof->LU->NoCar	-0.113	-0.111
IncomeOver50K->NoCar		-0.242	-0.172

Direct Effect	Indirect effect	ML Estimate	WLS Estimate
	IncomeOver50K->LU->NoCar	0.038	0.028
FT->HBW		0.410	1.13
	FT->LU->HBW	0.002	not significant
	FT->LU->NoCar->HBW	0.007	0.011
1adult->HBW		-0.068	-0.266
	1adult->LU->HBW	0.005	Not significant
	1adult->LU->NoCar->HBW	0.014	0.023
	1adult->NoCar->HBW	0.062	0.103
Manual->HBW		0.126	0.336
	Manual->LU->HBW	not significant	Not significant
	Manual->LU->NoCar->HBW	not significant	Not significant
	Manual->NoCar->HBW	0.048	0.076
SkillManual->HBW		not significant	0.071
	SkillManual->LU->HBW	-0.005	not significant
	SkillManual->LU->NoCar->HBW	-0.014	-0.023
	SkillManual->NoCar->HBW	-0.033	-0.064
Prof->HBW		not significant	not significant
	Prof->LU->HBW	-0.004	not significant
	Prof->LU->NoCar->HBW	-0.013	-0.021
	Prof->NoCar->HBW	-0.034	-0.055
IncomeLess25k->HBW		not significant	not significant
	IncomeLess25k->NoCar->HBW	0.063	0.105
IncomeOver50K->HBW		not significant	-0.193
	IncomeOver50K->LU->HBW	0.001	not significant
	IncomeOver50K->LU->NoCar->HBW	0.004	0.005
	IncomeOver50K->NoCar->HBW	-0.028	-0.033
LU->HBW		0.023	Not significant
	LU->NoCar->HBW	0.07	0.115
Male->Sh		-0.270	-0.549
	Male->HBW->Sh	not significant	not significant
FT->Sh		-0.148	-0.335
	FT->LU->Sh	-0.004	not significant
	FT->LU->HBW->Sh	not significant	not significant
	FT->LU->NoCar->HBW->Sh	not significant	-0.001
	FT->HBW->Sh	-0.013	-0.067
1adult->Sh		0.119	0.308
	1adult->LU->Sh	-0.008	not significant
	1adult->LU->HBW->Sh	not significant	not significant
	1adult->LU->NoCar->HBW->Sh	not significant	-0.001
	1adult->NoCar->HBW->Sh	-0.002	-0.006
Manual->Sh		-0.091	-0.113
	Manual->HBW->Sh	not significant	-0.020
	Manual->NoCar->HBW->Sh	-0.002	-0.004
SkillManual->Sh		-0.127	-0.301
	SkillManual->LU->Sh	0.008	not significant
	SkillManual->LU->HBW->Sh	not significant	not significant
	SkillManual->LU->NoCar->HBW->Sh	not significant	0.001
	SkillManual->HBW->Sh	not significant	not significant
	SkillManual->NoCar->HBW->Sh	0.001	0.004
Prof->Sh		not significant	-0.114
	Prof->LU->Sh	0.007	not significant
	Prof->LU->HBW->Sh	not significant	not significant

Direct Effect	**Indirect effect**	**ML Estimate**	**WLS Estimate**
	Prof->LU->NoCar->HBW->Sh	not significant	0.001
	Prof->NoCar->HBW->Sh	0.001	0.003
IncomeLess25k->Sh		not significant	0.106
	IncomeLess25k->HBW->Sh	not significant	not significant
	IncomeLess25k->NoCar->HBW->Sh	-0.002	-0.006
IncomeOver50K->Sh		not significant	-0.082
	IncomeOver50K->LU->Sh	-0.002	not significant
	IncomeOver50K->HBW->Sh	0.002	0.011
	IncomeOver50K->NoCar->HBW->Sh	0.001	0.002
LU->Sh		-0.039	not significant
	LU->HBW->Sh	-0.001	not significant
	LU->NoCar->HBW->Sh	-0.002	-0.007
NoCar->Sh		-0.183	-0.210
	NoCar->HBW->Sh	-0.004	-0.011
1adult->Oth		0.163	1.65
	1adult->NoCar>Oth	-0.218	-0.877
IncomeLess25k->Oth		-0.052	0.323
	IncomeLess25k->NoCar>Oth	-0.223	-0.896
LU->Oth		-0.047	0.449
	LU->NoCar>Oth	-0.248	-0.982
FT->Oth		-0.163	-1.281
	FT->HBW>Oth	-0.03	-0.513

B4. Variations in influences over time

Figures ... to ... below shows the trend of changes in significant influences over time from 2002 to 2010. In order to make comparison across years, influences and models, the coefficients for the year 2002 is set to 100 as the benchmark index and the rest are shown relative to those. The systematic changes over time are bolded and their associated trendline are shown with dotted lines. Reassuringly, the influences on car ownership has been very similar across the models so only those from travel distance model is shown.

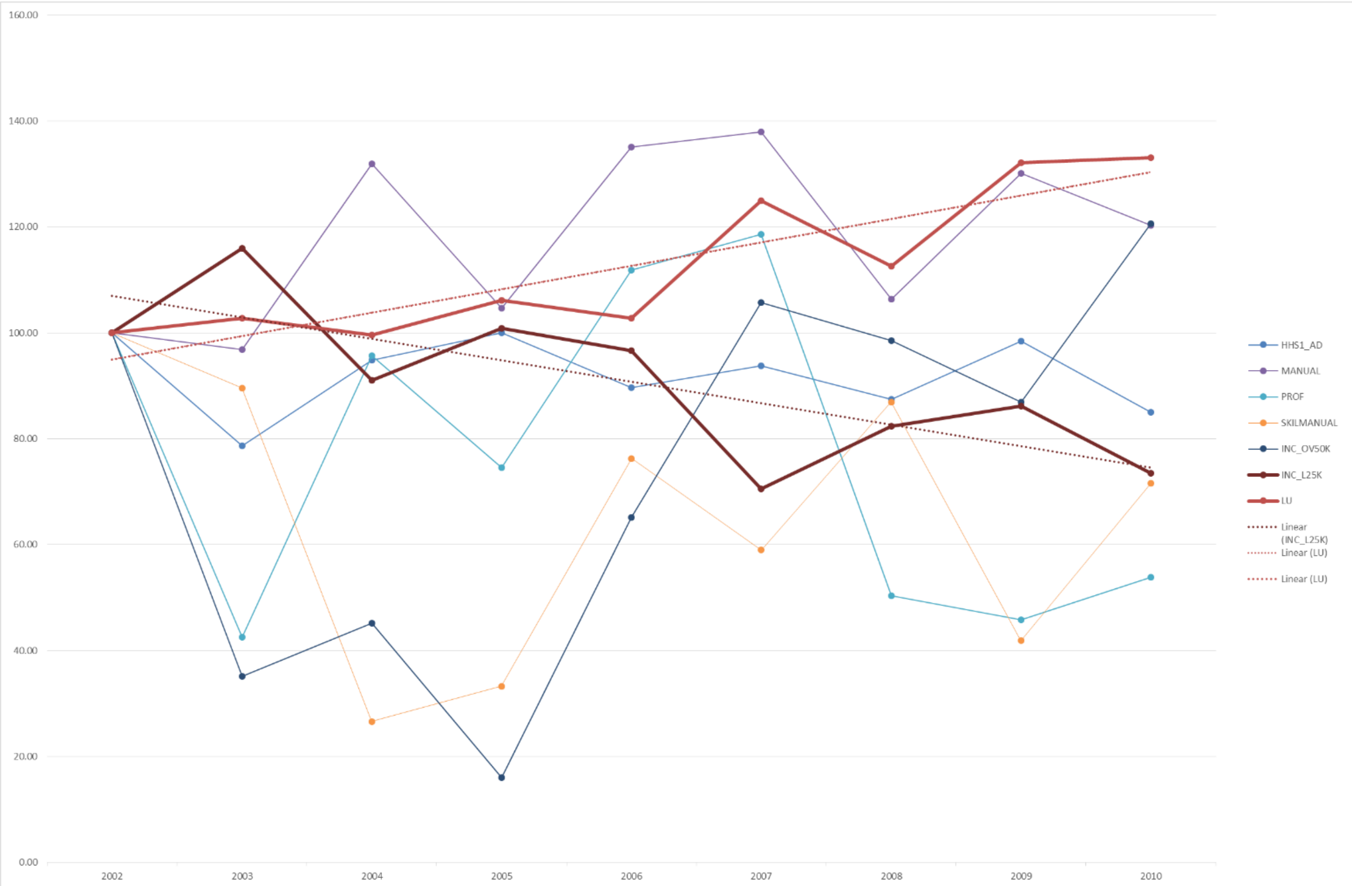

Figure 19 Trend of changes in influences on car ownership over time

Figure 20 Trend of changes in influences on Home Based Work travel distance over time

Figure 21 Trend of changes in influences on Shopping travel distance over time

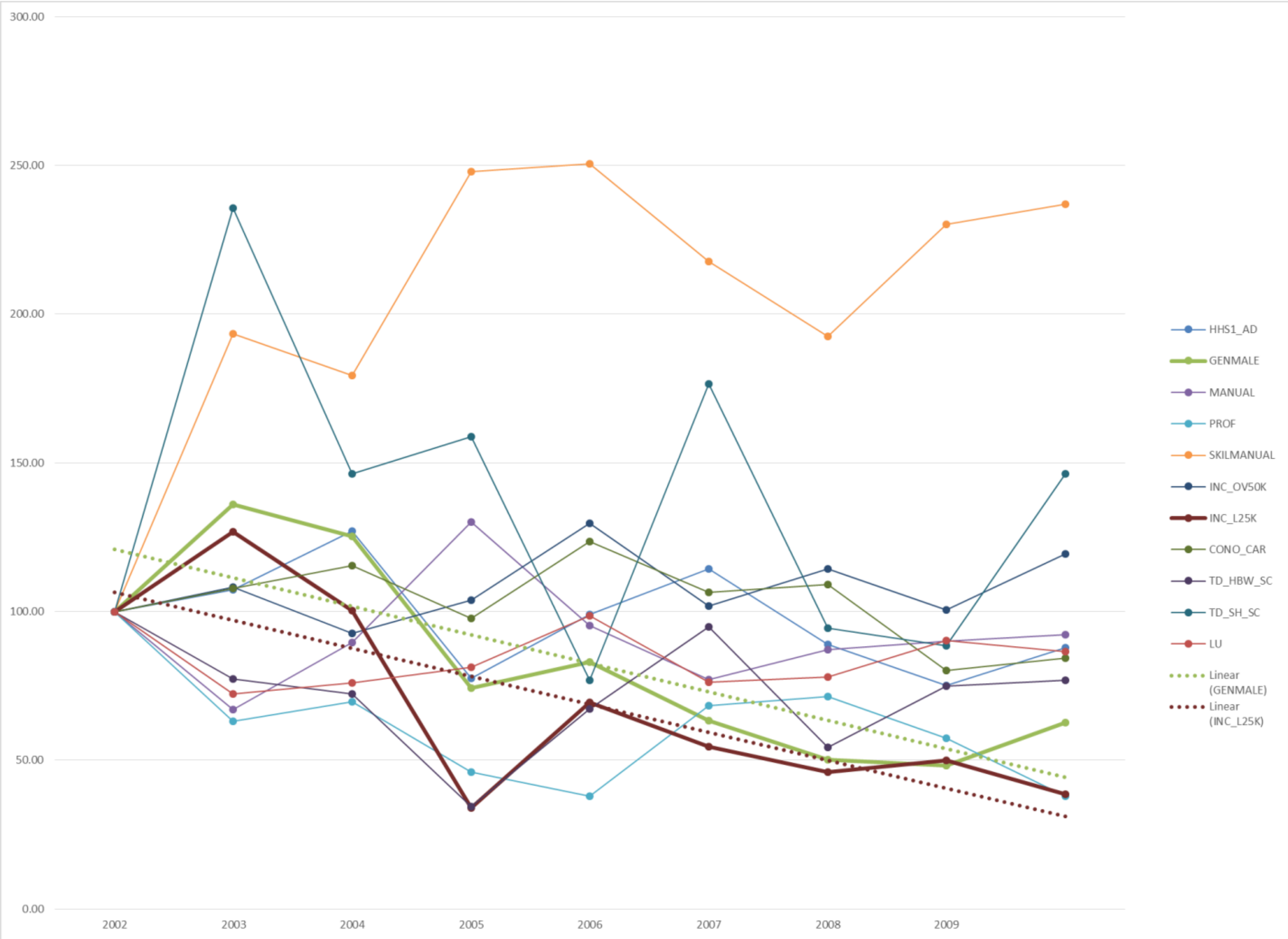

Figure 22 Trend of changes in influences on Other Purposes travel distance over time

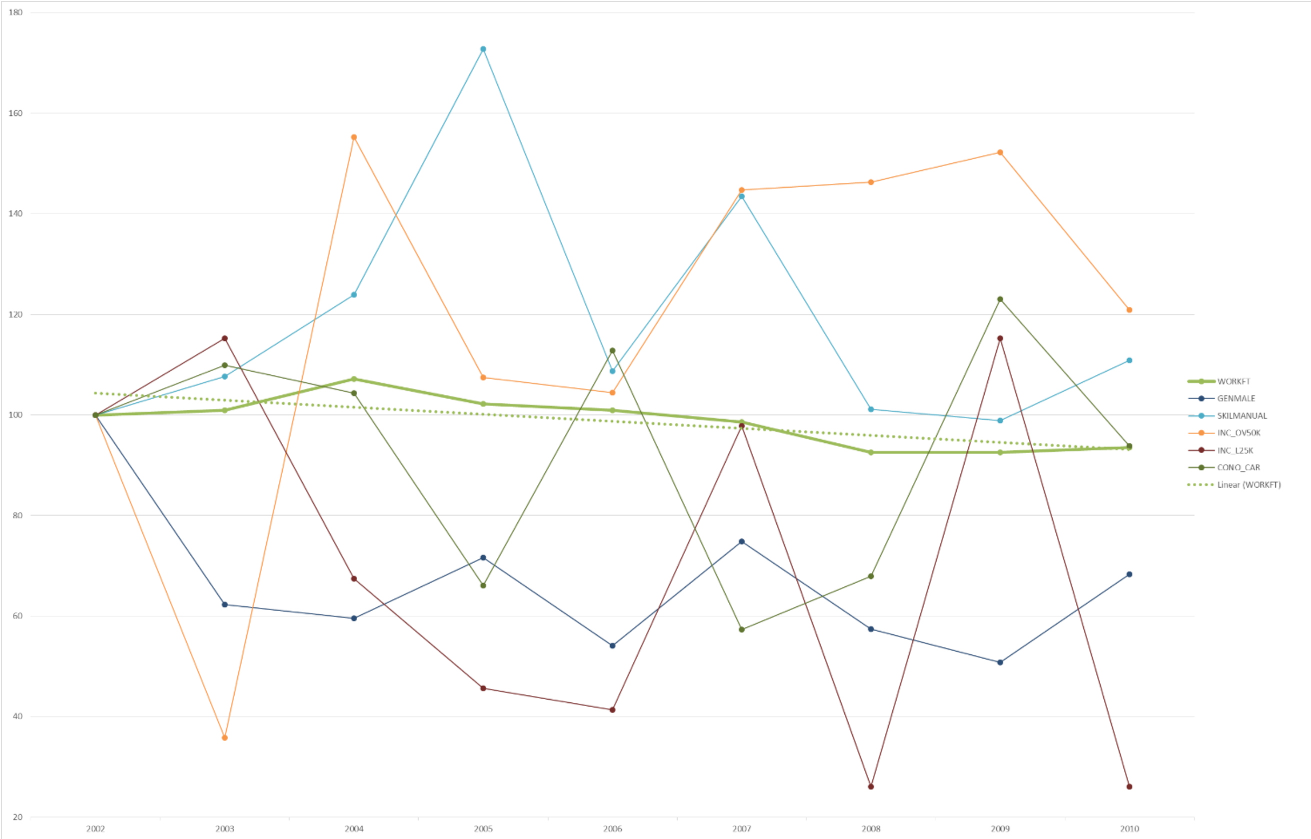

Figure 23 Trend of changes in influences on Home Based Work travel time over time

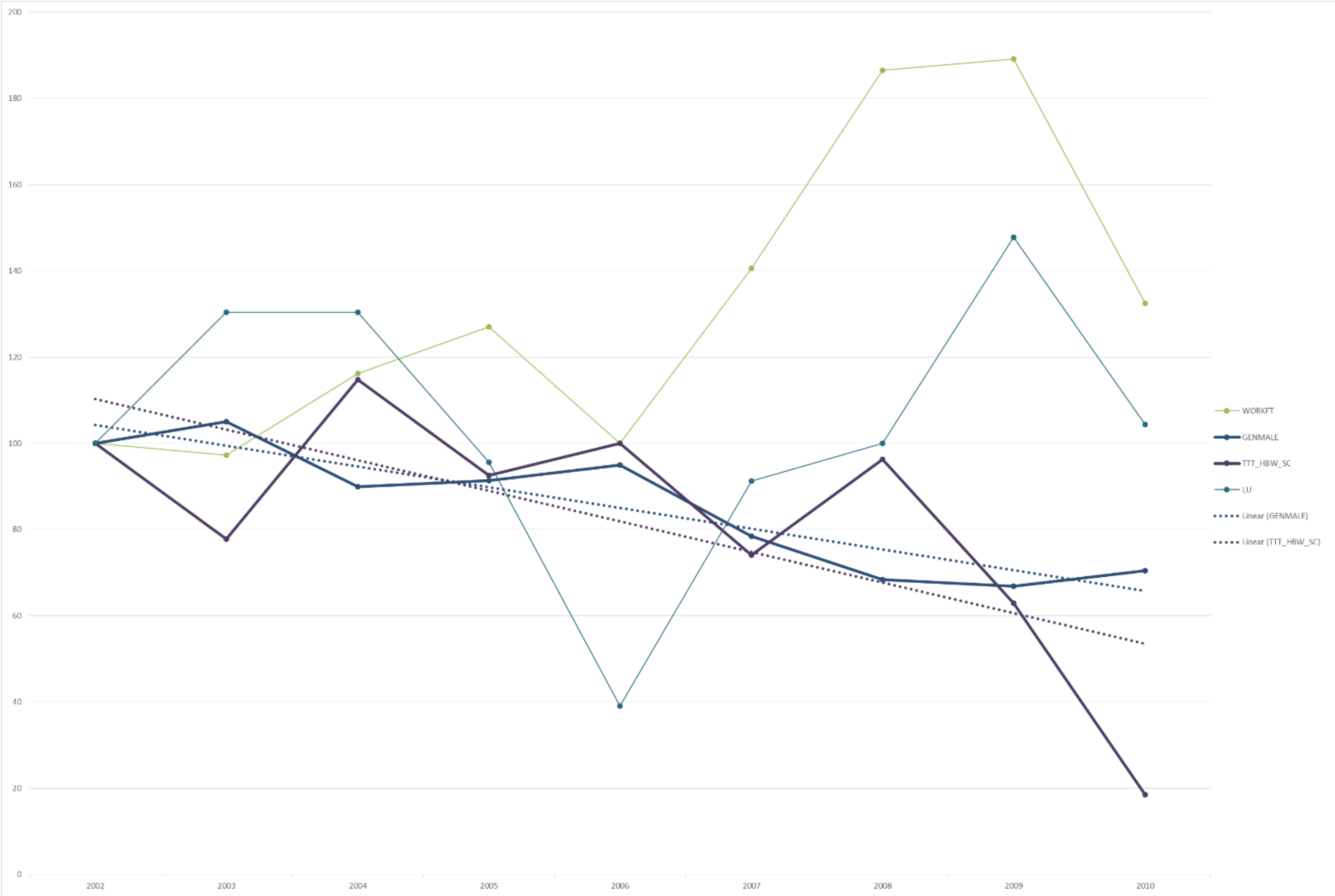

Figure 24 Trend of changes in influences on Shopping travel time over time

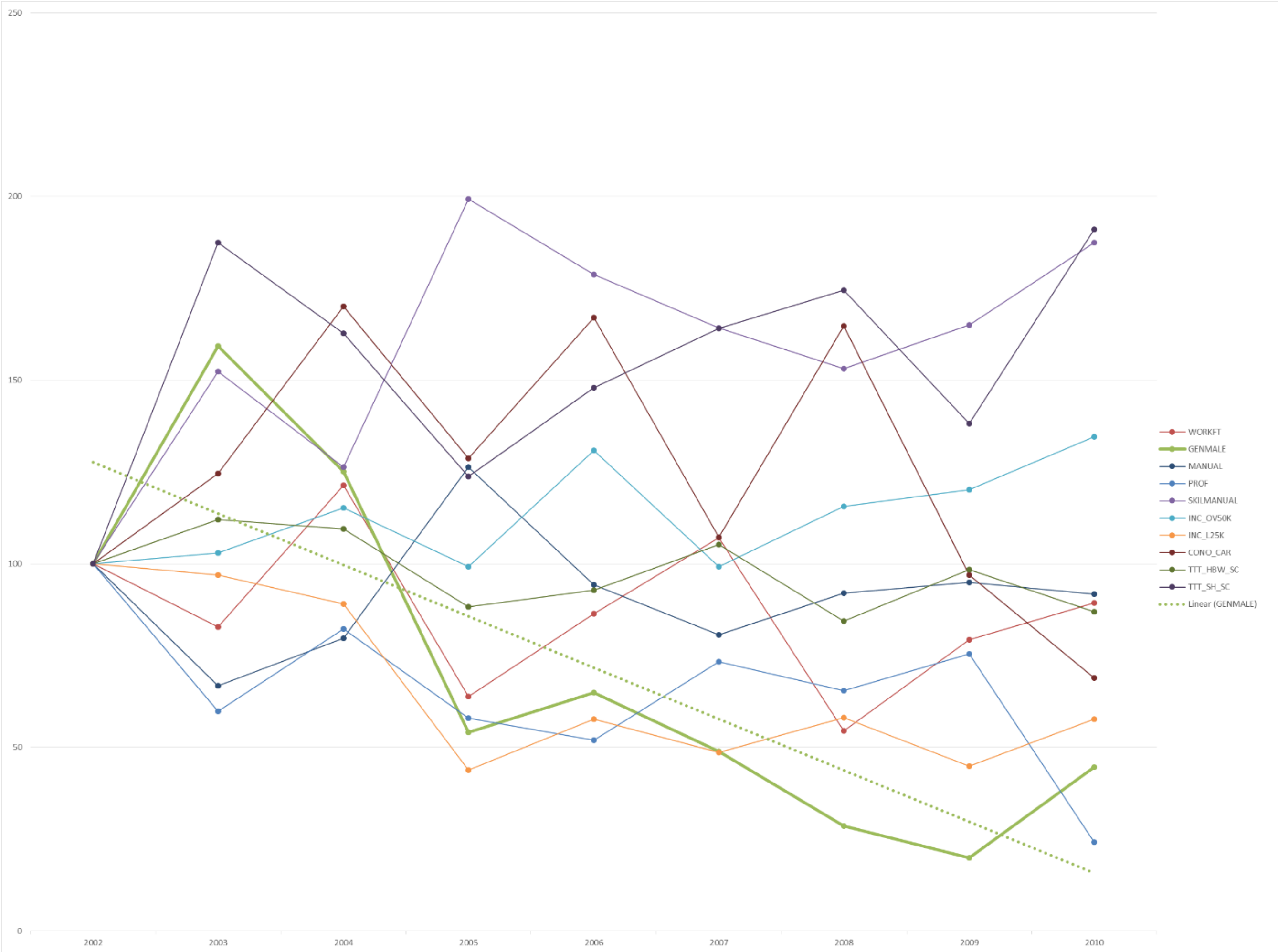

Figure 25 Trend of changes in influences on Other Purposes travel time over time

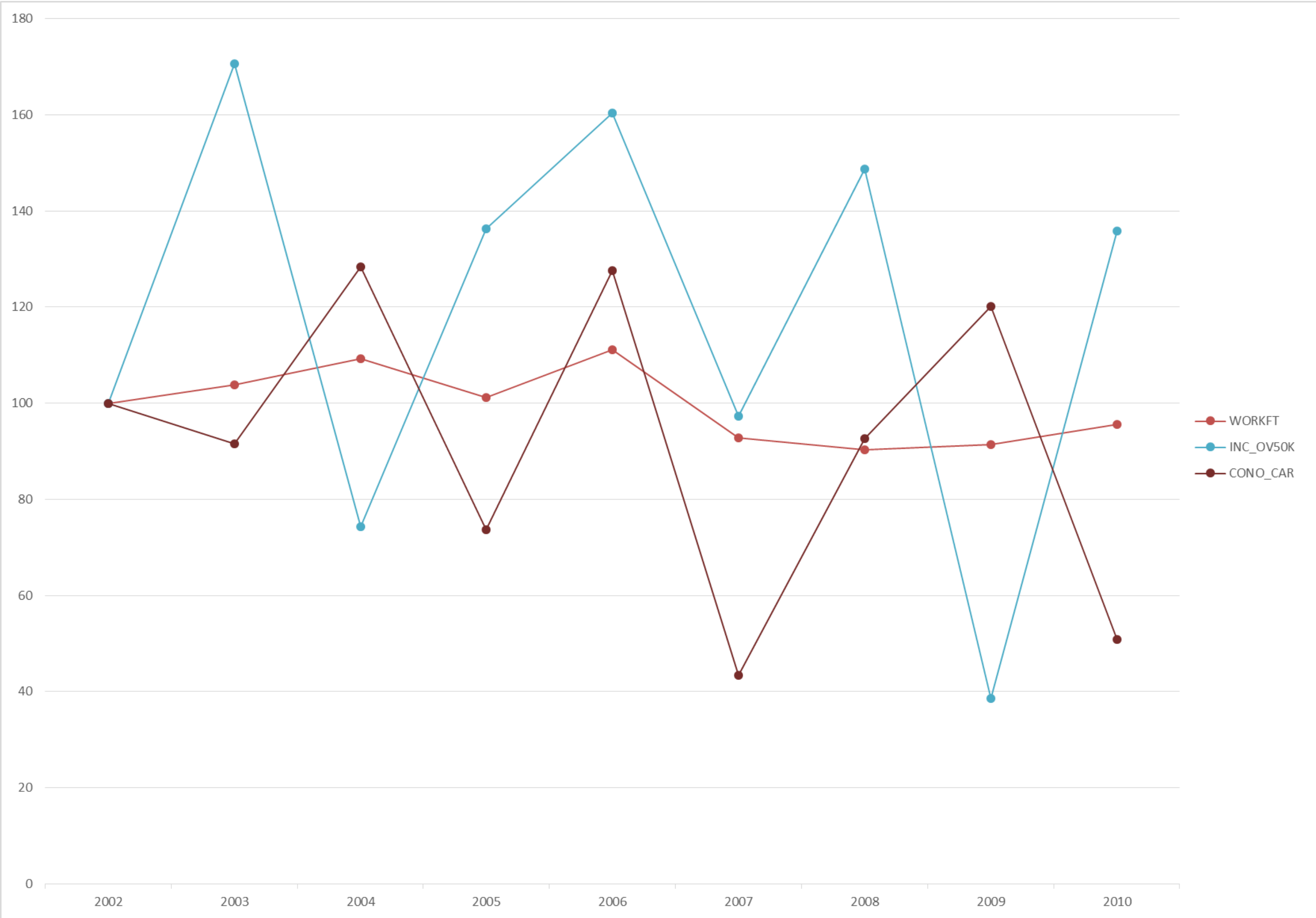

Figure 26 Trend of changes in influences on Home Based Work trip frequency over time

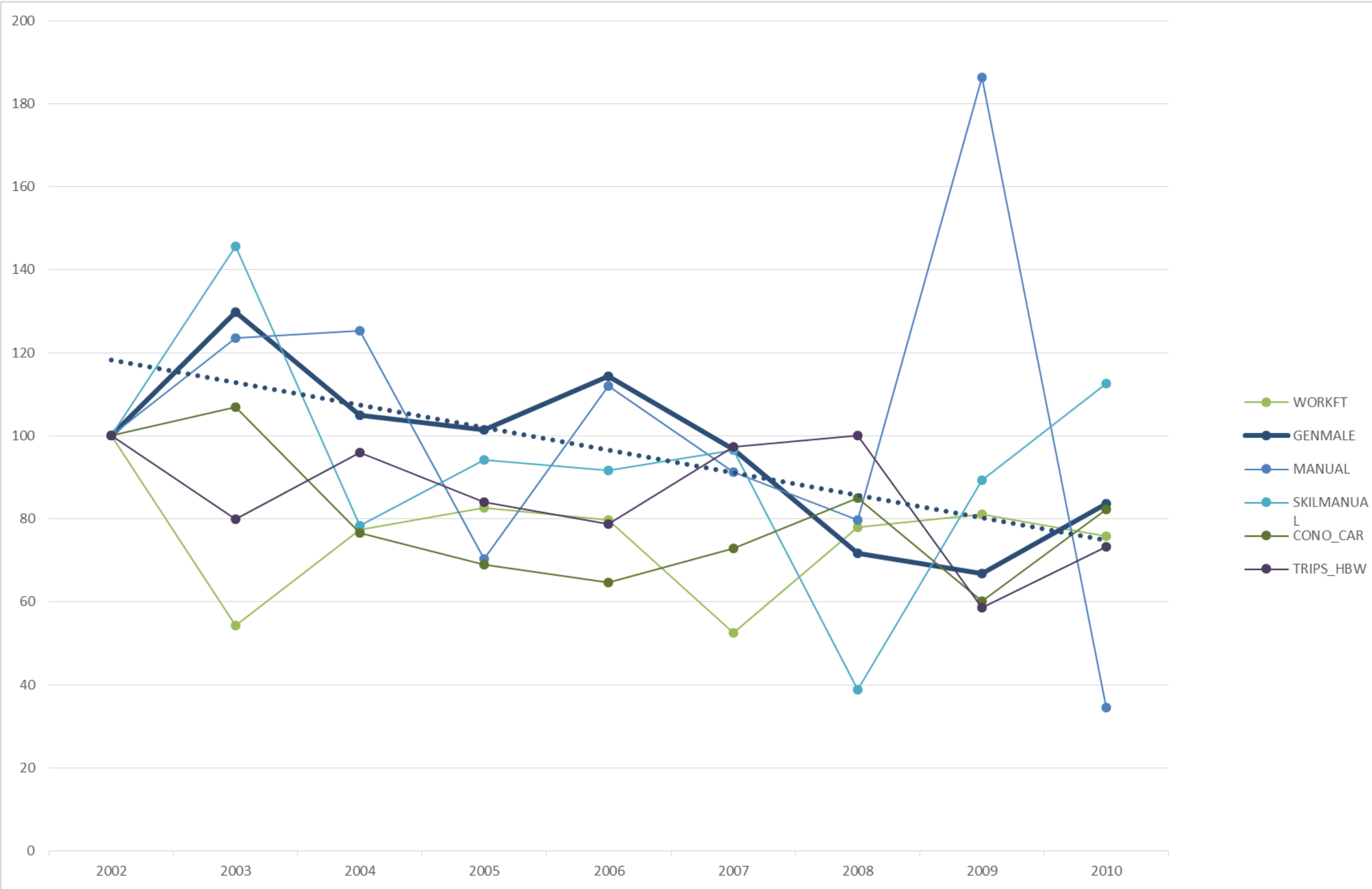

Figure 27 Trend of changes in influences on Shopping trip frequency over time

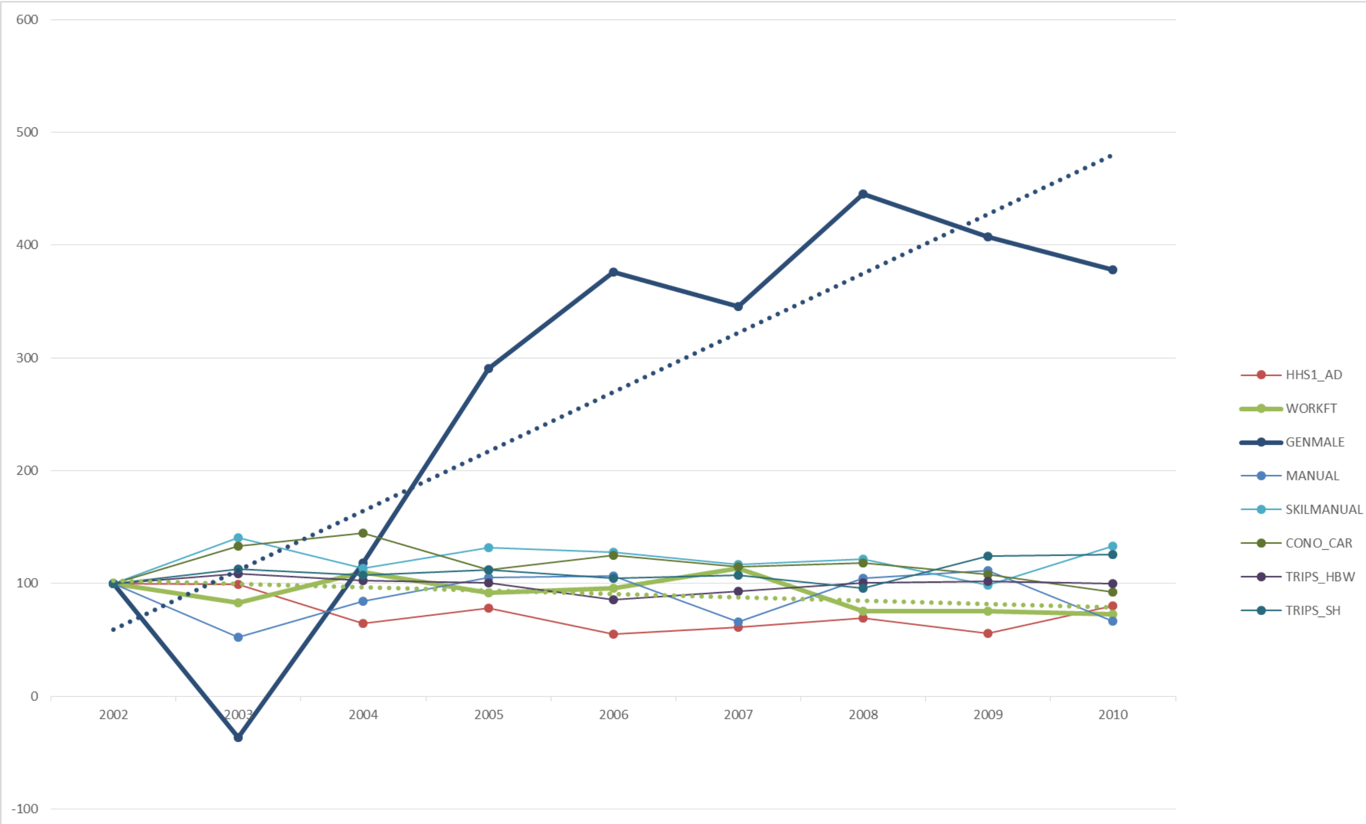

Figure 28 Trend of changes in influences on Other Purposes trip frequency over time

B5. WLS estimation and graphical representation of the travel time model coefficient changes before and after year 2007

Table 37 compares the travel time estimated coefficients before and after 2007 by the WLS estimator.

Table 37 The comparison of estimation before and after 2007 by WLS estimator (travel time unit is in minutes)

Direct Effect		Estimate (P.Value) After 2007	Estimate (P.Value) Before 2007
LU Measured BY			
	AREATYPE	2.045 (0.000)	1.983 (0.000)
	BUS_FREQ	1.079 (0.000)	1.11 (0.000)
	POPDEN	1.729 (0.000)	1.783 (0.000)
LU regressed ON			
	GENMALE	-0.014 (0.261)	-0.016 (0.104)
	WORKFT	0.092 (0.000)	0.104 (0.000)
	HHS1_AD	0.162 (0.000)	0.214 (0.000)
	MANUAL	0.008 (0.821)	-0.068 (0.011)
	SKILMANUAL	-0.237 (0.000)	-0.168 (0.000)
	PROF	-0.216 (0.000)	-0.178 (0.000)
	INC_L25K	0.044 (0.107)	0.029 (0.153)
	INC_OV50K	0.071 (0.007)	0.048 (0.028)
NoCar regressed ON			
	GENMALE	0.04 (0.123)	-0.026 (0.207)
	WORKFT	-0.017 (0.686)	0.019 (0.512)
	HHS1_AD	0.528 (0.000)	0.535 (0.000)
	MANUAL	0.388 (0.000)	0.372 (0.000)
	SKILMANUAL	-0.35 (0.000)	-0.326 (0.000)
	PROF	-0.18 (0.001)	-0.312 (0.000)
	INC_L25K	0.479 (0.000)	0.599 (0.000)
	INC_OV50K	-0.254 (0.000)	-0.079 (0.135)
	LU	0.652 (0.000)	0.544 (0.000)
TT_HBW regressed ON			
	GENMALE	10.4 (0.000)	12.7 (0.000)
	WORKFT	38.8 (0.000)	42.6 (0.000)
	HHS1_AD	-2.1 (0.302)	-5.7 (0.000)
	MANUAL	-3 (0.178)	-9 (0.000)
	SKILMANUAL	-6.5 (0.001)	-8.7 (0.000)
	PROF	5.7 (0.001)	3.1 (0.022)
	INC_L25K	-8.5 (0.000)	-11.3 (0.000)
	INC_OV50K	12 (0.000)	7.6 (0.000)
	LU	6.1 (0.000)	5.5 (0.000)
	CONO_CAR	11.5 (0.000)	12.1 (0.000)
TT_Sh regressed ON			
	GENMALE	-9.6 (0.000)	-13.4 (0.000)
	WORKFT	-6.2 (0.000)	-4 (0.000)

Direct Effect		**Estimate (P.Value) After 2007**	**Estimate (P.Value) Before 2007**
	HHS1_AD	1.9 (0.053)	2.2 (0.02)
	MANUAL	-4.1 (0.000)	-3.7 (0.000)
	SKILMANUAL	-1.3 (0.206)	-4.8 (0.000)
	PROF	-0.2 (0.855)	-1.5 (0.059)
	INC_L25K	1.3 (0.191)	-1 (0.245)
	INC_OV50K	0.5 (0.595)	0.7 (0.38)
	LU	-3.6 (0.000)	-2.1 (0.000)
	CONO_CAR	1.7 (0.006)	-0.4 (0.393)
	TT_HBW	-1.7 (0.000)	-2.6 (0.000)
TT_Oth regressed ON			
	GENMALE	7.5 (0.001)	22.6 (0.000)
	WORKFT	-13.1 (0.000)	-15.7 (0.000)
	HHS1_AD	40.9 (0.000)	48.8 (0.000)
	MANUAL	-40.5 (0.000)	-39.3 (0.000)
	SKILMANUAL	-49.9 (0.000)	-46.6 (0.000)
	PROF	13.6 (0.000)	15.1 (0.000)
	INC_L25K	-10 (0.011)	-12.6 (0.000)
	INC_OV50K	26.8 (0.000)	25 (0.000)
	LU	1.5 (0.514)	2.3 (0.185)
	CONO_CAR	-14 (0.000)	-18.5 (0.000)
	TT_HBW	-28 (0.000)	-32 (0.000)
	TT_SH	37.7 (0.000)	31.4 (0.000)

Figure 29 to Figure 33 provide additional comparative analysis by graphically presenting the ML output of changes in the travel time model coefficients before and after the year 2007 (travel time are in minutes). For each path arrow, the first line is the coefficient value post 2007 together with its p-value, and similarly the second line reports those for pre 2007. Numbers in square brackets, which are only reported for some arrows, are the Wald test p-values for measuring indifferences – if it is below 0.01 then the shift in coefficient value is statistically significant.

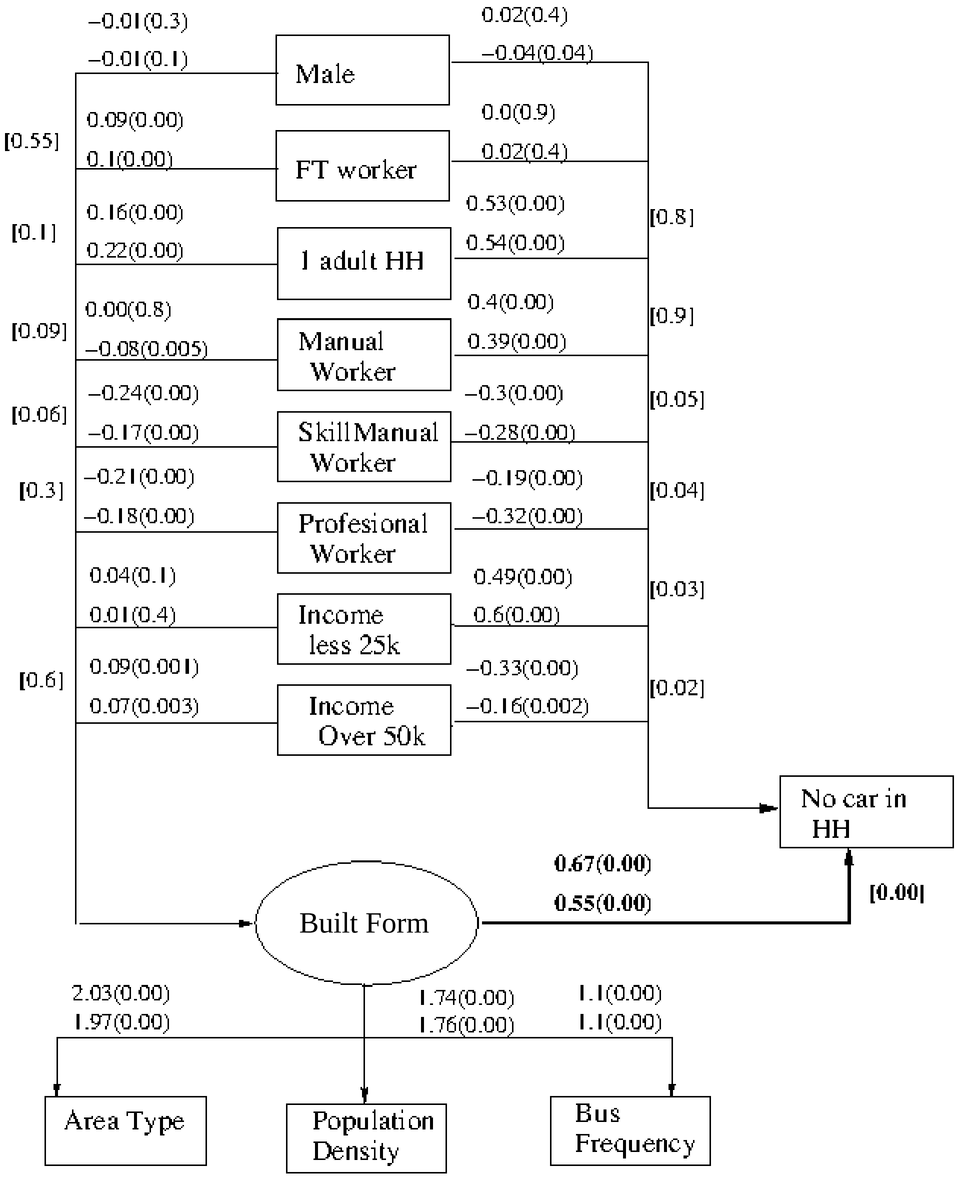

Figure 29 Comparing influences on travel time before and after 2007-on Car Ownership

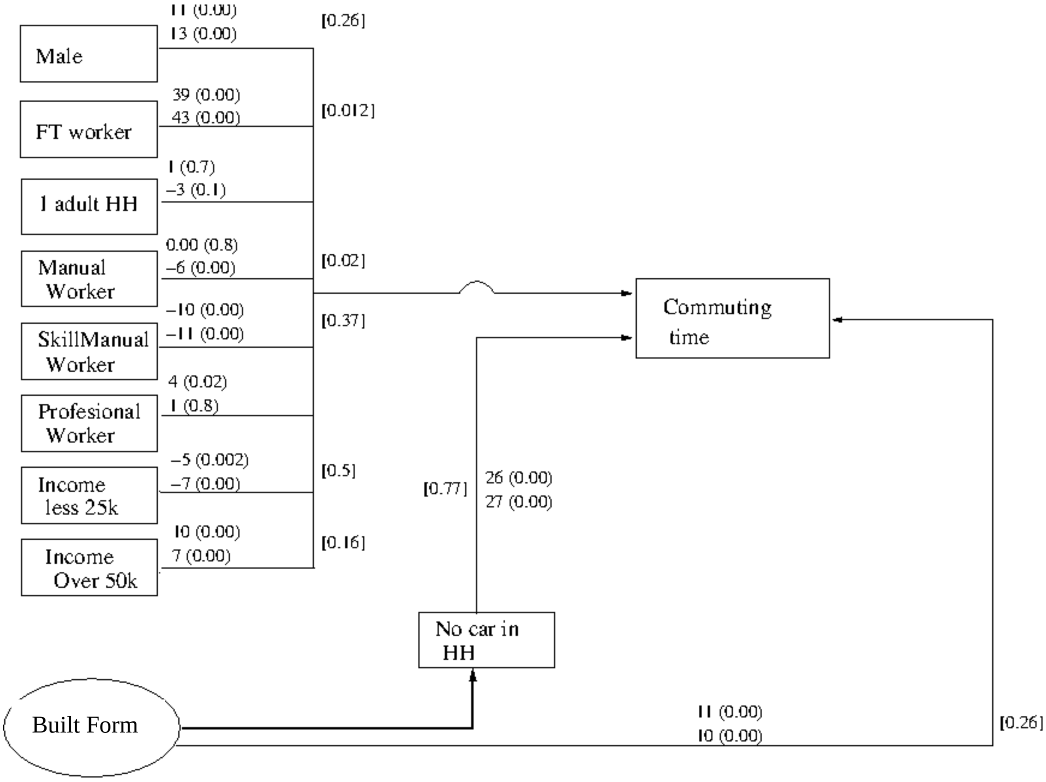

Figure 30 Comparing influences on travel time before and after 2007-on Commuting Time

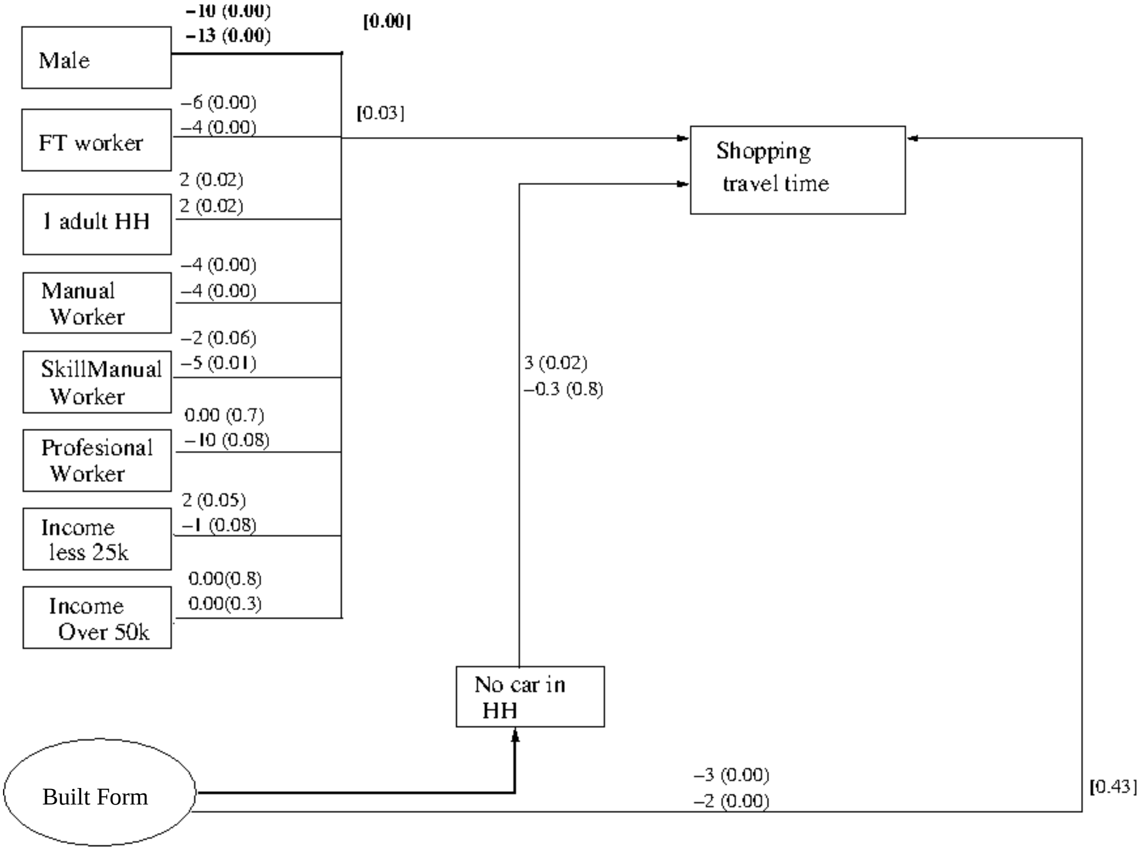

Figure 31 Comparing influences on travel time before and after 2007-on shopping travel time

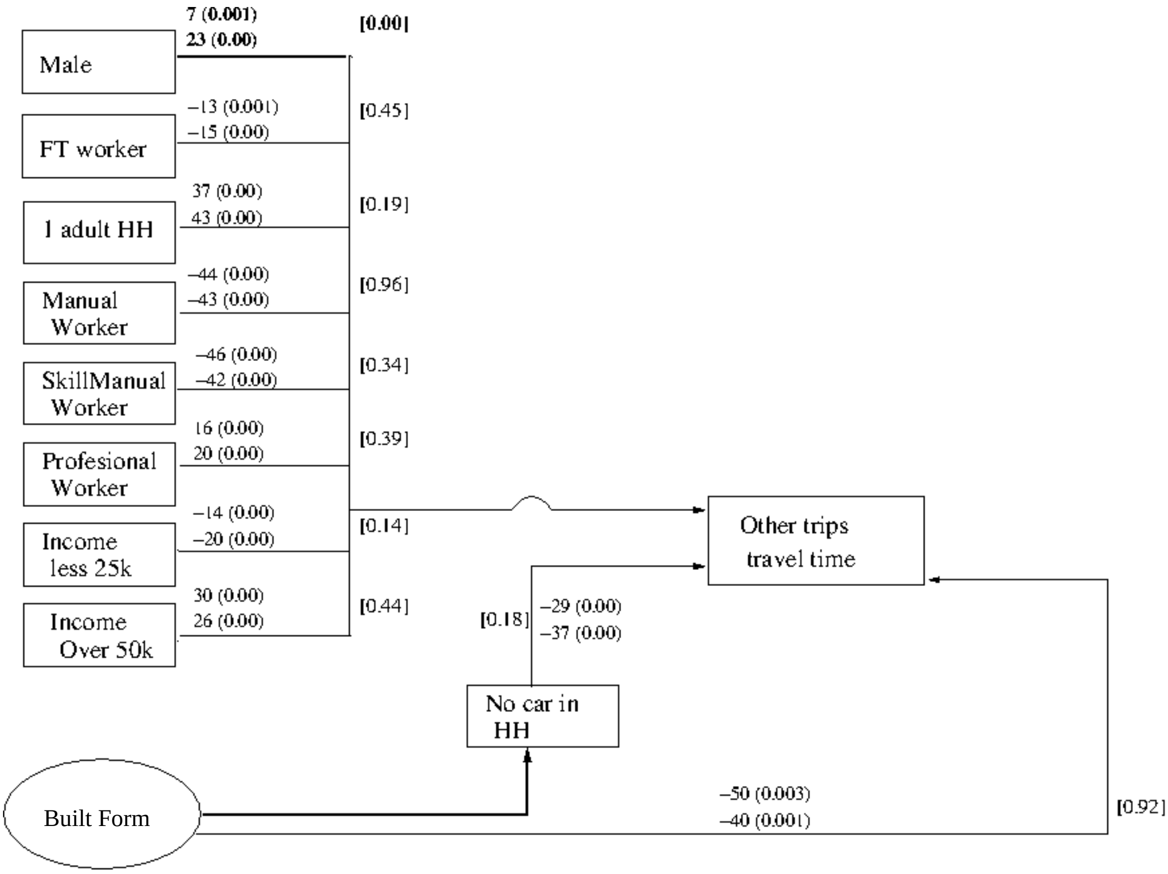

Figure 32 Comparing influences on travel time before and after 2007- on other trips travel time

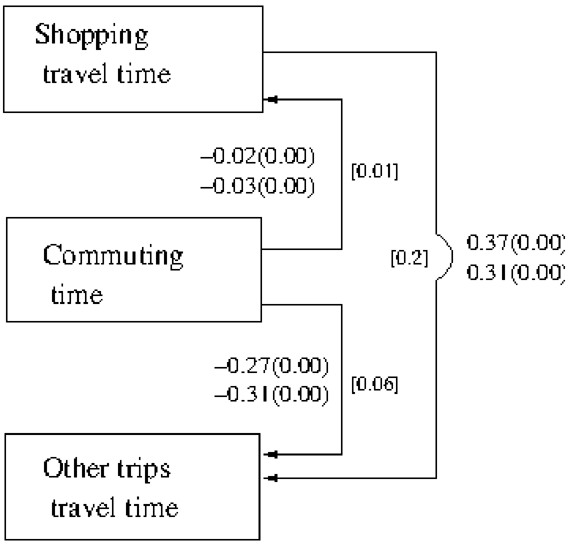

Figure 33 Comparing influences on travel time before and after 2007- interactions between travel purposes

B6. WLS goodness of fit statistics

In this dissertation, we have used Maximum likelihood (ML) estimator to estimate all models for the main results. However, the use of the ML means that absolute goodness of fit statistics such as chi-square (which are available from SEM models using WLS) are no longer available. Instead, the ML algorithm report relative fit statistics which in essence compare the logliklihood of different models against each other. To provide an approximate benchmark of the model fit, Table 38 below reports the absolute goodness of fit for the path-diagram based SEM of travel time using the WLS. The parameter estimation of this WLS model is reported in Table 37.

Table 38 Path diagram based SEM goodness of fit statistics for Constrained vs Grouped model (travel time)

	Chi-Square value (P-Value)/df	RMSEA[39]	Chi-Square value of base line model/df
Constrained model	407.831 (0.00)	0.006	108036.549
Grouped model	357.517 (0.00)	0.010	108036.549

WLS analysis of LCA-SEM and two-level SEM is not feasible due to the complexity of the models.

[39] Root Mean Square Error of Approximation – Maccallum et al (1996) have used 0.01, 0.05, and 0.08 to indicate excellent, good, and mediocre fit, respectively.