Understanding the impact of built environment on travel behaviour with activity-based modelling: Evidence from Beijing

Lun Liu

Pembroke College

Department of Land Economy

University of Cambridge

This dissertation is submitted for the degree of Doctor of Philosophy

September 2017

Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text. It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. It does not exceed the prescribed word limit for the relevant Degree Committee.

The following part of the research is a collaborative work with another two scholars and has been published:

—Material derived from Chapter 4:

Liu L., Silva E.A., Wu C., Wang H. 2017. A machine learning-based method for the large-scale evaluation of the qualities of the urban environment. Computers Environment and Urban Systems, 65, 113-125.

Lun Liu

Pembroke Collage

Cambridge

September 2017

Acknowledgements

First and foremost, I would like to thank my supervisor, Dr Elisabete A Silva, for all she has done for me throughout the course of my PhD. It includes but is not limited to: showing interest in my research topic and admitting me to the PhD programme, equipping me with various research skills and mind sets, working close with me in our every (to be) published work, offering the opportunity to lecture in the Tripos and PGR programmes, helping refine the final thesis into a real one... Her own endeavour as a female faculty in pursuing the academic perfection is also a great encouragement to me—I indeed learned a lot.

This research is made possible through the support of many organisations and academians. One of the most important comes from the China Scholarship Council and the Cambridge Overseas Trust, who generously provided full scholarship (plus extra fieldwork funding) and made the PhD experience a financially worriless one. Prof Pete Tyler and Prof Douglas Crawford-Brown, my internal examiners, gave constructive feedbacks at several stages of the research. Dr Hui Wang and Chunyang Wu collaborated with me extensively in developing the machine learning algorithms in Chapter 4. Dr Hui Wang also shared his vast knowledge on the urban development of Beijing and gave much advice in interpreting and explaining the simulation results. I should also thank Dr Xuesong Gao and his students, for sparing their valuable time and helping with the data collection for the meta-analysis in Chapter 2. Last but not least, I thank all the staffs in Pembroke College, especially Dr Becky Coombs, for always responding to my needs and making my life in Cambridge as comfortable as possible.

Finally, thank you to all the friends that kept me accompany throughout these years: members of the LISA lab, Ransford, Helin, Yu, Qian, Chaowei, Felix; friends in other groups of the department and the university, Maomao, Aileen, Ziyou, Xiye, Qiaoling. Finally and finally, a big thank you to all my (extended) family members.

Summary

The built environment has long been considered as a potentially influential factor in shaping and changing people's travel behaviour. However, many gaps still exist in the understanding of the direction, size and mechanism of this influence. This thesis explores the complexities in the influence of the built environment on daily travel using a behaviour-oriented, activity-based modelling approach based on the notion of utility maximisation. The model simulates the full process of decision making in daily activity participation and travel, which involves the decisions on the type and frequency of activity participation, the sequence of activities, the choice of destinations and the time and mode of travel. Moreover, the thesis also addresses the lack of understanding on the influence of the 'third dimension' of the built environment — the street facades. A machine learning-based method is proposed to automatically evaluate the qualities of street facades from street view images.

Scenario analyses using the proposed model show that, both commute and non-commute travel are more sensitive to the built environment in proximity to home (in my experiment, 500 metre buffer zone). In the context of Beijing, the total car use and commute car use of a person is significantly affected by the level of land use mix and the continuity of street facades around home, among all built environment features. Non-commute car use is significantly affected by employment density, retail density, accessibility to commercial clusters, bus coverage, road density and the quality and continuity of street facades. Similar effects on the final outcomes of travel behaviour (such as total car use) by different built environment features can happen through diverse processes and have different implications for people's actual experience and the urban system. Some of the results are consistent with theoretical assumptions and some are not, which provides alternative insights into the relationship between the built environment and travel behaviour.

Table of Contents

List of Tables

List of Figures

Chapter 01 Introduction

1.1 Motivation and objectives

Transportation is one of the sources of many vexing urban problems, namely, congestion, pollution, inequality and reliance to fossil fuels (Hanson & Giuliano, 2004). Among the many approaches in tackling transport-related problems, a host of urban planning and design philosophies — new urbanism, transit-oriented development, traditional town planning—have gained popularity as ways of shaping travel demand (Cervero & Kockelman, 1997). The theory of consumer choice is used as the theoretical base for the potential influences of the built environment, by assuming that the built environment impacts on (relative) trip costs (Boarnet & Crane, 2001).

Despite of the appealing potential of the built environment in modifying travel behaviour, the true effects need to be materialised with robust empirical evidences. As a consequence, the relationship between the built environment and travel behaviour has been extensively examined and become one of the most heavily researched subjects in urban planning (Ewing & Cervero, 2010). More than two hundred studies have been produced since the 1990s and more are still emerging.

However, existing research tend to focus on the influences of the built environment on the synthesised outcomes of travel (e.g. VMT, walking distance, see the meta-analysis in Section 2.3), while the behavioural processes that give rise to these outcomes have received much less attention. For instance, if a built environment change is found to be related with 10% reduction of VMT, then whether the reduction comes from lower activity frequency, or the choice of closer destinations, or smaller share of driving, or etc. From the practical perspective, different behavioural processes would have different implications for people's actual experience and can be related with different policy goals. Besides, gaps also exist in terms of the quite mixed, sometimes contradictory results produced by existing research (Salon, Boarnet, Handy, Spears, & Tal, 2012), as well as the lack of research on fast growing, high density Asian cities (P. Zhao, Lu, & De Roo, 2011) (see the next section for a detailed description of research gaps). Practically, these gaps would also pose challenges to the reliability of the research findings for planning policy making, which is particularly an issue considering that built environment interventions are often costly and long-standing (Cao, 2015a). Therefore, more research efforts are needed in improving the understanding of the relationship between the built environment and travel behaviour from both the theoretical and practical points of view.

The aim of this thesis is thus to explore the influence of the built environment on travel with special emphasis on the behavioural processes and mechanisms. The research will use Beijing as the empirical case. The overarching research question is ***what impacts the urban built environment has on people's daily travel behavior***. The thesis is guided by the hypothesis that ***various aspects of daily travel behavior can be influenced by the built environment***. More detailed hypotheses in terms of the relationship between different pairs of built environment features and travel outcomes are put forward following a discussion on travel gains and costs in Section 2.1.4.

1.2 Research gaps and related questions

As mentioned before, despite of an overwhelming number of existing research on this topic, there still exist many research gaps. The gaps are related with multiple facets of research, including the methodology adopted, the variables and the cases used and the results. Some of the gaps will be further addressed in the literature review.

Gap in findings

The gap in the findings of existing research is twofold. First, as mentioned before, a large proportion of existing findings are about synthesised outcomes of travel, such as VMT, total walking time and so on (see the review by Cervero and Ewing, 2010). It is plausible since these are key indicators of travel behaviour that can be linked with more general policy goals such as energy consumption, emission reduction and public health. However, as explained in the last section, the behavioural processes that give rise to these outcomes and influences have received much less attention (e.g. the travel frequency, destination choice and mode choice that together lead to the outcome of VMT).

Second, there exist a lot of differences in terms of the direction, significance and magnitude of the impacts of the built environment on travel behaviour (see Cervero and Ewing, 2010 for a summary of empirical results produced before 2010 and Appendix A for results produced between 2010 and 2016). It is especially the case in terms of the magnitude of the impacts. For instance, the effect size of the population density in one's neighbourhood on VMT can range from -0.01 to -0.31, the effect size of land use diversity on VMT can range from -0.10 to -0.36 (these numbers are from only a subset of existing research, see Appendix A). Sources for the differences in findings include the strategies used for data collection, the measurements of built environment features, the statistical models, and more systematically, the varying nature of built environment-travel relationship in different urban contexts. This inconsistency causes a lot vagueness in the understanding on the built environment-travel relationship. Questions remain in terms of to which extent the built environment can direct people towards more sustainable patterns of daily travel, as well as the relative importance of various built environment factors in fulfilling this target (Joh, Nguyen, & Boarnet, 2012; Knuiman et al., 2014).

Gap in methodology

Related to the first gap in findings, a large proportion of existing research is conducted through regressions between synthesised outcomes of travel and a set of socioeconomic and built environment explanatory variables (see Appendix A for a summary of regressions in prior research). The regressions are methodologically sound and robust but when used alone, usually cannot probe into the detailed behavioural processes.

On the other hand, developments in the field of transport simulation and time geography gave rise to the activity-based transport modelling approach. It is underpinned by the notion that travel is derived from the necessity to participate in activities, which in turn reflect needs, desires and commitments of individuals and households, subject to a set of spatial, temporal, institutional, spatial–temporal and possibly budget constraints (Castiglione, Bradley, & Gliebe, 2015; Rasouli & Timmermans, 2014a). Although prototypes began to emerge as early as in the 1970s and substantial progress has been made in developing practical models since the 1990s (Rasouli & Timmermans, 2014a; Yasmin, Morency, & Roorda, 2015), the built environment factors are seldom sufficiently account for in the model systems (see **Table 2-5** in Section 2.4 for a summary of built environment variables included in existing activity-based models).

Actually, the activity-based modelling approach can be developed into a helpful tool for the analysis of built environment-travel relationship. The influences of the built environment can be modelled at each detailed choice facet in the activity-travel decision making process (e.g. frequency of activity participation, travel distance for a specific purpose), which will enable a more nuanced understanding on the behavioural mechanisms underlying the observed influences.

Gap in the built environment factors analysed

Built environment factors analysed in existing research are usually sorted as 'D' variables, which was first put forward in the seminal work by Cervero and Kockelman (1997) as 3 'D's and then extended to 5 'D's or 6'D's (Ewing & Cervero, 2001, 2010; Ewing & Handy, 2009). The 'D' variables are:

- Density (population density, employment density, building density, etc.)
- Diversity (land use mix, job-housing balance, etc.)
- Design (street density, intersection density, percentage of 4-way/3-way intersection, percentage of cul-de-sac, etc.)
- Destination accessibility (distance to the city centre, distance to 'regional' sub-centres, job accessibility by auto/transit in certain time limits, etc.)
- Distance to transit (distance to bus stops, distance to subway stations, etc.)
- Demand management (parking supply)

Although these factors already provide a well-rounded account of the built environment, they are mainly two-dimensional and land use-related, while the factors related to the dimension of street facade have received much less attention. These factors can be termed as the seventh 'D', the design of street facade. The potential mechanism of the these factors' influence on travel behaviour can be at least both psychical and functional (Montgomery, 1998; Southworth, 2005). Psychically, some qualities of the facade design may foster positive or negative feelings and thus encourage or discourage physical activities in the urban space (Sarkar et al., 2015; Witten et al., 2012). Functionally, the design and layout may also have an impact on the level of convenience and the utility of travel, e.g. providing much room for street shops at the ground floor. Nonetheless, it should be noted that it is not absolutely rigorous to categorise density and diversity as two-dimensional, since they can also be related to factors like building height or vertical mix.

Gap in empirical cases

By far, most empirical studies on this topic are from American and European (plus a few Oceanian) cities, while evidences from fast growing, high density cities in Asia are relatively scarce (Eom & Cho, 2015; Zegras, 2010; P. Zhao et al., 2011). Such a bias could also undermine the reliability and generalisability of the conclusions made from this line of research, considering that daily travel behaviour involves lots of contextual specificity (Feng, Dijst, Prillwitz, & Wissink, 2013). Contextual differences that may affect the relationship between built environment and travel behaviour include the level of car ownership, the level of transport service, affordability, social culture, etc. (Feng et al., 2013; Giuliano & Dargay, 2006). Therefore, many studies have argued that scholars should be careful regarding the temporal and spatial transferability of spatial policies (e.g. Badoe and Miller, 1995; Ewing, Tian et al., 2015; Naess, 2015). For example, Ewing, Tian et al. (2015) warned that a study using data from, say, Portland or Houston, can be challenged for relevance to other regions of the US.

A few studies have reported quite different results on the built environment-travel relationship in different countries. For instance, a large number of Asian cities are featured by much higher density comparing with American and European cities (H. Chen, Jia, & Lau, 2008; Madlener & Sunak, 2011). Eom and Cho (2015) found that the impact of high density on reducing car use is greatly reduced when gross density reaches beyond a certain threshold, and some other built environment factors also demonstrate more or less different effects. Giuliano and Dargay (2006) also pointed out that the widespread conviction that higher densities are associated with less travel distance is more pronounced in the US than in Britain. Nonetheless, only very limited research efforts have been made to query the differences in built environment-travel relationship in different urban contexts (e.g. Giuliano and Narayan, 2003; Guiliano and Dargay, 2006; Gim, 2013; Milakis, 2008; Senbil et al., 2009; Feng et al., 2013).

In order to fill in the gaps, the research aims to answer the following questions.

(1) **To fill the gap in findings** - What conclusions can be drawn regarding to the impacts of built environment on travel behaviour from existing empirical studies? What are the detailed behavioural processes underlying the built environment's influences on the synthesised travel outcomes? It should be noted that the answer to the second question can also be context dependent, thus it is impossible to reach an ultimate conclusion in one research alone. However, the contribution lies in

raising this issue and proposing a modelling approach to probe into it.

(2) **To fill the gap in findings & methodology** - How can the influence of the built environment on daily travel behaviour be comprehensively modelled in an activity-based approach?

(3) **To fill the gap in the built environment factors analysed** – How can the street facade features be reliably measured? How do they impact travel behaviour?

(4) **To fill the gap in findings & empirical cases** – What are the impacts of the built environment on travel behaviour in Beijing? How are the findings consistent with or different from those from American and European cities? How do the results contribute to the theoretical understanding on the relationship of the built environment and travel utility?

1.3 Choice of the case

The city of Beijing is chosen as the case of study in this research. It is chosen as an example of high-density and fast-growing Asian city, which provides a quite different urban context comparing with American and European cities that have been extensively studied. Since the early 1980s, in parallel with the economic boom, Beijing has been undergoing rapid urban growth. The built-up area increased from 1106.1 square kilometres in 1990 to 2416.5 square kilometres in 2010 and during the same period, the population increased from 5.8 million to 24.2 million. Such growth has made Beijing one of the most-densely resided cities in the world (**Figure 1-1**).

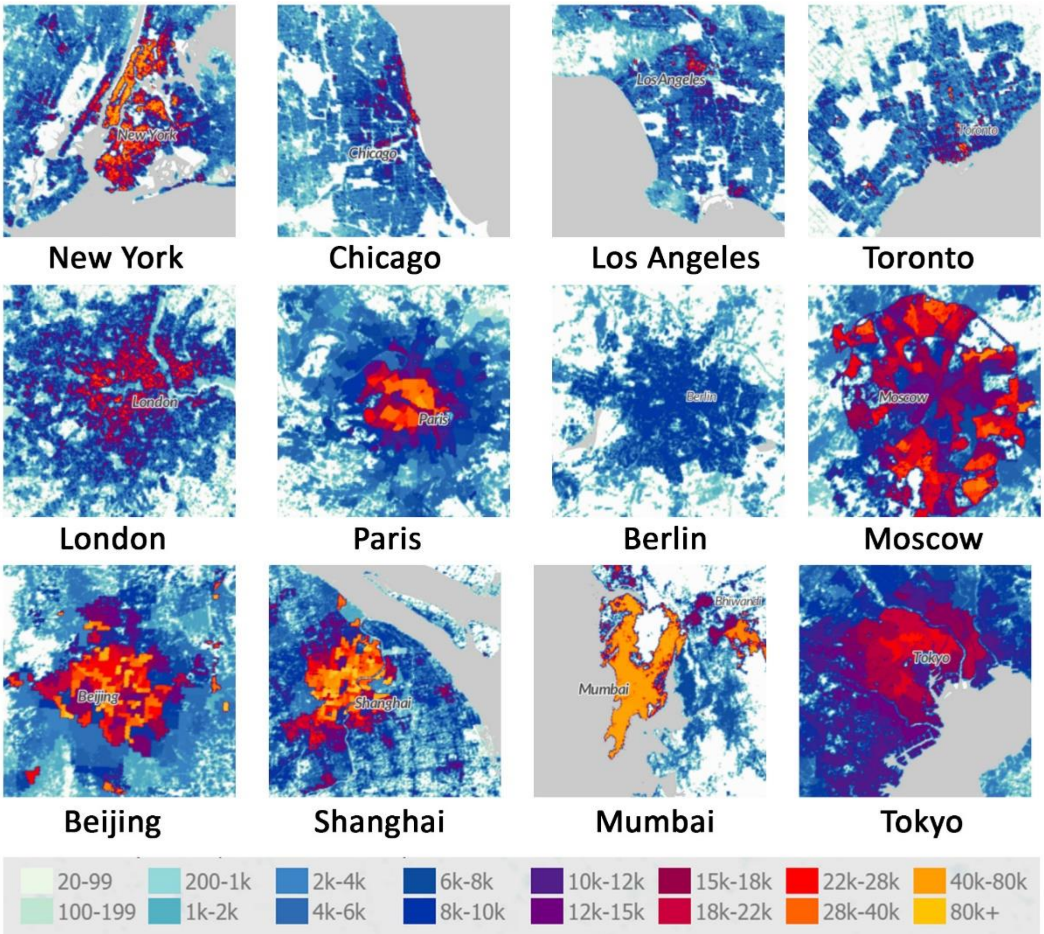

Figure 1-1 Urban population densities around the world

Note: residents per km^2, 2015, screen shot at the same scale

Source: http://luminocity3d.org/WorldPopDen/#9/43.7671/-79.5877

Along with urban growth, the process of motorisation began in Beijing at the end of the 1990s. Between 1999 and 2009, the number of vehicles registered in the city increased rapidly at an annual rate of 17.2%, which was to a large extent contributed by the increase of private cars (P. Zhao & Lu, 2011). During the same period, the total length of roads in Beijing increased from 2,441 to 6,248 kilometres. These trends inevitably changed people's travel behaviour: the share of driving increased from 5% in 1986 to 32.6% in 2012, and the share of cycling decreased from 62.7% to 13.9% (Beijing Transportation Research Center, 2013). Increasing motorised travel has become a key issue of concern for the sustainable urban development in Beijing: gasoline consumption increased from 26 to 120 toe (the tonne of oil equivalent) per 1000 inhabitants between 1995 and 2004 and road transport accounted for the major share of incremental energy consumption and CO2 emissions (P. Zhao & Lu, 2011). At the same time, urban growth drove more people out of the city centre to former suburban areas, which induced longer commute distances and triggered brisk demand for transportation (Z. Wang, Deng, & Wong, 2016). These conditions will provide a quite different urban context for the analysis of built environment-travel relationship. Besides, for the case itself, it may gain more from such research than the highly urbanised cities since new urban structures, forms and designs are quickly emerging, potentially influencing travel patterns for decades to come (Zegras, 2010).

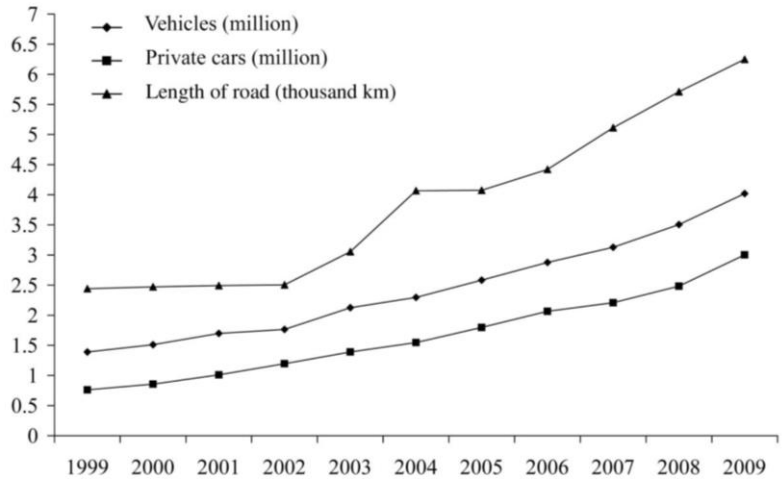

Figure 1-2 Trends of motorisation in Beijing

Source: P. Zhao & Lu, 2011

1.4 General methodology

The over-arching methodology in this thesis is activity-based travel modelling. It simulates activity-travel related decisions such as which activities are conducted when, where, for how long, with whom, and the transport mode involved (Arentze & Timmermans, 2004; Castiglione et al., 2015; Ma, Arentze, & Timmermans, 2012). The strength of the model developed in this research (named as Built Environment Activity-Travel Integrated Model, BEATIM) lies in the comprehensive incorporation of the built environment conditions in the decision making process. This modelling approach can help address the first and the second gaps in Section 1.2 by linking the activity-based modelling with the built environment-travel analysis and thus enabling a more behavior-oriented and decomposed analysis of the built environment's influence.

Activity-based models typically fall into one of two paradigms: utility-maximising econometric models and computational process models, though this categorisation is neither exclusive nor exhaustive (Bhat, Guo, Srinivasan, & Sivakumar, 2004; Pinjari & Bhat, 2011; Rasouli & Timmermans, 2014a; Yasmin et al., 2015). Some authors also mentioned a type of constraints-based models, which puts more emphasis on checking whether any given activity agenda is feasible in a specific space–time context (Rasouli & Timmermans, 2014a). The utility-maximising models and the computational process models bear different strengths. The former are more advantageous for the examination of alternative hypotheses regarding the causal relationships between activity-travel patterns, the built environment and socioeconomic characteristics of individuals (Bhat et al., 2004; Yasmin et al., 2015), while the latter are better at modelling decision making under incomplete information and imperfect rationality, and the learning process (Arentze & Timmermans, 2004; Auld & Mohammadian, 2012). The BEATIM model developed in this research generally takes the utility-maximising paradigm for its strength in examining the influences of the built environment, and also incorporates weak computational process features reflected in a series of action rules (see Chapter 5 for detailed description of the model).

Besides, in order to address the third gap, the machine learning method is employed to evaluate the street facade. Usually, this type of built environment features cannot be directly measured from various readily-available geodatabases, but by human field auditors through manual observation and recording (Brownson, Hoehner, Day, Forsyth, & Sallis, 2009). However, the manual nature makes this method inherently expensive and derives few economy of scale (Harvey, 2014). The machine learning method proposed in this research leverages state-of-the-art techniques and data sources (i.e. online street view images) to realise the automatic evaluation of this type of built environment features (see Chapter 4 for detailed descriptions).

Moreover, several sub-methods are also employed in this research, which include:

- **Travel diary survey** (implemented by the municipal government), on ca. 116,000 individuals, to collect information on people's 24-h activity-travel behaviour;
- **Questionnaire survey** (implemented by myself), on 200 individuals, to collect information on people's travel decision making process;
- **GIS-based spatial analysis**, on various sources of spatial data, to measure the two-dimensional, land use-related 'D' features of the built environment;
- **Discrete choice regressions**, to estimate the weights of built environment features on various choice facets of activity-travel (activity participation and organisation, location choice for primary destinations and intermediate stops, time of activity and mode choice);
- **Scenario analysis**, to simulate the impacts of built environment changes on travel behaviour with the proposed activity-based model.

1.5 Research outputs

The main outputs are on three levels:

Theoretical:

Link to the first and second gaps – The research will provide an understanding on the behavioural processes that give rise to the influences of the built environment on daily travel. The results will be examined against the assumptions on the relationship between the built environment and travel (dis)utilities.

Link to the third gap – The research will also provide an understanding on the impacts of the street facade design on travel behaviour.

Link to the fourth gap – The results from the case study of Beijing, a high density, fast growing Asian city, will be compared with the meta-analysis of the results from American and European cities. Theoretical reflections will be made based on the comparison.

Methodological:

Link to the first and second gaps – The research develops an activity-travel model that comprehensively incorporates the influences of the built environment on various aspects of daily activity-travel.

Policy-related:

Planning policy suggestions for various transport-related goals can be made from the simulation results.

Besides, there are two minor outputs. First, an updated meta-analysis of the effect sizes of built environment features in existing research is provided (see Section 2.3). Second, a machine learning-based method for the automatic evaluation of the street facade design is proposed (see Chapter 4).

1.6 Thesis organisation and structure

After this introduction, Chapter 2 provides a review of the theoretical base and the empirical findings on the built environment-travel relationship, as well as the progresses in activity-based modelling. The theoretical review starts from reviewing the urban planning and design philosophies that advocates the use of the built environment to influence and modify the travel behaviour. The microeconomic theory of utility maximisation is employed to explain the mechanism of this influence, based on which a series of assumptions on the relationship between various built environment features and travel (dis)utilities are proposed. It is followed by a general review of empirical findings on this topic, and a more quantitative meta-analysis of the effect sizes. Last, the progresses in activity-based modelling are reviewed, with an emphasis on the treatment of built environment features in the existing model systems.

For this chapter, these detailed research questions will be answered:

- What are the theoretical bases for the impact of the built environment on travel behaviour?
- What have existing empirical studies found about the impact of the built environment on travel behaviour? What are the significance levels and magnitudes of the impacts reported by existing studies?
- How are activity-based models constructed?

Chapter 3 introduces the study area, the data sources and the measurement of the built environment features conventionally included in this line of research (the six 'Ds'). Specifically, the data sources involve two field surveys for the collection of travel-related information, one large travel diary survey on people's 24h travel behaviour and one small questionnaire survey on the processes of travel decision making.

For this chapter, these detailed research questions will be answered:

- What is the spatial extent of the study area
- How are the data for my research collected?
- How are the '6D' features measured from various data sources?

Chapter 4 deals with the measurement of the street facade design, which employs the state-of-the-art machine learning techniques. Two specific features are selected as key factors that could potentially influence the travel behaviour: the construction and maintenance quality of building facade (a building-level feature) and the continuity of street wall (a street-level feature).

For this chapter, these detailed research questions will be answered:

- What features of street facade design can be considered as key factors that could potentially influence the travel behaviour?
- How can street facade features be measured? How is the performance of machine learning algorithms?

Chapter 5 develops the activity-based model, which simulates a sequence of decision making related to daily activity participation and travel, and more importantly, the impacts of the built environment within the process. The design of the model framework, the construction of the four sub-models and the validation results will be discussed in detail.

For this chapter, these detailed research questions will be answered:

- How can an activity-based model be designed to effectively simulate the decision making process of daily travel and incorporate the impacts of the built environment?
- How can the model parameters be estimated?
- How well can the model perform to approximate the reality?
- In which situations does the model perform well and in which situations does it produce relatively large errors?

Chapter 6 applies the model to simulate the impacts of various built environment changes on travel behaviour through scenario analysis. Two types of scenarios are designed: ‘local’ scenarios, referring to a built environment change in one single TAZ, and ‘regional’ scenarios, referring to the built environment changes in varying buffer zones from a TAZ. As emphasised before, the analysis not only examines the synthesised travel outcomes, but also the behavioural processes. A few policy suggestions are drawn from the simulation results.

For this chapter, these detailed research questions will be answered:

- What are the impacts of various ‘local’ built environment changes on the activity participation and travel behaviour of residents at where the changes take place (e.g. activity frequency, distance of travel, mode choice, etc.)? Specifically, what are the impacts of the two newly-added street facade features?
- What are the impacts of ‘regional’ built environment changes?
- How are the results consistent with or different from those from American and European cities?
- How are the results consistent with or different from theoretical assumptions? What are the implications?
- What are the proper policies for various goals, such as reducing total car use or reducing the travel distance needed for fulfilling daily needs?

In the end, Chapter 7 concludes and discusses the limitations of this research, and points to future directions of research.

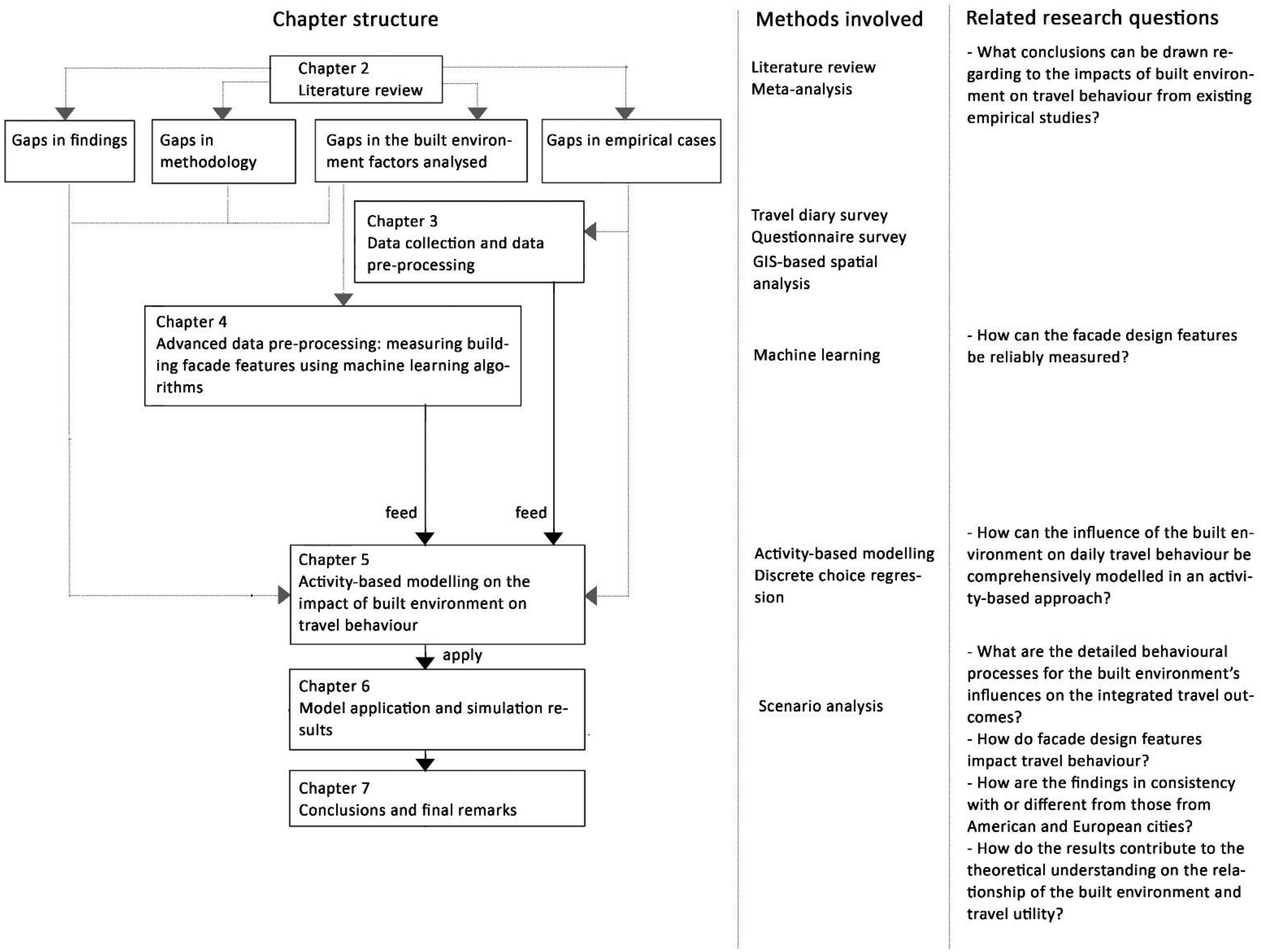

Figure 1-3 Thesis organisation and structure

Chapter 02 Literature review

2.1 Theories and assumptions

2.1.1 Philosophies of urban planning and design

It was observed in many cities, especially in the North America, that the increase in car travel has occurred hand in hand with urban sprawl. Consequently, planners plausibly assume that a reversal of this relation by compact urbanisation, densification, and mixed-use development, will reduce the need to travel—particular by car (Maat, Wee, & Stead, 2005). As a result, the idea of using the built environment to influence and modify travel behaviour has found their way into and become one of the key concerns of diverse urban planning and design concepts ever since the early 20th century (Maat et al., 2005; Zegras, 2010). Besides, in the field of transport engineering, the original 'predictive' focus of land development-transport analysis evolved to include an increasingly 'prescriptive' purpose—i.e. modifying land development patterns explicitly to influence travel demand (Boarnet & Crane, 2001; Zegras, 2010). The key ideas of some of the most influential planning and design strategies that aim at modifying the travel behaviour are briefly reviewed below.

- **Jobs-housing balance:** advocates that promoting the spatial matches between housing and jobs could help counter the trend of widening separation of suburban workplaces and the residences of suburban workers and increasing peak-period traffic congestion in the US (Cervero, 1989, 1996).
- **New Urbanist (or neo-traditional) design:** calls for a return to compact neighbourhoods with a combination of neighbourhood design elements including grid-like street patterns, mixed land uses and pedestrian amenities (Lund, 2003; Talen, 2013). It is believed that the neighbourhoods designed in such a manner are

less oriented toward automobile travel and more conducive to walking, bicycling and transit riding, especially for non-commute trips (Cervero & Radisch, 1996; Joh et al., 2012).

- **Smart growth:** aims to channel new development into existing urban areas and away from undeveloped areas and to improve the viability of alternatives to the car (Handy, Cao, & Mokhtarian, 2005), which are expected to counteract many of the negative effects associated with urban sprawl, including long vehicle travel and congestion (Tracy, Su, Sadek, & Wang, 2011). According to the American Planning Association, compact, transit accessible, pedestrian-oriented, mixed use development patterns and land reuse epitomise the application of the principles of smart growth (American Planning Association, 2002). Actually, there are many overlaps and common technical features between new urbanism and smart growth, as well as other related philosophies such as the walkable city, the compact city and so on (Lund, 2003).
- **Compact city:** the key idea is to bring activities closer to residents so that they can fulfil their needs and, because the distances are smaller, this allows slower modes (walking and cycling) and public transport to play a bigger role in their travel alternatives and reduces energy consumption and pollution (Aditjandra, Mulley, & Nelson, 2013; Neuman, 2005).
- **Walkable city:** advocates for better quality of the walking environment in transport planning and design to promote movement by foot and bicycle (Southworth, 2005). It also involves some of the planning and design doctrines mentioned above such as high density, mixed-use, pedestrian-oriented design (further includes high connectivity, safety, high quality of path, etc.), which are supposed to be conducive to walking (Sarkar et al., 2015; Southworth, 2005). Walkability can also be further linked with issues of public health, based on the notion that higher level of physical activity could lead to lower obesity and fewer weight-related chronic conditions, and better overall health (Doyle, KellySchwartz, Schlossberg, & Stockard, 2006).
- **Transit-oriented development:** seeks to maximise access to mass transit and non-

motorised transportation with centrally located rail or bus stations surrounded by relatively high-density commercial and residential developments (Dittmar & Ohland, 2004), which is believed to enhance the attractiveness of public transport services as a whole over the car (Kamruzzaman, Baker, Washington, & Turrell, 2013).

These ideas appear to have made a great impact on modern urban planning and design, both academically and practically (Crane, 1995). Many of the planning and design doctrines have been gradually integrated into the curriculum at the top planning and architecture schools and also incorporated into numerous development plans and projects (Knaap & Talen, 2005). The large influence has naturally given rise to the need for stringent empirical research to testify the promised benefits. As a result, more than two hundred studies have been devoted to analysing the influence of the built environment on travel, particularly the doctrines advocated by the strategies mentioned above. The following sections will provide a review of the relevant research. Section 2.1.2 and 2.1.3 will focus on the theoretical foundation. Section 2.2 to 2.3 will focus on empirical findings.

2.1.2 The built environment, causality and travel

Despite of a large number of studies on examining the assumed influence of the built environment on travel, Naess (2015) critised that “theories explaining why correlations exist between built environment characteristics and travel behaviour are rarely exposed, let alone reflected on”. According to Naess, “compared to the efforts spent on applying the statistical analyses in an as impeccable way as possible, the literature usually spends considerably less space on discussing which variables to include in the statistical models and their order in a causal chain”.

The establishment of a causal relationship is a long standing issue of social research, since many of them seek to determine what causes what in the complex open system of human society. In fact, the notion of causality itself makes a profound philosophical issue and there is still little agreement on the nature of causation (Beebee, Hitchcock, & Menzies, 2009, p. 1). For instance, some authors think that causation is a relatively non-fundamental feature of the world and can be understood in terms of other more fundamental features such as 'regularities'; some think that in some sense causation is not a feature of reality at all; others think that causation is about as fundamental as it gets (Beebee et al., 2009, pp. 1-2).

In this thesis, I am not going to dig into the complexity of the philosophical debates on causality, but rather take a more practical framework for the understanding of causation in social research, which is used in the methodological guidebook by Schutt (2011) and Singleton Jr, Straits et al. (1993). According to the framework, there are basically two types of causal explanation in social research, which find their roots in the two intellectual tendencies to knowledge coined by Wilhelm Windelband, termed as *nomothetic* and *idiographic* (Schutt, 2011, p. 182). A nomothetic causal explanation identifies common influences on a number of cases or events, which is usually associated with quantitative methods (Schutt, 2011, p. 182). A causal effect from the nomothetic perspective refers to that the variation in one phenomenon leads to or results, on average, in the variation in another phenomenon (Schutt, 2011, p. 182). In contrast, the idiographic causal explanation is more about concrete, individual sequence of events, thoughts, or actions, which can be classified as narrative reasoning and more commonly associated with qualitative methods (Richardson, 1990; Schutt, 2011, p. 184). In the field of built environment-travel research, most existing research are conducted in a quantitative manner and can be considered as reflecting the nomothetic tendency, except for a few works by Naess and collaborators, who consciously took a highly narrative approach from the philosophical position of critical realism (2001, 2015),.

Schutt (2011, pp. 188-189) put forward five criteria when deciding whether a nomothetic causal connection exists, which are: (1) empirical association, (2) appropriate time order (cause precedes effect), (3) non-spuriousness (a relationship between two variables is not due to variation in a third variable), (4) identification of causal mechanism, (5) identification of causal context. While the first criterion is easy to identify with statistical methods, the second and the third requires more sophisticated strategies of research design and data collection. The fourth and the fifth criteria, though not necessarily involve methodological complexity, are much less discussed in existing research.

To be more specific, the time precedence criterion requires approaches that permit multiple directions of causality and/or involve longitudinal measurements. The non-spuriousness criterion needs explicit inclusion of people's travel attitudes, which is considered to be the major source of spuriousness and the consequence is termed as 'self-selection' in this field of research (Cao, Mokhtarian, & Handy, 2009; Mokhtarian & Cao, 2008) (see Cao et al., 2009 for more details in travel attitude-related spuriousness). However, constrained by the costs of data collection, only a small proportion of existing research have, to some extent, incorporated these approaches and more tightly examined the relationship (see Cao et al., 2009 and Mokhtarian et al., 2008 for a review).

The latter two criteria are more about logical deduction for why and under what circumstances the alleged cause should produce the observed effect (Mokhtarian & Cao, 2008). As mentioned at the beginning of this section, Naess (2015) strongly criticised the considerably less research attention on the causal mechanisms. Such a gap may be explained by a general attitude of researchers in this field that takes the existence of a causal mechanism as granted—as Cao et al. (2009) wrote, that "all the statistical methods used in the studies in this field can rely on the travel price changes suggested by Boarnet and Crane (2001) as a plausible causal mechanism". However, the travel price-based causal mechanism put forward by Boarnet and Crane (2001) is subject to further refinement in many ways. Therefore, the next section will be spent on reviewing and developing the theoretical explanation for the mechanism of the built environment's influence on travel behaviour, which, though insufficient alone for the establishment of a causal relationship, could at least help improve the understanding of quantitative results.

2.1.3 Travel as an outcome of utility maximisation

Basically, the aggregate-level relationships between the built environment and travel emerge from a set of transport rationales at the individual-level (Naess, 2015). As mentioned before, most existing works tend to resort to the economic consumer theory that explains individual behaviour and motivations as a result of utility maximisation (Ben-Akiva & Lerman, 1985; Boarnet & Crane, 2001, p. 66; Cervero & Kockelman, 1997; Domencich & McFadden, 1975; Maat et al., 2005; McFadden, 1974; Zegras, 2010). Although the notion of 'rational man' and utility maximisation is commonly questioned, the fact that humans are not entirely rational decision-makers with complete information does not imply that they do not at all use instrumental rationality (Naess, 2013). This behavioral paradigm has served as the basis for a rich production of models in transportation-related choice analysis, including the mode of travel, destinations to visit, the household residence, etc. (see examples in Ben-Akiva and Lerman, 1985) (Goulias, 2009). Besides, the repetitive nature of daily travel indicates that people may have already searched and compared many alternatives and optimised their choices from day to day, so that the observed behaviour can be considered as an outcome under near-complete information.

In the framework by Boarnet and Crane (2001), travellers perform trade-offs among available alternatives as in any situation where decisions are made concerning the allocation of scarce resources, whether or not they involve actual money (pp. 61-62). The built environment influences trip-making through impacts on (relative) trip costs (Boarnet & Crane, 2001, pp. 61-62). A later work by Maat et al. (2005) extends this framework including the benefit side of travel, i.e. activity realisation, following the notion of transport as a derived demand from activity participation, which has an intellectual link with the activity-based approach of travel analysis (Axhausen & Garling, 1992) and Hägerstrand's time geography (1970). The main contribution of Maat et al.'s framework is twofold. First, it assumes that people do not make separate decisions considering only trips, but that they try to schedule activities in a daily pattern, and therefore, they do not maximise utility for separate travel choices, but optimise their entire activity pattern (Maat et al., 2005). This more integral framework more realistically accounts for the possibility that people may use travel time saved from better accessibility of one activity on participating more other activities (Maat et al., 2005). Second, following the key concept of time geography that both space and time are scarce resources and constrain daily activity patterns (Axhausen & Garling, 1992; Hägerstraand, 1970), the activity-based approach also takes into account the fact that individuals are not free to choose any alternative but are constrained by the total amount of time in a day (Maat et al., 2005).

However, although these frameworks are helpful in understanding the motivations of travel decision making, they do not involve explicit assumptions on the influence of the built environment on travel choices. For instance, Maat et al. (2005) merely briefly discussed the change of the travel utility curve in a condition that is broadly described as 'compact design'. A related problem is that few published papers in this field ever explicitly stated the hypothesis of research in terms of how travel behaviour would be influenced by built environment (e.g. see Næss, 2013, Cervero & Kockelman, 1997 as examples of those that provided hypothesis) (Naess, 2010). It is especially an issue with the activity-based framework. The assumption of 'overall utility maximisation' may lead to more than one mechanism of influence of the causal powers. Some causal mechanisms may amplify each other while others may neutralise or reduce each other's influences (Naess, 2013). Besides, to the best of my knowledge, none of existing frameworks mentions the gains from travel itself, such as health gains from walking and cycling, or ‘psychic’ gains from aesthetically pleasing streetscapes (Boarnet & Crane, 2001; Naess, 2013). Therefore, building upon existing works, I would like to put forward a framework which more comprehensively discusses the gains and costs of travel and makes deductions about the influences of the built environment.

Figure 2-1 shows the subdivision of daily activity-travel behaviour based on the activity-based framework, and how they together contribute to the synthesised travel outcomes such as total travel distance and VMT. First, an individual needs to have a general idea about the amount of activity participation in a given day. Activities can be categorised into commute activities (work and go to school), which are usually quite routine and tend to take place at fixed locations, and non-commute activities (all other activities), which are more flexible both in terms of the frequency and the location. For each of the activities that one chooses to participate, a location, a travel mode and a time of activity need to be further selected. These decisions may not have a clear priority and can be mutual influential. For instance, people may perform trip chaining for higher efficiency, in which case multiple activities are combined into one tour so that the total travel distance could be shorter. Besides, considering that people are constrained by the total time budget, the travel costs of one activity (usually more obligatory ones, such as working) can also affect the participation of other activities (usually more discretionary ones). All these subdivided facets together contribute to the synthesised travel outcomes.

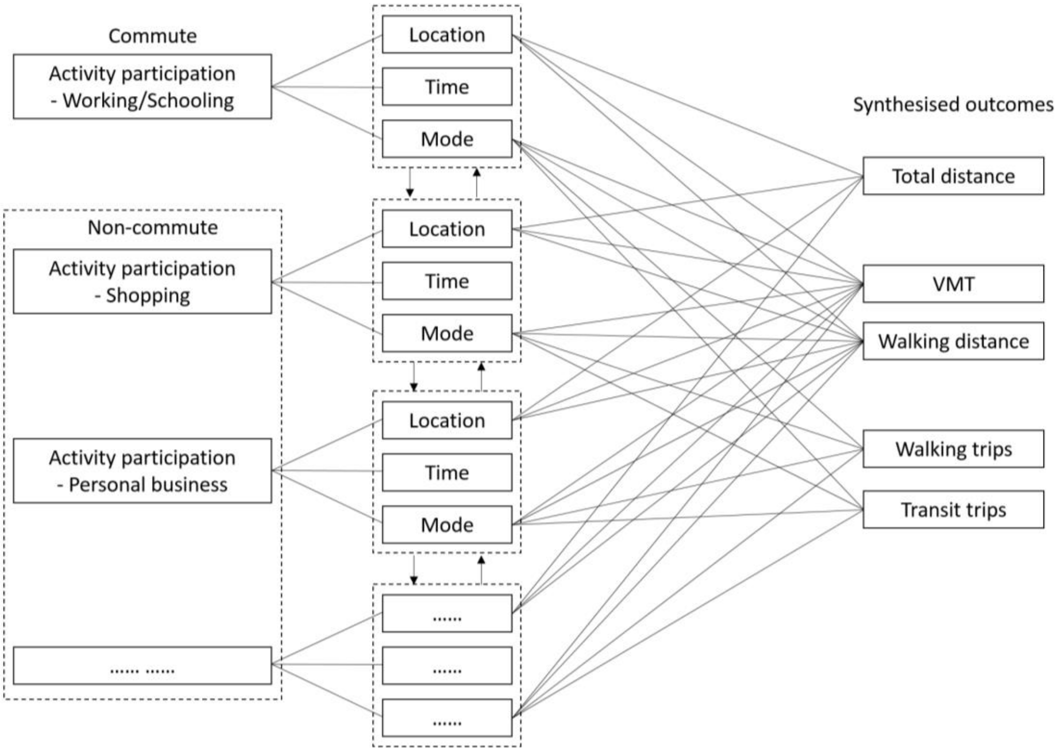

Figure 2-1 Subdivision of daily activity-travel behaviour

Maat et al. (2005) analysed the net utility of travel as a function of travel time since they thought that individuals are not primarily interested in travel distance, but rather in the costs of bridging that distance, particularly the time cost. While totally agree with this notion, I still think that it is more appropriate to use travel distance as the independent variable. First, travel distance is more straightforward and predictable given the built environment conditions of an area, while travel time could vary with the traffic speed. Second, travel time is already mediated by the built environment as a function of the travel distance and the road network or the placement of public transit stations, etc., thus it can be considered as a 'second-order' variable while travel distance is 'first-order'.

The diagram in **Figure 2-2** illustrates the assumed changes in travel gains and costs as functions of travel distance. As mentioned before, two types of travel gains are considered, gains from conducting activities and gains from travel itself. For non-commute activities, the activity gains generally increase with travel distance since the further one travels, the more opportunities are within reach and the bigger the chance of being able to reach an opportunity with a higher utility, e.g. cheaper price or more specialised goods (Boarnet & Crane, 2001; Maat et al., 2005). It is assumed that the increase would slow down as the travel distance gets longer, because the additional benefits of travelling longer can be subject to the law of diminishing returns. For example, the second-nearest supermarket might be more attractive than the nearest, perhaps because of lower prices or more variety in products, but the additional benefits of the fifth-nearest compared with the fourth-nearest might be smaller (Maat et al., 2005). It should be noted that, in reality, due to the heterogeneity of the urban space, the curve may not be smooth and continuous but instead bumpy and discontinuous (see Maat et al., 2005 for an example of net travel utility in mixed and concentrated uses). For commute activities, the relationship between travel gains and travel distance can be very bumpy, since the gains can only be realised at places where there are suitable job positions. The more professional and specialised a job is, the more heterogeneous is the distribution of suitable positions.

The travel gains are shown with dotted line since they may not exist if an individual does not value these benefits or does not take a proper travel mode to realise the benefits. For instance, the health benefits of travel are mainly related to active travel modes (walking and cycling) and slightly related to public transit since the trip to and from bus/metro stations may also involve active travel. However they are hardly related to driving or taking taxi. Moreover, these benefits may not even be considered if an individual does not care about health issues. Similarly, the psychic benefits of travelling in an enjoyable urban environment are more experienced if an individual travels in a slow mode and may not actually exist if an individual does not appreciate such qualities. It is assumed that these travel gains are proportional to the travel distance.

The costs of travel are mainly associated with the time spent, the monetary costs and the physical efforts. The curves are differentiated among different travel modes. For simplicity, three types of modes are considered, which are driving, taking public transit (bus or subway) and active travel (walking or cycling). For driving and taking public transit, the intercepts are above zero since usually some initial actions are required, such as walking to the parking lot, starting the engine, finding a parking space at destination and traveling to and from the subway/bus station, etc. (Maat et al., 2005). Either of the intercepts can be larger than the other, affected by factors such as the distance to public transit stations, the availability of parking spaces, etc. The slope is the highest for active travel since longer time and more physical energy are needed to cover a same distance. Either of the slopes of public transit and driving can be larger than the other, depending on the traffic speed and so on. All curves are supposed to be concave because longer travel could result in growing tiredness, which is the strongest for active travel, medium for taking public transit and the mildest for driving. Another reason for the concave shape is that time is scarce resource so there could be a growing marginal cost. In practice, the shapes of these lines could be more complicatedly affected by the structural conditions of the society, the specific mind set of a person, and the specific conditions of a trip (Naess, 2013). For instance, in a circumstance that drivers are rude and careless, or a person is particularly sensitive to safety risks, the cost slope of active travel could be steeper.

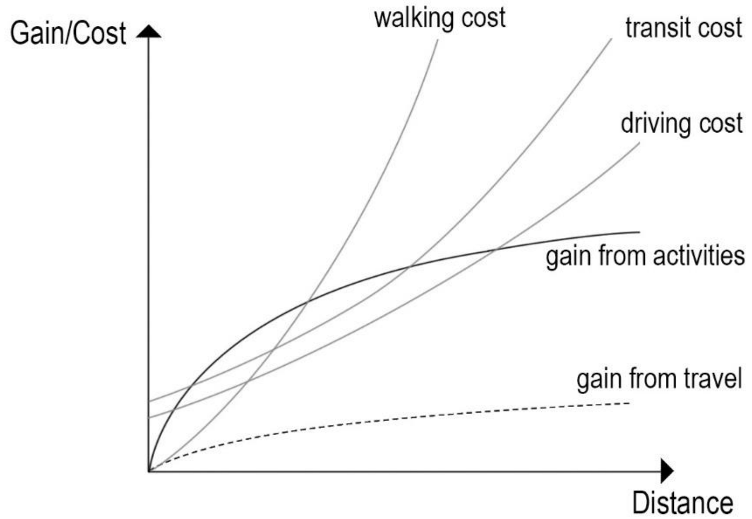

Figure 2-2 Gains and costs of travel in relation to travel distance

The net utility can then be derived by minusing the costs from the gains as shown in **Figure 2-3**. An individual would choose the combination of travel distance and travel mode that produces the highest net utility, if such an alternative is feasible (e.g. not inhibited by car ownership or time budget). It should be noted that, as mentioned before, the shapes of the curves and the positions of the peaks and intercepts may not take the exact form as shown in the figure, but are affected by many factors including the institutional conditions, personal preferences, and more importantly for this research, the built environment settings. In the next part, the impacts of built environment factors on travel gains and costs and the choice outcomes will be discussed.

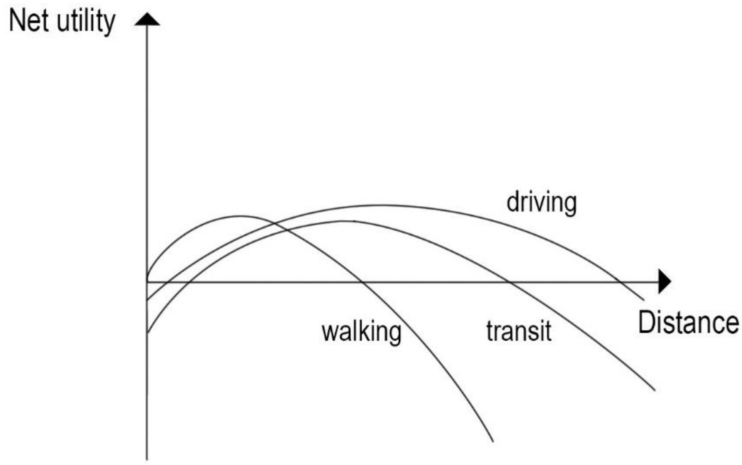

Figure 2-3 Net travel utility in relation to travel distance (an example)

2.1.4 Assumptions on the influence of the built environment

In travel research, the built environment have often been described through features named with words beginning with D. The ‘three Ds’, coined by Cervero and Kockelman (1997), are density, diversity and (road network) design, followed later by destination accessibility and distance to transit (Ewing & Cervero, 2001; Ewing et al., 2015). Parking supply is also sometimes coined as the sixth ‘D’, namely demand management (Ewing & Cervero, 2010). Besides, as mentioned in the introduction, in order to fill in the gap that the influence of street facade features is seldom analysed (see Section 1.2), this research also introduces a seventh ‘D’, street facade design. Although these D-variables are rough categories, divided by ambiguous and unsettled boundaries that may involve overlaps, they are still useful for the description of the built environment (Ewing & Cervero, 2010). Their assumed effects on travel gains and costs are discussed below.

It should be noted that, since people tend to optimise their entire activity pattern, all the effects that result in shorter travel distance and travel time may be compensated by inducing more activities (Maat et al., 2005). However, the compensation effect cannot be explicitly plotted on the diagram but is taken into account in the summary of hypotheses in **Table 2-1**.

(1) Higher density

- On gains of commute activities: employment density may enhance the gains by increasing the chance for an individual to find a suitable job;
- On gains of non-commute activities: can enhance the gains by increasing the amount and, potentially, the quality of goods and services in a given travel distance (residential density can exert this influence by expanding the consumer base and attracting more businesses);
- On gains of travel itself: no obvious effect;
- On costs of driving: may increase the cost by increasing the traffic flow and lowering the traffic speed (steeper slope), as well as the competition for parking space (larger intercept);
- On costs of taking public transit: may reduce the cost if the service provider enhances the level of service in response to the larger consumer base (smaller intercept and gentler slope);
- On costs of active travel: may reduce the cost if the increased traffic flow lowers down the traffic speed and thus improves the road safety (gentler slope).

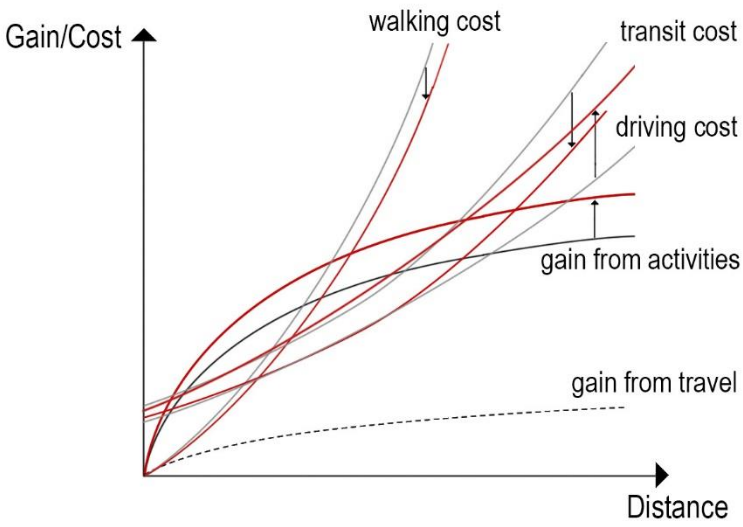

Figure 2-4 Assumed changes in travel gains and costs following enhanced density, diversity and destination accessibility

(2) Higher diversity

- On gains of commute activities: may enhance the gains by providing more office space and increasing the chance for an individual to find a suitable job;
- On gains of non-commute activities: can increase the gains by providing a larger variety of activity opportunities in a given distance (Feng et al., 2013);
- On gains of travel itself: no obvious effect;
- On costs: similar to the effects of higher density, by attracting more consumers and inducing more traffic.

(3) Higher destination accessibility (to city/sub-centres)

- On gains of commute activities: may enhance the gains by increasing the number of jobs available within a given distance, especially professional and specialised jobs that tend to agglomerate at central business areas of the city;
- On gains of non-commute activities: can enhance the gains when the travel distance is long enough to reach the city/sub-centres, where there is usually a concentration of facilities and services;
- On gains of travel itself: no obvious effect;

- On costs: similar to the effects of higher density and diversity, if a place is close enough to the city/sub-centres to be affected by the traffic flow induced by the activities at the centres; no obvious effect if not.

(4) Higher road density/connectivity

- On gains of commute activities: no obvious effect;
- On gains of non-commute activities: no obvious effect;
- On gains of travel itself: no obvious effect;
- On costs of driving: can reduce the cost by providing more direct routes and increasing the speed by diverting traffic flow (gentler slope);
- On costs of taking public transit: may reduce the cost by providing more direct routes to and from bus/subway stations (smaller intercept). However, if the increase of road density is mainly associated with high-level roads such as express ways or primary roads which are less pedestrian/cyclist-friendly, the effect could be inverse (larger intercept);
- On costs of active travel: may reduce the cost by providing more direct routes (gentler slope) and also could be inverse for the same reason explained above (steeper slope).

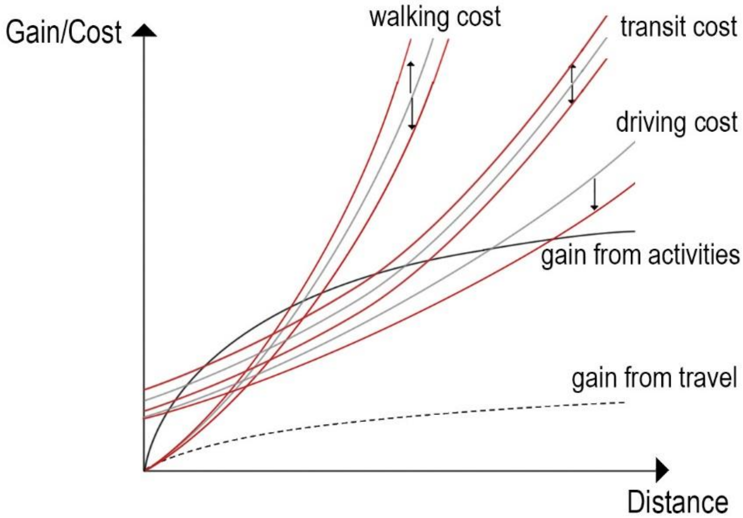

Figure 2-5 Assumed changes in travel gains and costs following enhanced road

density/connectivity

(5) Shorter distance to transit

- On gains of commute activities: no obvious effect;
- On gains of non-commute activities: no obvious effect;
- On gains of travel itself: no obvious effect;
- On costs of driving: no obvious effect;
- On costs of taking public transit: can reduce the cost by decreasing the 'fixed cost' of travelling to the stations (smaller intercept);
- On costs of active travel: no obvious effect.

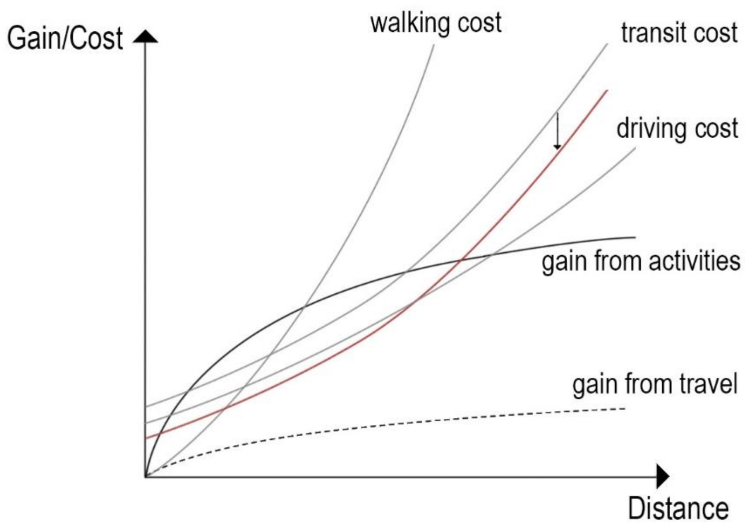

Figure 2-6 Assumed changes in travel gains and costs following reduced distance to transit

(6) More parking provision

- On gains of commute activities: no obvious effect;
- On gains of non-commute activities: no obvious effect;
- On gains of travel itself: no obvious effect;
- On costs of driving: can reduce the cost by cutting the 'fixed cost' of searching for parking space (smaller intercept);

- On costs of taking public transit: no obvious effect;
- On costs of active travel: no obvious effect.

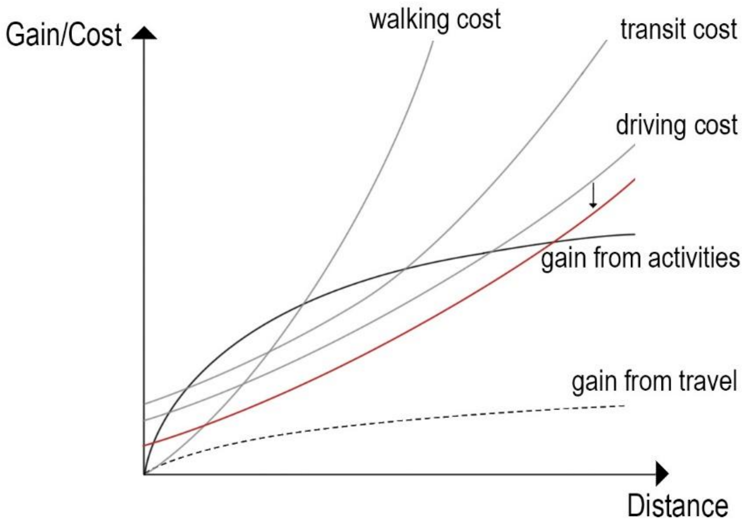

Figure 2-7 Assumed changes in travel gains and costs following increased parking space

(7) Better street facade design

- On gains of commute activities: no obvious effect;
- On gains of non-commute activities: no obvious effect;
- On gains of travel itself: can increase the gains by enhancing psychic enjoyment, especially for slow modes (walking/cycling) and trips that involve slow modes (e.g. walking/cycling to transit stations);
- On costs: no obvious effect.

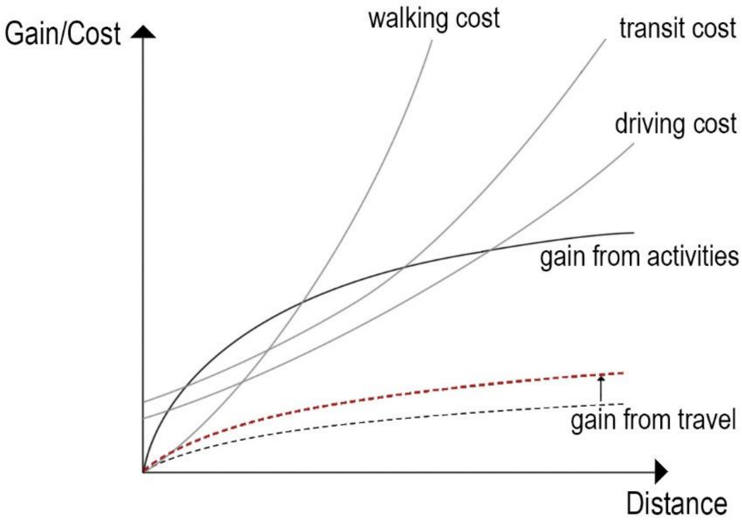

Figure 2-8 Assumed changes in travel gains and costs following enhanced street facade design

However, these gains and costs can hardly be directly observed and measured. Instead, the direct observation is people's travel behaviour as an optimised choice out of the gains and costs. Therefore, the effects of these built environment changes on the behavioural outcomes are further deduced based on their expected effects on travel gains and costs (see **Table 2-1**). The behavioural outcomes are represented with two main indicators of travel behaviour: total travel distance and VMT. These assumptions can be directly examined against the simulation results later in the thesis.

Table 2-1 Assumed effects of built environment changes on travel behaviour

Built environment changes	**Assumed effects on total travel distance**	**Assumed effects on VMT**
Higher density	For one trip: reduce, by increasing the gains of non-commute activities (and possibly also commute activities) in a given distance For total travel in a day: ambiguous, since shorter travel distance may induce more activities	For one trip: reduce, by reducing the travel distance and potentially discouraging driving For total travel in a day: ambiguous
Higher diversity	For one trip: reduce, by increasing the gains of non-	For one trip: reduce, by reducing the travel distance

Built environment changes	Assumed effects on total travel distance	Assumed effects on VMT
	commute activities (and possibly also commute activities) in a given distance For total travel in a day: ambiguous, since shorter travel distance may induce more activities	and potentially discouraging driving For total travel in a day: ambiguous
Higher destination accessibility	For one trip: reduce when the travel distance exceeds a threshold to reach the centres, by increasing the gains of non-commute and commute activities For total travel in a day: ambiguous, since shorter travel distance may induce more activities	For one trip: reduce, by reducing the travel distance and potentially discouraging driving For total travel in a day: ambiguous
Higher road density/connectivity	For one trip: likely to increase by reducing the travel costs For total travel in a day: likely to increase	For one trip: ambiguous, since both the costs of motorised and non-motorised travel are reduced For total travel in a day: ambiguous
Shorter distance to transit	For one trip: slightly likely to increase, if the optimised choice is a longer trip with public transit For total travel in a day: slightly likely to increase	For one trip: reduce, by encouraging transit use for both commute and non-commute activities For total travel in a day: reduce
More parking provision	For one trip: slightly likely to increase, if the optimised choice is a longer trip by car For total travel in a day: slightly likely to increase	For one trip: increase, by encouraging driving for both commute and non-commute activities For total travel in a day: increase
Better street facade design	For one trip: slightly likely to reduce, if the optimised choice involves a nearer destination For total travel in a day: slightly likely to reduce	For one trip: likely to reduce, by increasing the gains of active travel and probably also public transit for both commute and non-commute activities For total travel in a day: likely

Built environment changes	Assumed effects on total travel distance	Assumed effects on VMT
		to reduce

2.2 Summary of empirical research

2.2.1 Method of literature search

As mentioned before, large efforts have been made on applying statistical methods and evaluating the significance and magnitude of the built environment's influence on travel behaviour. Therefore, as a first step to probe into this research topic, it is necessary to obtain an overview of the findings from the existing work.

Two rounds of literature search were conducted, first in October to December 2013 and then in April 2016 for an update. The search was conducted through the Web of Knowledge and the Google Scholar using 'travel', 'travel behaviour', 'transport', 'trip', 'activity', 'nonwork travel', 'non-commute travel', 'built environment', 'urban form' and 'land use' as the keywords. I also used the 'snowballing' approach and tracked the references of seminal works, as well as the papers that cited those seminal works. Considering the rapid expansion of the studies on this topic, an emphasis was given to the studies published within five years from when the search was did. The first search resulted in approximately two hundred studies. The second search found another seventy studies published between 2013 and 2016. Around forty studies were removed for various reasons, e.g. they are pure qualitative studies, or they deal with very specific topics, such as school travel (e.g. McMillan, 2007; Yarlagadda & Srinivasan, 2007) , mode switching behaviour (e.g. Wang & Chen, 2012) , travel of homemakers (e.g. C. Chen & McKnight, 2007) . The findings of these approximately two hundred studies are briefly summarised below. For a more concrete and quantitative synthesis of the findings, please refer to the meta-analysis in Section 2.3.

2.2.2 A brief overview of findings

First of all, most of the large number of studies on this topic employ statistical methods and regress between synthesised measurements of travel behaviour (e.g. VMT, total walking distance) and built environment and socioeconomic conditions. This corresponds to the first gap mentioned in Section 1.2 that the behavioural processes that give rise to these synthesised outcomes have received much less attention.

Besides, questions remain about the quantifiable influence of the built environment on travel behaviour, since large divergence exists in the findings from existing research (Zegras, 2010). For instance, some studies found small to no effect of density on VMT (e.g. Ewing et al., 2009; Salon, 2015) , whereas some reported more prominent effects (e.g. Heres-Del-Valle and Niemeier, 2011, reported elasticities between 0.14 and 0.19; Guerra, 2014, reported elasticities between 0.21 and 0.31) . Causes for the divergence in findings could include the method of data collection, the measurements of the built environment features, the analytical approaches and control variables employed, the effects of the Modifiable Areal Unit Problem and so on (Zegras, 2010). Besides, contextual factors may also play an important role, which could activate or inactivate, and influence the relative strength of the multi-mechanisms of the interaction between the built environment and travel.

Nonetheless, the findings still share some common features. First, the effect sizes are generally small (Ewing et al., 2015; P. Zhao, 2011). According to a meta-analysis by Ewing and Cervero (2010), the elasticities between built environment variables and travel outcomes are mostly smaller than 0.5, and many are smaller than 0.2. However, it does not imply that built environment measures are irrelevant or ineffective in modifying the travel behaviour, since multiple measures can be combined and produce larger effects (Ewing & Cervero, 2010). Besides, although the influence may seem small at the individual level, the increments accumulated over time and across populations could potentially produce large benefits (McCormack et al., 2012).

Another common feature is the small R-squared values of regressions (when applicable). For instance, vehicle use models, estimated on data for cities in the U.S. have displayed R-squared values in the range of 0.04 to 0.17 (Zegras, 2010). In the seminal work by Cervero and Kockelman (1997), the models on VMT have R-squared values in the range of 0.17 to 0.20 and only 0.03 to 0.05 is contributed by built environment variables. However, for this research topic, such R-squared values can be considered to be fairly good explanatory power, particularly considering the disaggregate nature of the data and the fact that only a single day's travel behaviour is predicted (Zegras, 2010).

Besides, the findings share some similarities in terms of the significance and the direction of the influences. Generally speaking, built environment features that are related to higher accessibility to activity opportunities and a more friendly environment for active travel and transit use usually induce shorter travel distance and less car use. For instance, the features that are frequently found to significantly reduce VMT include: higher accessibility (Ewing & Cervero, 2010; Krizek, 2003; Nasri & Zhang, 2015), higher density (Ewing et al., 2015; Guerra, 2014; Nasri & Zhang, 2012), higher mix of uses (Cervero & Duncan, 2006; Kockelman, 1997; Nasri & Zhang, 2015), higher street connectivity (Chatman, 2008; Salon, 2015), better access to public transit (Kamruzzaman et al., 2013; Jie Lin & Long, 2008; Susilo, Williams, Lindsay, & Dair, 2012) and so on. It should be noted that these relationships may not hold in every individual study and even reverse results may occur. For instance, Kamruzzaman et al. (2013) found no independent effect of land use diversity on travel and Knuiman et al. (2014) found that land use mix could even discourage walking .

It remains controversial to infer causality based on the research design and statistical method of a large proportion of existing research, since the criteria of 'non-spuriousness' (see Section 2.1.2) is usually confronted with the confounding factor of 'residential self-selection’ (Mokhtarian & Cao, 2008). Self-selection refers to that people may choose residential locations that are consistent with their travel preference (attitude) or capability (e.g. the financial capability of affording a car or physical capability of driving a car). In this situation, it may not be the built environment but the preference or capability that causes the observed behaviours (Cao, 2015a). If the self-selection effect exists but is not controlled for through research design and/or econometric models, we may misestimate the impact of the built environment on travel behaviour, leading to erroneous policy implications (Cao, 2015a).

However, this caveat may not be as important as it sounds to be (Ewing et al., 2015). A review on 38 studies that controlled for self-selection showed that nearly all of these studies still found statistically significant associations between the built environment and travel, independent of self-selection influences (Cao et al., 2009). Later individual studies also had similar findings (e.g. Giles-Corti et al., 2013; McCormack et al., 2012). Nonetheless, self-selection does attenuate the effects of the built environment (Cao et al., 2009), but the latter still plays a dominant role (Cao, 2015a; Cao & Fan, 2012; Ewing et al., 2015). For instance, Bhat and Eluru (2009) found that 87% of the VMT difference between households residing in conventional suburban and traditional urban neighbourhoods is due to ‘true’ built environment effects (Bhat & Eluru, 2009); Zhou and Kockelman (2008) found that the built environment accounted for 58% to 90% of the total influence of residential location on VMT (Zhou & Kockelman, 2008); Cao and Fan (2012) found that density and self-selection contributed to 72% and 28% of the observed impact of density on person miles travelled (Cao & Fan, 2012). Although there do exist contradictory findings that preferences and attitudes are more influential than the built environment (Kamruzzaman et al., 2013), these studies are relatively scarce. Moreover, some studies even find that controlling for self-selection is more likely to enhance than diminish built environmental influences (Chatman, 2009; Ewing & Cervero, 2010). These findings indicate that, the lack of control for self-selection, although prohibits causal inferences, does not seem to substantially affect the validity of the conclusions.

2.3 A meta-analysis

2.3.1 Method

In order to tackle the divergence in the findings and clarify the state of knowledge, some researchers have employed the method of meta-analysis and provided synthesised analysis of the effect sizes in existing findings. This method can provide more generalisable and reliable estimates across studies and help make sense of differing results. Among them, Ewing and Cervero's work (2010) is one of the most comprehensive, which involved 62 empirical cases selected from more than 200 works published before 2010 and provided weighted average elasticities between the outcomes of daily travel and different built environment measures.

In this section, I build on the work of Ewing and Cervero (2010) and expand the meta-analysis to include studies published afterwards (January 2010 to January 2016), so that a more comprehensive and up-to-date overview of existing findings can be made. This work follows the statement of Ewing and Cervero that they 'aimed to seed the meta-study of built environments and travel, expecting that others would augment and expand their database over time'. This update can help people examine whether the findings on this topic are stable enough and remain consistent. More importantly, since most of the existing research are on non-Asian cities (most North American or European, few Oceanian), the synthesised results can be used to compare with the simulation results from Beijing later in my research, which could help address the gap of the lack of understanding on the regional differences in the built environment-travel relationship. For this purpose, the few studies on Chinese and other Asian cities (e.g. Korean) are deliberately excluded from this meta-analysis (e.g. Eom & Cho, 2015; Huang et al., 2016; P. Zhao, 2014; P. Zhao, 2015) .

The literature search described in the last section returned around fifty papers published after January 2010. The meta-analysis follows the method used by Ewing and Cervero (2010), which calculates the weighted average elasticities. An elasticity refers to the ratio of the percentage change in one variable associated with the percentage change in another variable, which is the most widely used measure of effect size in economic and planning research (Ewing & Cervero, 2010). The elasticities are obtained either by directly copying them from published articles where they were reported explicitly, or calculating them on myself from regression coefficients and the mean values of dependent and independent variables (see **Table 2-3** for formulas of calculation). More than half of the studies are excluded from the meta-analysis for not being able to calculate the elasticities (see **Table 2-2**). The most common reason is not providing the mean values of either the dependent or the independent variables. Other reasons include using statistical methods, such as structural equation models, from which simple summary effect size measures could not be calculated (Ewing & Cervero, 2010), or using some special measurements of the built environment that are not comparable to other studies (e.g. transform the continuous measure of the built environment into categorical variables).

Table 2-2 Studies published between Jan 2010 and Jan 2016

	Study sites	**Methods**	**Controls**	**Self-selection**	**In meta-analysis? (Reason if no)**
(Zegras, 2010)	Santiago de Chile, U.S.	OLS /MNL	SE/OT	No	Yes
(Boarnet, Forsyth, Day, & Oakes, 2011)	Irvine Minnesota, U.S.	-	-	-	No (does not provide sufficient mean values and uses detailed descriptors of pedestrian environment which are not comparable with other studies)
(Boarnet, Joh, Siembab, Fulton, & Nguyen, 2011)	Los Angeles, U.S.	NBR/PRR	SE/AT	Yes	Yes
(Tracy et al., 2011)	Buffalo, New York, U.S.	-	-	-	No (does not provide sufficient mean values)
(Cao & Fan, 2012)	North Carolina, U.S.	-	-	-	No (density is treated as categorical - high density/low density)
(De Vos, Derudder, Van Acker, & Witlox, 2012)	Flanders, Belgium	-	-	-	No (does not directly analyse the influence of the built environment)
(Joh et al., 2012)	Los Angeles, U.S.	NBR	SE/CR	No	Yes
(McCormack et al., 2012)	Perth, Australia	-	-	-	No (built environment features not comparable to other studies)
(Nasri & Zhang, 2012)	6 metropolitan areas, U.S.	-	-	-	No (does not provide sufficient mean values)

	Study sites	Methods	Controls	Self-selection	In meta-analysis? (Reason if no)
(Salon et al., 2012)	-	-	-	-	No (literature review)
(Susilo et al., 2012)	UK	-	-	-	No (does not provide sufficient mean values)
(T. Wang & Chen, 2012)	Puget Sound, U.S.	-	-	-	No (uses structural equation models)
(Witten et al., 2012)	Christchurch and Wellington and Waitakere and North Shore in New Zealand	HLM	SE/AT/OT	Yes	Yes
(Aditjandra et al., 2013)	Tyne and Wear, North East England	-	-	-	No (does not provide sufficient mean values)
(Freeman et al., 2013)	New York, U.S.	-	-	-	No (uses a combined walkability scale)
(Giles-Corti et al., 2013)	Perth, Australia	-	-	-	No (built environment changes are represented in discrete numbers)
(Gim, 2013)	-	-	-	-	No (literature review)
(Kamruzzaman et al., 2013)	Brisbane, Australia	-	-	-	No (does not provide sufficient mean values)
(Millward, Spinney, & Scott, 2013)	Halifax, Canada	-	-	-	No (does not include built environment variables)
(Scheiner & Holz-Rau, 2013)	Cologne, Germany	-	-	-	No (uses structural equation models)

	Study sites	Methods	Controls	Self-selection	In meta-analysis? (Reason if no)
(Song, Preston, & Brand, 2013)	Cardiff, Kenilworth and Southampton, UK	FLG	SE	No	Yes
(Ellder, 2014)	Sweden	-	-	-	No (uses multilevel regression)
(Guerra, 2014)	Mexico City, Mexico	TOR/LGR/OLS	SE	No	Yes
(Hong, Shen, & Zhang, 2014)	Puget Sound, U.S.	-	-	-	No (does not provide sufficient mean values)
(Knuiman et al., 2014)	Perth, Australia	-	-	-	No (uses 'whether conduct any transport walking' as the dependent variable, which is not comparable to other studies)
(Lee, Nam, & Lee, 2014)	Houston-Galveston, U.S.	MNL	SE	No	Yes
(Cao, 2015a)	Twin city, U.S.	OLR	SE/AT	Yes	Yes
(Cao, 2015b)	Twin city, U.S.	OLR	SE/AT	Yes	No (uses 'urban' and 'suburban' to indicate the built environment)
(Cho & Rodriguez, 2015a)	Twin cities and Montegory County, U.S.	-	-	-	No (uses categorical variables to indicate the built environment)
(Cho & Rodriguez, 2015b)	Washington, U.S.	-	-	-	No (uses categorical variable to indicate the built environment)
(De Vos, 2015)	Flanders and Netherlands	-	-	-	No (descriptive analysis)
(Eom & Cho, 2015)	Seoul, Korea	-	-	-	No (Asian city)
(Ewing et al., 2015)	15 regions, U.S.	HLM/LGR/OLS/NBR	SE	No	Yes

	Study sites	Methods	Controls	Self-selection	In meta-analysis? (Reason if no)
(Jahanshahi, Jin, & Williams, 2015)	UK	-	-	-	No (uses structural equation models)
(Kim & Wang, 2015)	Hamilton County, Ohio, U.S.	-	-	-	No (uses multilevel regression and does not provide sufficient mean values)
(Lamiquiz & Lopez-Dominguez, 2015)	Madrid, Spain	-	-	-	No (does not provide sufficient mean values)
(Manaugh & El-Geneidy, 2015)	Montreal, Quebec, Canada	-	-	-	No (does not include built environment variables)
(Merlin, 2015)	U.S.	-	-	-	No (does not provide sufficient mean values and focuses on nonwork activity participation only)
(Naess, 2015)	-	-	-	-	No (qualitative)
(Nasri & Zhang, 2015)	19 metropolitan areas, U.S.	-	-	-	No (does not provide sufficient mean values)
(Salon, 2015)	California, U.S.	TOR/ OLS /MNL	SE	Yes	Yes
(Sarkar et al., 2015)	London, UK	-	-	-	No (built environment features are transformed into quartiles)
(L. Yang et al., 2015)	Missouri, U.S.	-	-	-	No (does not provide sufficient mean values)
(Klinger & Lanzendorf, 2016)	Germany	-	-	-	No (does not provide sufficient mean values and the built

	Study sites	Methods	Controls	Self-selection	In meta-analysis? (Reason if no)
					environment features used are not comparable with other studies)

Note: I use the following abbreviation

Method:

FLG = fractional logit model

HLM = hierarchical linear modelling

LGR = logistic regression

MNL = multinomial logit model

NBR = negative binomial regression

OLR = ordered logit regression

OLS = ordinary least squares

PRR = probit regression

TOR = Tobit regression

Controls:

AT = attitudinal variables

CR = crime variables

OT = other variables

SE = socioeconomic variables

It should be noted that this analysis also runs the common risk of meta-analysis that 'put orange and apple in the same basket', which refers to the problem that dissimilar studies and variables are combined (Martinussen & Kroger, 2013). For instance, following the practice of Ewing and Cervero (2010), results on walk mode choice and walk trips per person are mixed, as well as results with various measurements of land use diversity (e.g. job-housing balance, entropy index, distance to nearest store, etc.). However, such mix can hardly be avoided, to acquire a reasonable minimum sample size. In order to fully inform the readers of potential bias, the elasticities from individual studies are also presented, with descriptions of the dependent and independent variables (see Appendix A).

Table 2-3 Elasticity estimation formulas

Regression specification	**Elasticity**
Linear	$\beta * \frac{\overline{x}}{\overline{y}}$
Log-log	β
Log-linear	$\beta * \overline{x}$
Linear-log	$\frac{\beta}{\overline{y}}$
Logistic[a]	$\beta * \overline{x}(1 - \left(\frac{\overline{y}}{n}\right))$
Poisson	$\beta * \overline{x}$
Negative Binomial	$\beta * \overline{x}$
Tobit[b]	$\beta * \frac{\overline{x}}{\overline{y}}$

Reference: (Ewing & Cervero, 2010)
Note: β is the correlation coefficient on the built-environment variable of interest, $\overline{y}$ the mean value of the travel variable of interest, and $\overline{x}$ the mean value of the built environment variable of interest.

a $\left(\frac{\overline{y}}{n}\right)$ is the mean estimated probability of occurrence.

b Applied only to positive values of the Tobit distribution.

Following the practice of Ewing and Cervero (2010), weighted average elasticities are calculated from individual studies using sample size as the weighting factor. The elasticities are estimated between travel outcomes and built environment features that are frequently analysed, so that a sample of at least three studies can be available. For each dependent/independent variable pair, two elasticities are compared: one from studies published before 2010 (directly obtained from Ewing & Cervero, 2010) and one updated analysis plus studies published between 2010 and 2016.

2.3.2 Results

Table 2-4 shows the results of elasticity estimation, which are calculated from the individual results presented in Appendix A. The weighted average elasticities on VMT turn out to be relatively stable when later studies are added to the estimates. The elasticities of five of the ten built environment factors show differences smaller than 0.02 between the updated estimates and the original ones produced by Ewing and Cervero (2010). The largest differences are 0.05, which are the elasticities on the job accessibility by auto and the percentage of four-way intersections.

The results turn out to be less stable on walking and transit use. The updated estimates can be more than two times of the original estimates or even with an opposite sign. Part of the reason could be that the criteria of consistency in the measurement of dependent and independent variables is relaxed when dealing with walking and transit use, since otherwise there cannot be enough studies to support the analysis. For instance, only studies using total VMT as the dependent variable are included in the analysis on VMT, while those using commute or non-commute VMT are excluded. However, for walking and transit use, studies on commute or non-commute trips are mixed with those on all types of trips (see the Appendix of Ewing and Cervero, 2010). Moreover, the instability of the influence itself could also be part of the reason. Substantial differences can be observed among individual results, even between those using quite consistent measurements of dependent and independent variables. For instance, the elasticity of population density on walk/bike mode choice for nonwork trips is 0.48 in Lee et al., 2014, in contrast to 0.01 in Rajamani et al., 2003; the elasticity of job density on transit mode choice for work trips is 0.35, in contrast to 0.09 in Zhang, 2004 (see **Table A-6** and **Table A-11** in Appendix A).

In summary, existing findings seem to be more consistent in the effect sizes of the built environment on VMT. Although these results should be only used as ballpark estimates (Ewing & Cervero, 2010), we can more confidently draw the results on VMT to make further implications. I will not discuss in detail about the relative strongness of various built environment features and their applications in the planning practice, since Ewing and Cervero (2010) had already written a lot on these issues. My focus will be to use these results as the 'benchmark', since all the studies included in this meta-analysis are based on non-Asian cities (most North American or European, few Oceanian), for a comparison with my own empirical results on Beijing later in the thesis.

Table 2-4 Summary of elasticities derived from existing studies

	All studies		**Studies before 2010**[a]	
	Number of studies	**Weighted average elasticity**	**Number of studies**	**Weighted average elasticity**
On VMT				
Household/population density	14	-0.04	9	-0.04
Job density	7	-0.03	5[b]	0
Land use mix (entropy index)	14	-0.07	10	-0.09
Jobs-housing balance	5	-0.03	4	-0.02
Street/intersection density	11	-0.09	6	-0.12
%4-way intersections	4	-0.07	3	-0.12
Job accessibility by auto	8	-0.15	5	-0.20
Job accessibility by transit	4	-0.07	3	-0.05
Distance to downtown	5	-0.22	2[c]	-0.22
Distance to nearest transit stop	8	-0.05	5[b]	-0.05
On walking				
Household/population density	18	0.15	10	0.07
Job density	11	0.14	6	0.04
Commercial floor area ratio	-	-	3	0.07
Land use mix (entropy index)	19	0.28	7[d]	0.15
Jobs-housing balance	-	-	4	0.19
Distance to a store	-	-	5	0.25
Street/intersection density	12	0.40	7	0.39
%4-way intersections	8	0.30	5	-0.06
Job within one mile	-	-	3	0.15
Distance to nearest transit stop	4	0.10	3	0.15

On transit use				
Household/population density	13	0.15	10	0.07
Job density	8	0.07	6	0.01
Land use mix (entropy index)	9	0.26	6	0.12
Street/intersection density	5	0.41	4	0.23
%4-way intersections	6	0.48	5	0.29
Distance to nearest transit stop	-	-	3	0.29

a This part is from the work of Ewing and Cervero (2010).
b Was six in the published paper, however I can only five from the material provided by the authors.
c Was three in the published paper, however I can only two from the material provided by the authors.
d Was eight in the published paper, however I can only seven from the material provided by the authors.

2.4 The built environment in activity-based travel models

As mentioned in the introduction, the activity-based modelling approach can be developed into a helpful tool for the analysis of built environment-travel relationship, with special strength in simulating the detailed behavioural processes. Activity-based models have been put forward as a superior alternative to the widely-used four-step models, which was the dominant method in the field of transport modelling (McNally, 2007; Rasouli & Timmermans, 2014a; Yasmin et al., 2015; Yasmin, Morency, & Roorda, 2017). The four-step model is a kind of spatial interaction model that predicts the aggregate trip productions and trip attractions of traffic zones based on the propensity to travel and the travel impedance (time and/or cost), which finds its theoretical roots in social physics (Batty, 2009; McNally, 2007). The main critique of four-step models lies in that it is aggregate in nature and does not involve any behavioural mechanism—the unit of measurement is not an individual, but rather the number of trips emanating from any particular zone (Rasouli & Timmermans, 2014a). As a consequence, the lack of behavioural mechanism also makes four-step models patently inadequate when it comes to accounting for the effects of the built environment on person travel (Ewing et al., 2015).

The major advantage of activity-based model is claimed to be the behavioural realism and the integrity, which allows for a comprehensive prediction of the sequence of activities and the associated travel, where, when, for how long, subject to a set of spatial, temporal and institutional constraints (Acheampong & Silva, 2015; Rasouli & Timmermans, 2014a). This modelling approach provides a means of forecasting the impacts of a given policy at the disaggregate level, so that a wider set of more detailed policies can be tested in ways that are generally infeasible with the conventional four-step approach (Bhat et al., 2004; Goulias, 2002).

Although practical models in this strand started to increase since the 1990s, particularly after 2000, the theoretical underpins can be traced back to the 1970s. Chapin (1968) first put forward the idea to relate human activity systems to the spatial structure of the city as a critique to the disposition to rely wholly on land rent theory and the market mechanism in the study of urban structure and processes. Hägerstrand (1970) introduced the time-space concept which also emphasised the importance of understanding the micro-situation of human activities in studying the large scale aggregate outcomes such as traffic generation.

Activity-based models typically fall into one of two categories: utility-maximising econometric models and computational process models. The former involves using systems of equations to capture the relationships among activity and travel attributes, and to predict the probability of decision outcomes (Bhat et al., 2004). The latter approach is, on the other hand, a computer program implementation of a production system model, which is a set of rules in the form of condition-action (if-then) pairs that specify how a task is solved (Gärling, Kwan, & Golledge, 1994; Shabanpour, Javanmardi, Fasihozaman, Miralinaghi, & Mohammadian, 2017). However, it is important to note that the above two approaches have been neither exclusive nor exhaustive. Several other approaches, including: (a) time-space prisms and constraints, (b) operations research/mathematical programming approaches, and (c) agent-based approaches have been employed, either in combination with the above approaches or separately, to develop activity-based model systems (Pinjari & Bhat, 2011).

This review is not going into the details of the many paradigms, frameworks and techniques in activity-based modelling, which are well reviewed by Henson, Goulias, & Golledge (2009), Pinjari & Bhat (2011) and Rasouli & Timmermans (2014), etc. Instead, I will particularly focus on the treatment of built environment features in the existing model systems, e.g. what built environment features are included and how they are accounted for in the behavioural process. This issue is reviewed and summarised based on the models that emerged or are actively updated after 2000, which may not be exhaustive, but it is not very possible that highly influential and referenced models could be left out (see **Table 2-5**).

It turns out that most existing model systems include only zonal 'size' variables in the aspect of built environment, such as population, employment by sectors or number of commercial establishments. Size variables are usually used to weight the likelihood for a zone to be selected as the location of an activity, either through simple statistical distribution, or econometric models, or some other functions. This is conceptually similar to four-step models which estimate the attractions of traffic zones and distribute travel demands based on zonal 'size' characteristics (McNally, 2007). When a model incorporates a dynamic module of route choice and traffic flow estimation, the road network can also be considered as an included built environment feature (e.g. in RAMBLAS, MATSIM). The SACSIM model (and probably other models in the same 'family') takes most account of built environment features. It is related to the fact that the model system is composed of a series of discrete choice models, which is particularly convenient and straightforward for the inclusion of an extended set of explanatory variables. Nonetheless, to the author's knowledge, the 'D-variables' in the built environment-travel research as mentioned before are never fully accounted for in existing model systems.

A reason for this gap could be that there are basically two groups of people working in the two fields: those interested in the impacts of nuanced built environment features are more urban planning and design oriented, while those building models tend to have a stronger background in transport planning and civil engineering. Therefore, my research will build on this gap and link the activity-based modelling approach with the analysis of the built environment-travel relationship. The development of an activity-based model that takes full account of the built environment will, on one hand, improve the comprehensiveness and realism of travel modelling, and on the other hand, enable the analysis of the detailed and decomposed influence of the built environment on the behavioural processes of daily travel.

Table 2-5 Summary of the inclusion of built environment features in existing activity-based models

Model names and key references	What built environment features are included and how
RAMBLAS (Veldhuisen, Timmermans, & Kapoen, 2000)	**Features:** Land use per zone, population per zone, dwellings by type **How:** The destinations for shopping and services are drawn based on the distribution of employment in the relevant services. The destinations for social participation and social contacts are drawn based on the distribution of households.
SIMAP (Kulkarni & McNally, 2000)	**Features:** density by land use **How:** to assign a selection likelihood to candidate activity locations.
CEMDAP (Bhat et al., 2004)	**Features:** zone-level land use, zonal basic/service/retail employment levels **How:** as independent variables in the econometric model for household activity-generation and activity location choice.
PCATS & FAMOS (which incorporates PCATS as the activity-travel module) (Kitamura, 1996; Pendyala, Kitamura, Kikuchi, Yamamoto, & Fujji, 2005)	**Features:** zone size, population density, commercial employment **How:** as explanatory variables in the nested logit model of destination-mode choice.
TASHA & ILUTE (which incorporates TASHA)	**Features:** population per zone, employment per zone, whether the zone is the city core

(Miller & Roorda, 2003; Roorda, Miller, & Habib, 2008; Salvini & Miller, 2005)	**How:** to estimate the probability of choosing a zone as the location of an activity.
SACSIM (Bowman & Bradley, 2005)[b]	**Features:** mixed use density, intersection density, purpose-specific size in parcels, parking and employment mix, accessibility from home, accessibility to nearest transit stop **How:** as explanatory variables in the econometric models for car ownership, activity generation, mode choice and destination choice.
AURORA & PUMA (which incorporates an updated version of AURORA) (Ettema, de Jong, Timmermans, & Bakema, 2007)	**Features:** NA **How:** to calculate the attractiveness of a location, which is then used to model the probability of location choice.
ALBATROSS & FEATHERS (Arentze & Timmermans, 2004; Bellemans et al., 2010)	**Features:** total amount of floor space and number of employees per sector per zone **How:** input to the decision tree.
MATSIM (Balmer et al., 2009)	**Features:** land-use information about the capacities of different activity types like 'work', 'shopping', 'education', etc. **How:** to indicate potential activity locations.
ADAPTS (Auld & Mohammadian, 2009; Auld & Mohammadian, 2012)	**Features:** zonal size variables, including the land-use area and employment by various categories. **How:** as explanatory variables in the multinomial logit models of destination choices.

a Note that many published articles do not describe every detail of the model, therefore the information in the table may not be absolutely complete.

b There are several other models in the same 'family', which include models for Portland Metro I/II, San Francisco SFCTA, New York NYMTC, Columbus MORPC, Atlanta ARC, etc. (Bradley & Bowman, 2006). The inclusion of built environment features can be more or less different from in SACSIM. However, it is difficult to find detailed technical documents of these models as the technical memos for SACSIM (http://jbowman.net/#Implementation).

c Full names of the models are:

RAMBLAS - Regional Planning Model based on the Micro-Simulation of Daily Activity Patterns

SIMAP - Microsimulation of Daily Activity Patterns

CEMDAP - A Comprehensive Econometric Micro-Simulator for Daily Activity-travel Patterns

PCATS - Prism-Constrained Activity Travel Simulator

FAMOS - Florida Activity Mobility Simulator
TASHA - Toronto Area Scheduling Model for Household Agents
ILUTE - Integrated Land Use, Transportation, Environment
SACSIM – Sacramento Activity-based Travel Demand Model
PUMA - Predicting Urbanisation with Multi-Agents
ALBATROSS - A Learning Based Transportation Oriented Simulation System
FEATHERS - Forecasting Evolutionary Activity-Travel of Households and their Environmental RepercussionS
MATSIM – Multi-Agent Transport Simulation
ADAPTS – Agent-based Dynamic Activity Planning and Travel Scheduling

2.5 Chapter summary

This chapter provides a glance at the theories, the empirical findings and the progresses in the activity-based modelling approach on the topic of the built environment-travel relationship. Besides, building on the notion of utility maximisation in travel decision making, a conceptual framework of travel utility changes in relation to changes in the built environment and a series of assumptions are proposed on the influence of various built environment features on the travel behavior (indicated by the total travel distance and VMT), which will be examined against the simulation results in Chapter 6.

Most existing research employ statistical methods and regress between synthesised measurements of travel behaviour (e.g. VMT, total walking distance) and built environment and socioeconomic conditions, which can be termed as a behaviourally 'top-down' analysis. On the contrary, there is hardly any research that takes a behaviourally 'bottom-up' approach that examines the decomposed influence of the built environment on various behavioural facets of daily travel. This corresponds to the major gap to be addressed in this research as mentioned in the Introduction. This gap is reinforced by the fact that the built environment is usually not sufficiently accounted for in existing activity-based models, which have the strength in simulating the behavioural process of daily activity-travel. This research will therefore build upon this gap and link the activity-based modelling approach with the analysis of the built environment-travel relationship, which will enable the analysis of the detailed and decomposed influence of the built environment.

Empirically, significant influence is found of the built environment on daily travel, but questions remain about the sizes of the influence and the relative importance of various built environment features, which corresponds to the second gap mentioned in the Introduction. This gap is partly addressed by the meta-analysis in Section 2.3. The meta-analysis updates the work of Ewing and Cervero (2010) and examines the stability of the weighted average elasticities from existing research. The results will be further used as representation of the built environment-travel relationship in European and American cities and be compared with my own simulation results on Beijing later in the thesis (see Chapter 6).

Chapter 03 Data collection and pre-processing

3.1 The study area

As mentioned in the Introduction, Beijing is selected as the case of study as an example of high-density and rapid-growing Asian city, which provides a quite different urban context comparing with North American and European cities that have been extensively studied. The municipality of Beijing covers an area of 16,410 square kilometres that includes both urban and rural lands. The basic urban structure is shaped by circular freeways: starting from the 2nd ring road that surrounds the old city core, and expanding to the 6th ring road that connects town centres of outer-urban districts (Z. Yang, Cai, Ottens, & Sliuzas, 2013). In terms of economic activities, landscape and lifestyle, the 4th ring road could originally be roughly regarded as the boundary between the urban built-up and the peri-urban areas of the municipality. However, the urbanised territory has dramatically expanded alongside the 5th ring road since the 1990s (Z. Yang et al., 2013). Therefore, the area within the 5th ring road is selected as the study area in this research, which is approximately 670 square kilometres and covers most of the built-up area in the city. The study area is home to around nine million people according to the Sixth National Population Census (Sixth National Population Census Office of State Council, 2010), meaning an average density of 13,400 people per square kilometre.

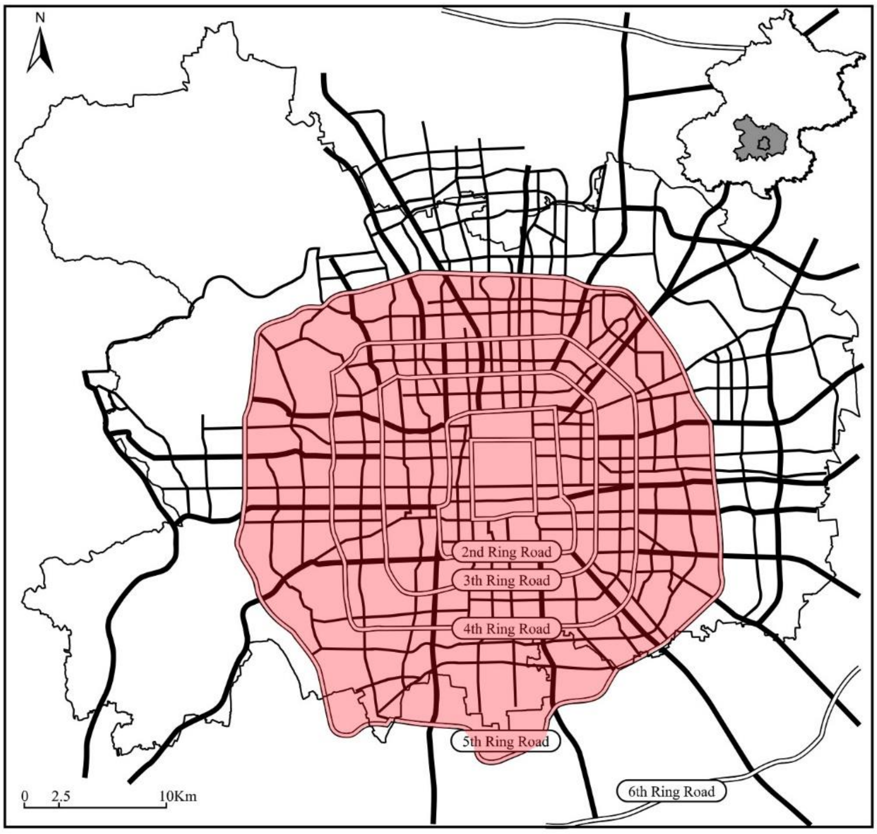

Figure 3-1 Base map of Beijing (coloured area indicates the study area)

Source: adapted from (P. Zhao, 2011)

3.2 Data sources

3.2.1 Data on travel behaviour

The data on travel behaviour in Beijing come from two field surveys: a large government-administered survey on people's 24-hour travel records and a small questionnaire survey on the decision making in daily travel conducted by myself.

The travel diary survey was conducted by the transport branch of the municipal government in 2010, as part of a series of large travel diary surveys in Beijing that have been conducted successively in 1986, 2000, 2005 and 2010. The survey area was the whole city, but a larger sampling weight was assigned to the central city according to the population density. The sample size was approximately 47,000 households containing around 116,000 individuals, which corresponds to a sampling rate of 1.5% of the total population. Both registered, long-term residents and unregistered migrants were included in the survey. The interviewees were selected using systematic sampling. The survey took the form of face-to-face survey that was fully administrated by interviewers, who read the questions to respondents and recorded the answers. It was required that all household members should be present so that the travel of the entire household can be recorded. The day of survey was evenly distributed from Monday to Sunday. The sampling was carefully controlled so that the samples collected on each day were spatially evenly distributed.

The spatial unit of the travel record is the Transport Analysis Zones (TAZ), which means that all the trip origins and destinations were recorded in TAZs instead of the exact coordinates. The whole city is divided into 1,911 TAZs by the transport authority. 652 of them are in the study area. The sizes of the TAZs are generally smaller in the city centre and larger in the inner and outer suburbs, which range a lot from 0.13 square kilometres (sqkm) to 382.03 sqkm. The variance in the sizes of the TAZs is much smaller in the study area, from 0.13 sqkm to 5.25 sqkm. The TAZs are delineated based on the following principles:

- They do not conflict with administrative boundaries.
- They are neither too big so that the traffic OD matrix can be generated in a high spatial resolution, nor too small so that there will not be too much random error.
- They are smaller where the density is high and the road network is dense (in the city centre) and larger where the density is low and the road network is sparse (in the suburbs).
- They do not extend across ‘natural boundaries’ such as rivers, railways.
- They do not extend across main roads and express ways.
- Special zones are delineated into separate TAZs such as the rail stations, large parks, and tourist sites.

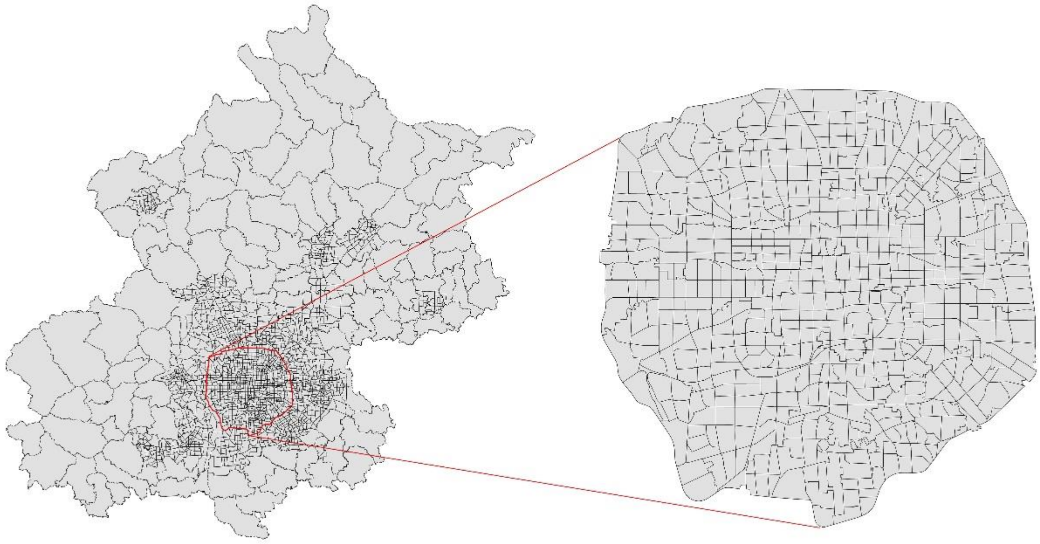

Figure 3-2 TAZs in the whole city and the study area

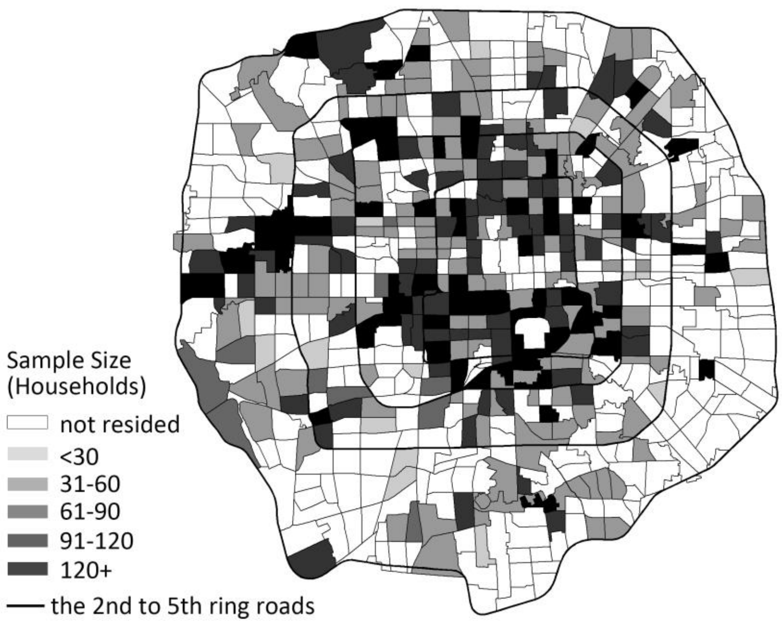

Figure 3-3 Spatial distribution of samples

The following information is recorded in the survey:

Demographic and socioeconomic information (household-level)

- Home location (in the unit of TAZ);
- Vehicle ownership, including car, motorcycle, bicycle, electric bicycle;
- Property right of the apartment/house that the interviewees are living in, choose

from self-owned, owned by a state-owned enterprise, rented, borrowed and others;

- Building type of the apartment/house that the interviewees are living in, choose from apartment, informal apartment, detached or semi-detached house and courtyard houses;
- Type of the apartment/house that the interviewees are living in, choose from commercial housing, houses built up by state-owned enterprises, affordable housing and others;
- Floor area of the apartment/house that the interviewees are living in;
- Household annual income.

Demographic and socioeconomic information (individual-level)

- Gender;
- Age;
- Residential registration status[1];
- Employment status, choose from full-time worker, part-time worker, full-time student, part-time student, pre-school child, retired, unemployed and others;
- Level of education, choose from pre-school, primary school, junior school, high school, technical school, bachelor and master and above.

Travel diary

- The start time of each trip in the day;
- The end time of each trip in the day;
- The origin of each trip in the day (in the unit of TAZ);
- The destination of each trip in the day (in the unit of TAZ);
- The purpose of each trip in the day, choose from sleeping, dining out, working, doing business, studying, personal business, housework, entertainment/sports,

[1] China imposes a residential registration system, in which each individual is officially registered to a place to live. If an individual migrates to a place where he/she is not registered at, he/she may not have access to a same level of public service or other rights as registered residents.

shopping, meeting friends, dropping off/picking up people, escorting people, dropping off/picking up goods;

- The travel mode of each trip in the day, choose from on foot, by car, by freight car, by motorcycle, by subway, by bus, by taxi, by shuttle bus, by school bus, by illegal taxi, by bicycle, by electric bicycle and others.

Table 3-1 Summary of the interviewees

Demographic and socioeconomic characteristics	**Distribution**
Household-level	
Car ownership	29.0%
Motorcycle ownership	3%
Bicycle ownership	63.0%
Electric bicycle ownership	14.0%
Property right of the apartment/house that the interviewees are living in	Self-owned: 68.9% Owned by a state-owned enterprise: 12.8% Rented: 16.1% Borrowed: 1.6% Others: 0.6%
Building type of the apartment/house that the interviewees are living in	Apartment: 84.4% Informal apartment: 2.4% Detached or semi-detached house: 0.2% Courtyard houses: 13.0%
Type of the apartment/house that the interviewees are living in	Commercial housing: 30.4% Houses built up by state-owned enterprises: 39.1% Affordable housing: 4.7% Others: 25.8%
Floor area of the apartment/house that the interviewees are living in	<50: 21.8% 50-75: 40.3% 75-100: 22.0% >100: 15.9%
Household annual income	<50 thousand RMB: 65.0% 50-100 thousand RMB: 27.6% 100-150 thousand RMB: 5.0% 150-200 thousand RMB: 1.4% 200-250 thousand RMB: 0.5% 250-300 thousand RMB: 0.2% >300 thousand RMB: 0.3%
Individual-level	
Gender	Male: 47.9%

	Female: 52.1%
Age	<=18: 10.7%
	19-40: 33.6%
	41-60: 35.4%
	>60: 20.3%
Residential registration status	Registered in the same district of the current residence in Beijing: 72.3%
	Registered in another district in Beijing: 10.1%
	Registered in another place in China: 17.4%
	Foreigner: 0.1%
	Others: 0.1%
Employment status	Full-time worker: 45.9%
	Part-time worker: 1.6%
	Full-time student: 7.3%
	Part-time student: 0.2%
	Pre-school child: 3.9%
	Retired: 29.0%
	Unemployed: 8.3%
	Others: 3.8%
Level of education	Pre-school: 3.9%
	Primary school: 10.1%
	Junior school: 21.6%
	High school: 18.1%
	Technical school: 23.2%
	Bachelor: 17.6%
	Master and above: 3.5%
	No education: 2.0%

It should be noted that although the face-to-face and interviewer-administered survey is advantageous in terms of the response rate and the correctness of the contents, it could also induce systematic errors. The most common error is that interviewees might get impatient and under report deliberately, which could result in an underestimation of the total amount of travel, especially short non-motorised trips (Kockelman, 1997). Besides, since only a one-day record was taken, occasional trips, which may happen on a weekly or monthly basis, are likely to be undersampled (Kockelman, 1997). These systematic errors are difficult to be rectified.

The small questionnaire survey was conducted in March and October 2015. The purpose of this survey was to collect information on the process of the decision making related to daily activity participation and travel. The sample size was two hundred individuals randomly drawn from eight residence compounds, twenty-five interviewees in each. The residence compounds were selected at both the city centre and the city fringe and covered a housing price range from 24,000 RMB per square metre to 88,000 RMB per square metre when the survey was conducted, as shown in **Figure 3-4** and **Table 3-2**. The sampling of the interviewees aimed to approximate the travel diary survey as well as possible. For this purpose, the distributions of gender, age and household type were controlled to be consistent with the travel diary survey, as shown in **Table 3-3**.

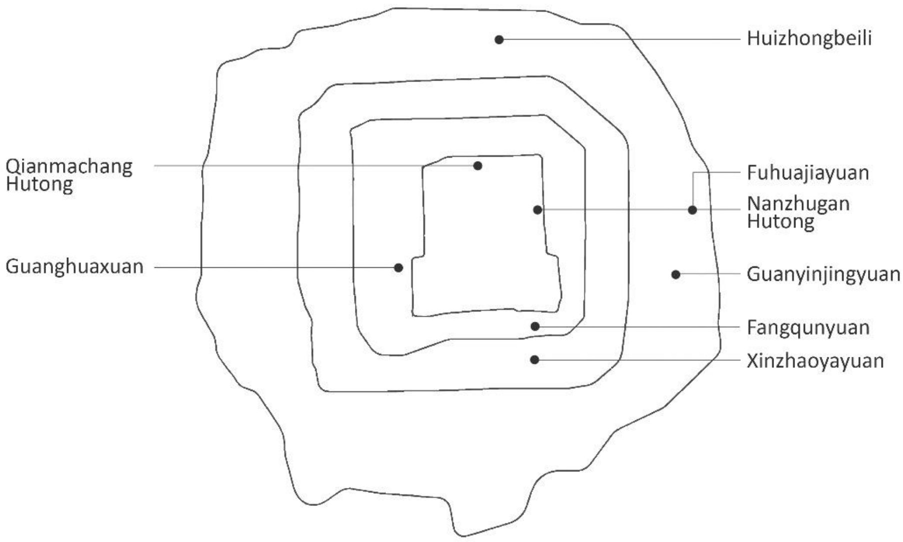

Figure 3-4 Locations of the selected residences in the small survey

Table 3-2 Housing prices of the selected residences in the small survey

Residence	**Average Price (RMB per square metre)**
Guanghuaxuan	40,000
Qianmachang Hutong	88,000
Nanzhugan Hutong	48,000
Fangqunyuan	39,000
Xinzhaoyayuan	37,000
Guanyinjingyuan	24,000
Fuhuajiayuan	26,000

Huizhongbeili	39,000

Table 3-3 Demographic characteristics of the interviewees in the small survey

Characteristics	**Proportion%**
Gender	
Female	50
Male	50
Sum	*100*
Age	
19-40	23
41-60	38
60+	39
Sum	*100*
Household type	
Single	12
Couple	34
Core family	32
Others	22
Sum	*100*

The questionnaire included two parts: basic information and the information on travel decision making. The former included gender, age, household type, car ownership, residential registration status, employment status and the level of education. The latter was composed of four questions as listed below:

- What is your first consideration, when you make plans about your activities (except work) on **weekdays**? Choose from 'what shall I do today', 'when shall I go', 'shall I go by car/subway/bus/walk/..', 'where shall I go' and 'how far shall I go'.
  What is your second consideration, if you could specify.
  What is your third consideration, if you could specify.
- What is your first consideration, when you make plans about your activities (except work) on **weekends**? Choose from 'what shall I do today', 'when shall I go', 'shall I go by car/subway/bus/walk/..', 'where shall I go' and 'how far shall I go'.
  What is your second consideration, if you could specify.
  What is your third consideration, if you could specify.
- When deciding about the activity destinations on **weekdays**, which do you prefer?

Choose from ‘decide all destinations together’ and ‘first decide long-stay/primary destinations and then short-stay/intermediate stops’.

- When deciding about the activity destinations on **weekends**, which do you prefer? Choose from ‘decide all destinations together’ and ‘first decide long-stay/primary destinations and then short-stay/intermediate stops’.

3.2.2 Data on the built environment

The information on the built environment is drawn from multiple sources, as listed below. Section 3.4 will provide detailed explanation on how these data serve the measurement of various aspects of the built environment. It should be noted that there are time gaps between a few built environment data sets and the travel diary survey due to constraints of data availability, which range from one to six years. However, considering that the change of the built environment is generally very slow (Batty, 2013), such time gaps do not seem to pose a major problem for the analysis. The data sets are:

- Population and employment data in 2010 from the Sixth National Population Census (Sixth National Population Census Office of State Council, 2010). The released data are at the Jiedao (sub-district) level, one of the smallest political divisions in China. Jiedao is a larger spatial unit than TAZ, which composes of six TAZs in average (there are 109 jiedaos and 652 TAZs within the study area).

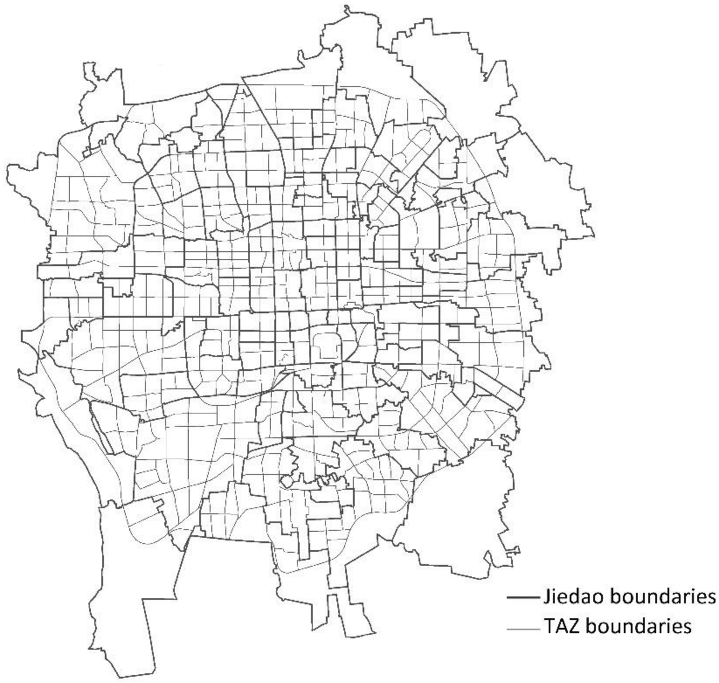

Figure 3-5 Boundaries of Jiedao and TAZ

- Point of Interest (POI) data in 2011 produced by NavInfo, a digital map producer in China. The data include twenty types of interest points, which are government buildings, airports/ports, railway and subway stations, bus stops, gas stations, parking lots, motorway service areas, highway toll stations, banks, commerce and office building, retail facilities, hotels, restaurants and entertainment facilities, hospitals, educational institutions, companies, parks and plazas, residences and others. Those that are of particular interest for this research are subway stations, bus stops, parking lots, retail facilities and restaurants and entertainment facilities. It should be noted that, similar to many new types of data, there is not yet thorough study on comparing the data with ground truth in the context of China. However, an advantage of the NavInfo POI lies in that, unlike voluntarily contributed geographic information (VGI), the data collection process was managed and controlled by the company and passed the ISO/TS16949[2] examination (NavInfo,

[2] The ISO/TS16949 is an ISO technical specification aimed at the development of a quality management system that provides for continual improvement, emphasizing defect prevention and the reduction of variation and

2009).

- Land use map in 2004 from the Beijing Master Plan (2004-2020).
- Open Street Map (OSM) data on road network, retrieved in 2013. Regarding to the issue of data quality , a few research have found that OSM information can be fairly accurate comparing with Ordnance Survey or ground truth (Haklay, 2010; Mashhadi, Quattrone, Capra, & Mooney, 2012).
- Data on parking lots and parking spaces, obtained from 51Parking in 2013, a service provider for real time parking information.
- Street view images obtained from Baidu Map, the Chinese equivalent of Google Map. The images were requested at an interval of 200 metres along all the streets in the study area in February 2016, resulting in 360,796 images in total (800*500 pixels). The treatment of the images will be explained in the next chapter.

3.2.3 Other data sets

Two other data sets are employed in this research. The first is the housing prices of residence compounds (RMB per square metre) in Beijing in 2011, obtained from the largest real estate website in China, Fang.com. The data are used in conjunction with the housing information of the interviewees in the travel diary survey to estimate the overall socioeconomic well-being of a household, which will be explained in detail in Section 3.3.

The second data set is the travel time between all TAZ pairs (652*652/2=212,552 pairs in total, measured from the centroids of TAZs), obtained from Baidu map. This information is inputted to the model of destination choice and mode choice as the expected travel time using different modes between two locations. The travel time of four travel modes are provided by the Baidu map, which are driving, taking public transport, cycling and walking. Since the travel time provided by Baidu is real time

---

waste in the automotive industry supply chain.

estimate based on the traffic conditions and is subject to fluctuations, the data query and download was restricted to non-peak hours and was done at different time points for three times to calculate the average.

3.3 Socioeconomic data pre-processing: creating an indicator of overall socioeconomic well-being

In the travel diary survey, several socioeconomic features are collected for each individual and household, which include: the residential registration status, the occupation, the education level, the annual income of the household, the vehicle ownership, the housing condition and the housing property right. Each of these features (except for vehicle ownership) does not seem to be explicitly linked with travel behaviour. However, the overall socioeconomic well-being indicated by some of these features may exert an influence on daily travel, which may be a combined effect of budget, pressure, life style and so on. Therefore, an effort is made in creating an indicator of overall socioeconomic well-being for the households in the data set. The well-being is analysed at the household-level instead of individual-level considering the dependency among household members.

3.3.1 Methods

Latent class analysis (LCA) is applied to stratify the sample households into different levels of socioeconomic well-being. LCA identifies unmeasured class membership from multiple observed characteristics. The number and the sizes of classes are taken as unknown. It is similar to standard cluster analysis techniques in that the goal is to form segments. However, LCA is more preferable for this task for the following reasons. First, LCA assumes the existence of a latent variable that induces spurious relationships among the observed variables rather than just looking for similarities (Hagenaars & McCutcheon, 2002). This corresponds to the notion that there is a latent social class membership that links to the differences in various socioeconomic features. Second, LCA is similar to standard cluster analysis techniques in that the allocation of objects to clusters should be optimal according to certain criteria. However, the choice of the criterion is more arbitrary in standard cluster analysis (Hagenaars & McCutcheon, 2002), such as a distance measure that is arbitrarily chosen. Other advantages of LCA include providing a probabilistic estimate of object class membership and being more flexible in terms of the data type. Due to these advantages, LCA is becoming a more popular clustering tool (Hagenaars & McCutcheon, 2002).

When specifying the LCA model, different numbers of latent classes and different combinations of variables are tested and compared. The best fitting model is selected based on performance indicators which include Akaike's Information Criterion (AIC), Bayesian Information Criterion (BIC), log-likelihood and G-square fit statistics (Linzer & Lewis, 2011). Models with low AIC, BIC, G-square fit and a high log-likelihood are preferred. The interpretability of the identified classes are also considered in the model selection process (Byles et al., 2016; Lanza, Flaherty, & Collins, 2010). The analyses is performed using R package 'poLCA' (Linzer & Lewis, 2011).

According to prior studies, factors related to socioeconomic differentiation in urban China include income, occupation, education level, housing condition and so on (Logan, Bian, & Bian, 1999; Wu & Li, 2005). Based on these findings and the data availability of the transport survey, the following variables are considered in the latent class analysis.

- **Housing condition**: Housing property accounts for nearly 70% of the total assets of Chinese households (Li, Luo, Lu, Deng, & Gan, 2016). The housing-price-to-income ratio was up to 15 in Beijing by 2010 (Tan & Zhao, 2012). Therefore, the socioeconomic well-being of a household is largely determined and reflected by their housing condition. Four variables related to housing condition are considered and tested in the analysis: housing type, housing floor area, housing floor area per capita and the market value of the property. The market value of the property is estimated from the housing price data obtained from the largest real estate website in China, Fang.com.

- **Car ownership**: Despite the fast increase in car ownership in the post-reform era, the rate of car ownership was still only 25% in Beijing by 2010 (Beijing Transportation Research Center, 2011). Therefore, car ownership may also be considered as a representation of a household's socioeconomic capability.
- **Education**: Higher education usually indicates a better income and occupation (Bian and Logan, 1996) and higher social status (Wu & Li, 2005). Two relevant variables are considered in the analysis: the average education level of all adult household members and the highest education level among all adult household members.

Table 3-4 Variables considered in LCA

Variables	Values
Housing floor area (square meters, sqm)	<50, 50~75, 75~100, 100~150, 150~200, >200
Housing floor area per capita (sqm)	<10, 10~20, 20~30, 30~40, 40~50, 50~60, 60~70, >70
Housing type	Old one-floor housing, affordable housing, matchbox housing, commercial housing
Housing market value (million RMB)	<200, 200~300, 300~400, 400~600, >600
Car ownership	No car, one car, more than one cars
Highest education in the household	No education, primary school, secondary school, high school, technical school, junior college, bachelor, post-graduate
Average education in the household	No education, primary school, secondary school, high school, technical school, junior college, bachelor, post-graduate

3.3.2 Results

By testing all possible combinations of these variables and different numbers of clusters, it is found that the model performs best (produces the lowest AIC, BIC, G-square fit statistics and the highest log-likelihood) when the class number is three. The combination of variables that produce the best performance are the highest education in the household, car ownership, housing market value and housing floor area. The three classes are labelled as ‘the best-off’ (Stratum 1), ‘the middle class’ (Stratum 2), and ‘the least well-off’ (Stratum 3).

The profiles of the three social strata are as follows. Households in the first stratum usually have at least one member with a bachelor’s degree or higher. More than half of these households own at least one private car. Around seventy percent of these households live in an apartment/house (owned or rented) with a market value of more than four million RMB. Besides, their apartments/houses are all larger than seventy-five sqm. In the second stratum, less than half of the households have a member with a bachelor’s degree or higher. Around thirty percent of them own a private car. About ninety percent of these households live in apartments smaller than seventy-five sqm and with market values of less than four million RMB. In the third stratum, only less than thirty percent of the households have a member with a bachelor’s degree or higher. Less than twenty percent of the households own a private car. Most of these households live in apartments/houses smaller than fifty sqm and with market values of less than two million RMB.

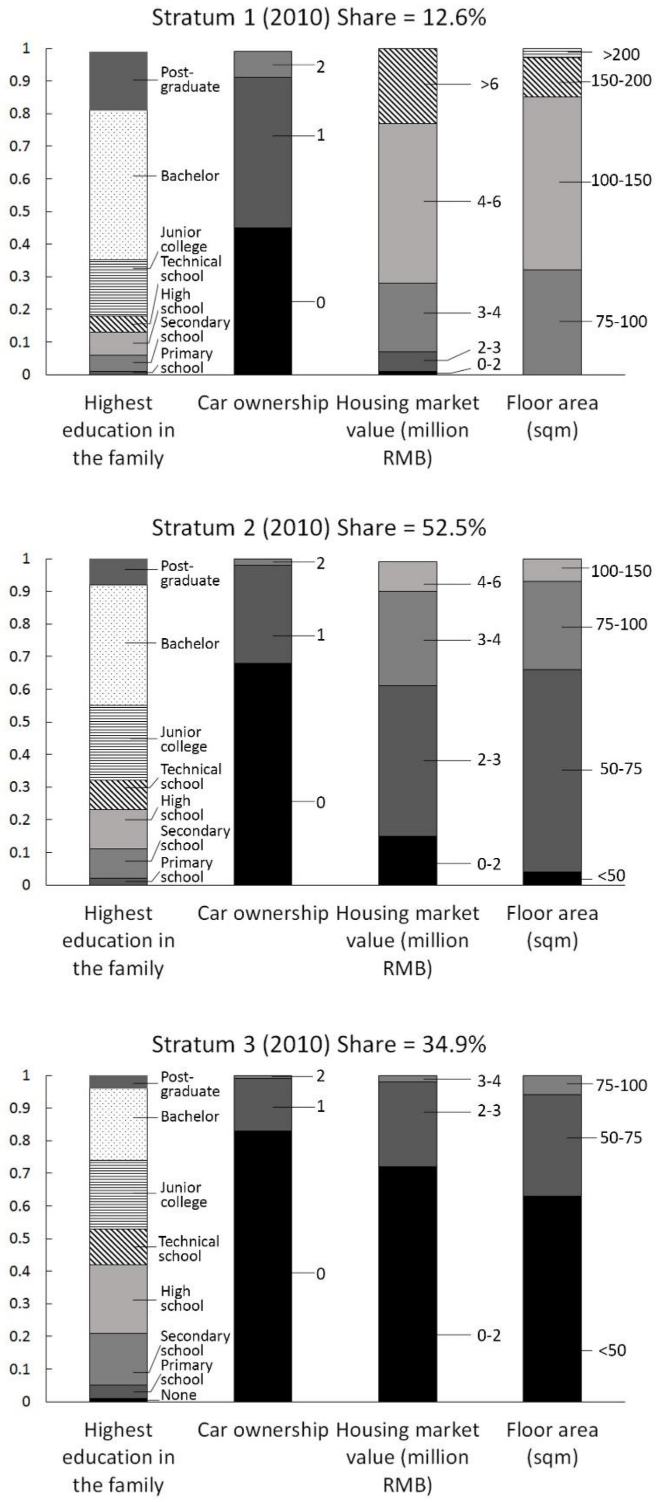

Figure 3-6 Socioeconomic characteristics of the three social groups identified by LCA

3.4 Built environment data pre-processing: measuring land use-related features using GIS

As mentioned before, in travel research, the built environment have often been described through features named with words beginning with ‘D’. Six types of ‘D-variables’ have been mentioned by existing research as potentially influential to travel behaviour, followed by a seventh ‘D’, street facade design, introduced by this research. This section will deal with the measurements of the first six D-variables, which are two-dimensional and land use-related, while the measurement of the street facade design, which involves more advanced methods, will be explained in the next chapter.

3.4.1 Measuring density

Density is always measured as the variable of interest per unit of area, which can be population, employment, dwelling units, or something else (Ewing & Cervero, 2010). In this research, four variables of interest are considered, which are population, employment, retail facilities and entertainment facilities (including restaurants). The data for population and employment density come from the Sixth National Population Census (Sixth National Population Census Office of State Council, 2010). The data for retail and entertainment density come from the POIs produced by NavInfo.

Actually, for retail and entertainment facilities, the difference between density and accessibility is vague, since a high density of facilities in an area is usually related to a high accessibility to facilities within this area, which is sometimes terms as ‘local accessibility’ (Handy, 1993). Therefore, while the measurements of population and employment density are straightforward, five indicators of retail and entertainment density/accessibility were tested and compared, which were:

- Density, calculated as the number of facilities in a TAZ divided by the area of the TAZ;

- Accessibility to facilities within 400 metres, calculated as the average number of facilities within a 400-metre radius from sample points in a TAZ, which are selected using a 200m * 200m grid;
- Accessibility to facilities within 800 metres, calculated as the average number of facilities within an 800-metre radius from sample points in a TAZ, which are selected using a 200m * 200m grid;
- Accessibility to facilities within 400 metres at the TAZ centroid, calculated as the number of facilities within a 400-metre radius from the centroid of a TAZ;
- Accessibility to facilities within 800 metres at the TAZ centroid, calculated as the number of facilities within an 800-metre radius from the centroid of a TAZ.

The choice of the 400-metre and 800-metre radius is based on previous findings on travel behaviour. 400 metres is identified to be the median distance of walking (Porta, Romice, Maxwell, Russell, & Baird, 2014). Besides, it is long held belief that a spatial scale of 400 metres represents the local walkable neighbourhoods (Appleyard, 1980; Sarkar et al., 2015). 800 metres is also a commonly used distance for creating buffers in the physical activity literature as an 'easy walking distance' (Colabianchi et al., 2007). It turns out that all of the five measurements are highly correlated (Pearson's correlation coefficient > 0.8). Therefore, the simplest measure, the density, is chosen.

Table 3-5 Correlation matrix of the five measurements of entertainment density

	Den	Ave_acc_400	Ave_acc_800	Cen_acc_400	Cen_acc_800
Den	1	0.94	0.89	0.88	0.89
Ave_acc_400	-	1	0.97	0.93	0.97
Ave_acc_800	-	-	1	0.87	0.99
Cen_acc_400	-	-	-	1	0.86
Cen_acc_800	-	-	-	-	1

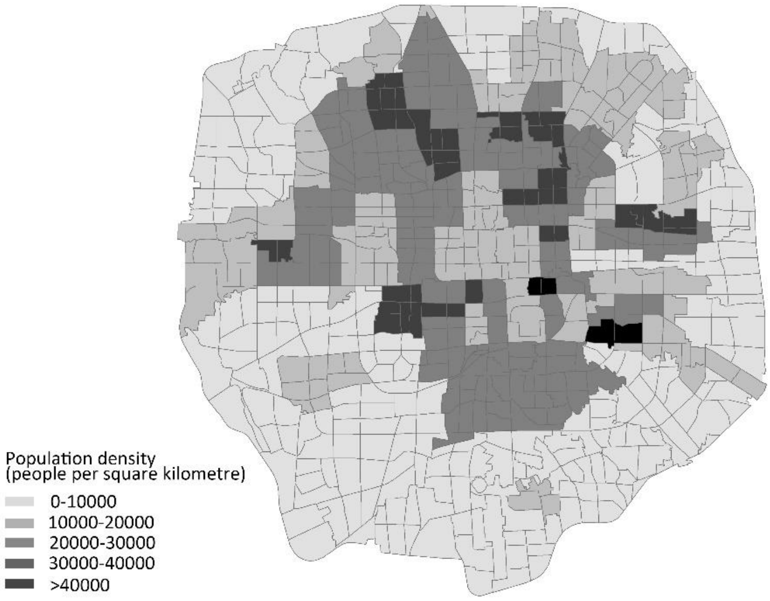

Figure 3-7 Population density

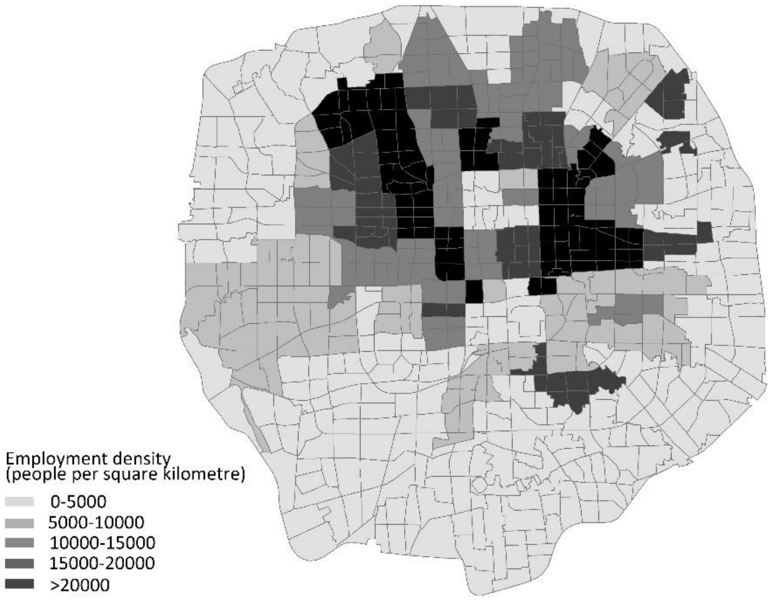

Figure 3-8 Employment density

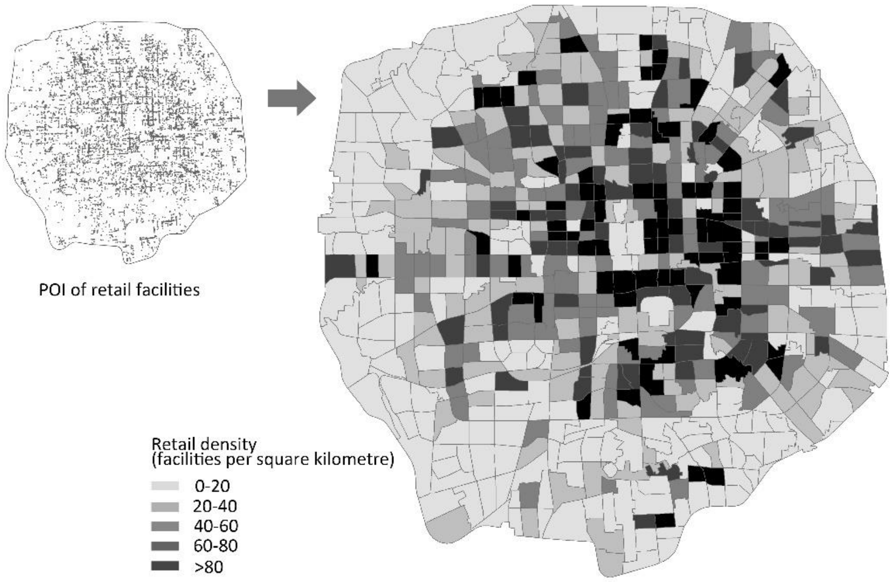

Figure 3-9 Retail density

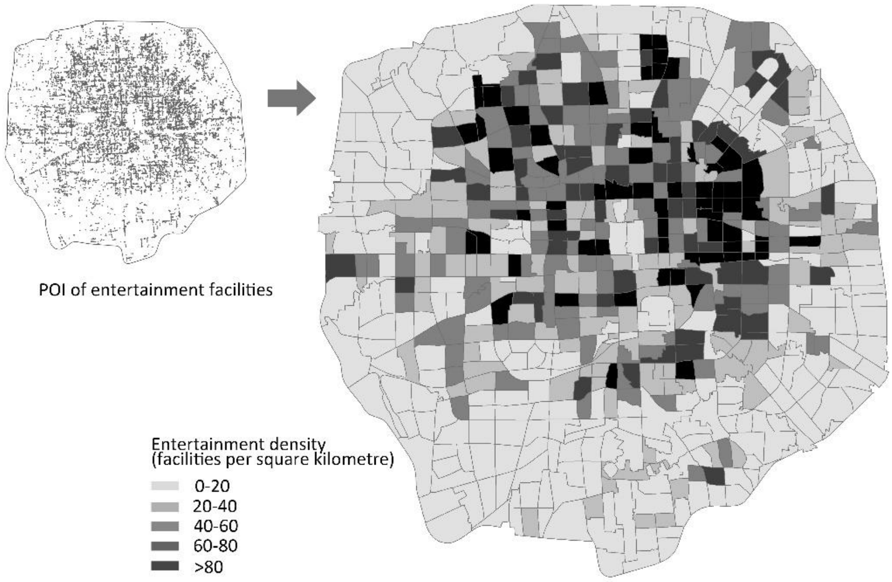

Figure 3-10 Entertainment density

3.4.2 Measuring diversity

Diversity measures pertain to the number of different land uses in a given area and the degree to which they are represented in land area, floor area, or employment (Ewing & Cervero, 2010). The measures of diversity range from simple functions of population to employment ratios (e.g. Rajamani et al. 2003; Bento et al. 2005) to more complex methods, such as zonal entropy-based methods (Greenwald, 2006). In this research, the entropy measure is used since it involves more types of urban functions than the simple job-residence mix and is more widely used in travel studies (Ewing & Cervero, 2010). The entropy is measured in terms of the land area in this research, which is derived from the land use map of 2004 published with the Beijing Master Plan (2004-2020). It should be noted that there is a temporal gap between the land use data and the travel survey is the largest in this research. However, because the land use information is treated as highly confidential in Beijing, the land use map in 2004 is already the most recent dataset that is publicly accessible. The seventeen land use types in the map[3] are collapsed into five functional categories, which are residential, commercial, educational, public service, and natural (greenery and waterbody). The entropy is calculated as

$$Entropy_j = \sum_1^5 -p_{ij} * Ln(p_{ij})$$

Where p_{ij} denotes the proportion of the land area of function i in TAZ_j. The value ranges from 0, which indicates single-use environments, to 1.61, which indicates perfect mix.

Figure 3-11 Land use mix

[3] The seventeen land use types are residence, public facility, commerce and finance, education, sports, industry, storage, railway, airport, road, square and parking lot, municipal facilities, greenery, special use, waterbody, agriculture, mix use.

3.4.3 Measuring destination accessibility

Destination accessibility measures the ease of access to trip attractions (Ewing & Cervero, 2010). It can be ‘local’ or ‘regional’ (Handy, 1993), which corresponds to the number of attractions in small or large buffer zones from a location. The measurements of accessibility range from simple distance to the nearest attractions to formulas that integrate the size of attractions and the impedance of travel for a given location (Handy, 1993). Although the latter measurements can provide a more comprehensive analysis, the indexes created from these formulas are less straightforward for planning policy making. Therefore, the simple distance measurement is used in this research. Since ‘local’ accessibility is already accounted for by the retail and entertainment density, two types of more ‘regional’ accessibility are measured here: the accessibility to commercial clusters and the accessibility to the city centre (can actually be considered as a ‘global’ accessibility). Commercial clusters refer to the concentrations of retail and entertainment facilities. This feature is measured in addition to the retail and the entertainment densities based on the notion that clusters may provide extra attraction to people since multi-tasks can be completed within short distance and a larger variety of goods and services can be found. While the measurement of the accessibility to the city centre is quite straightforward, which is simply the distance, the measurement of the accessibility to commercial clusters needs further explanation.

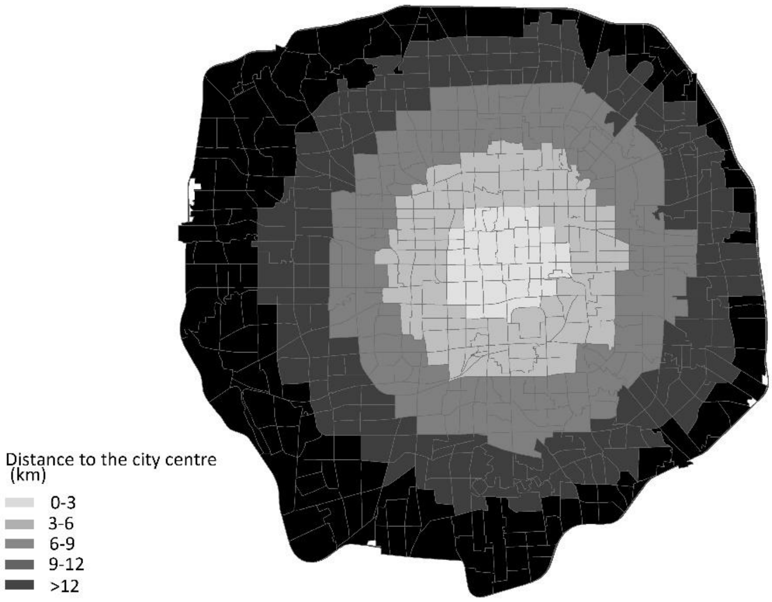

Figure 3-12 Distance to the city centre

First of all, commercial clusters need to be properly defined and identified, for which purpose the method of DB-SCAN is employed. DB-SCAN is a density-based method for discovering clusters in large spatial databases with noise (Ester et al. 1996). Two parameters need to be determined in the process: Eps, the search radius of neighbourhood, and MinPts, the minimum number of POIs in an Eps-neighbourhood (Ester et al. 1996). The clustering result produced with *Eps=x* and *MinPts=y* means that each point within a cluster should have at least *y* points within *x* distance from it. Actually, similar results can be produced by various combinations of Eps and MinPts values. A large Eps with a large MinPts may produce similar clustering results as a small Eps combined with a small MinPts. Therefore, the strategy used in this research is to fix the value of Eps to 200 (metres) and adjust the value of MinPts. The result produced with Minpt=40 is selected since the clusters are neither too few and too small nor too large and too many.

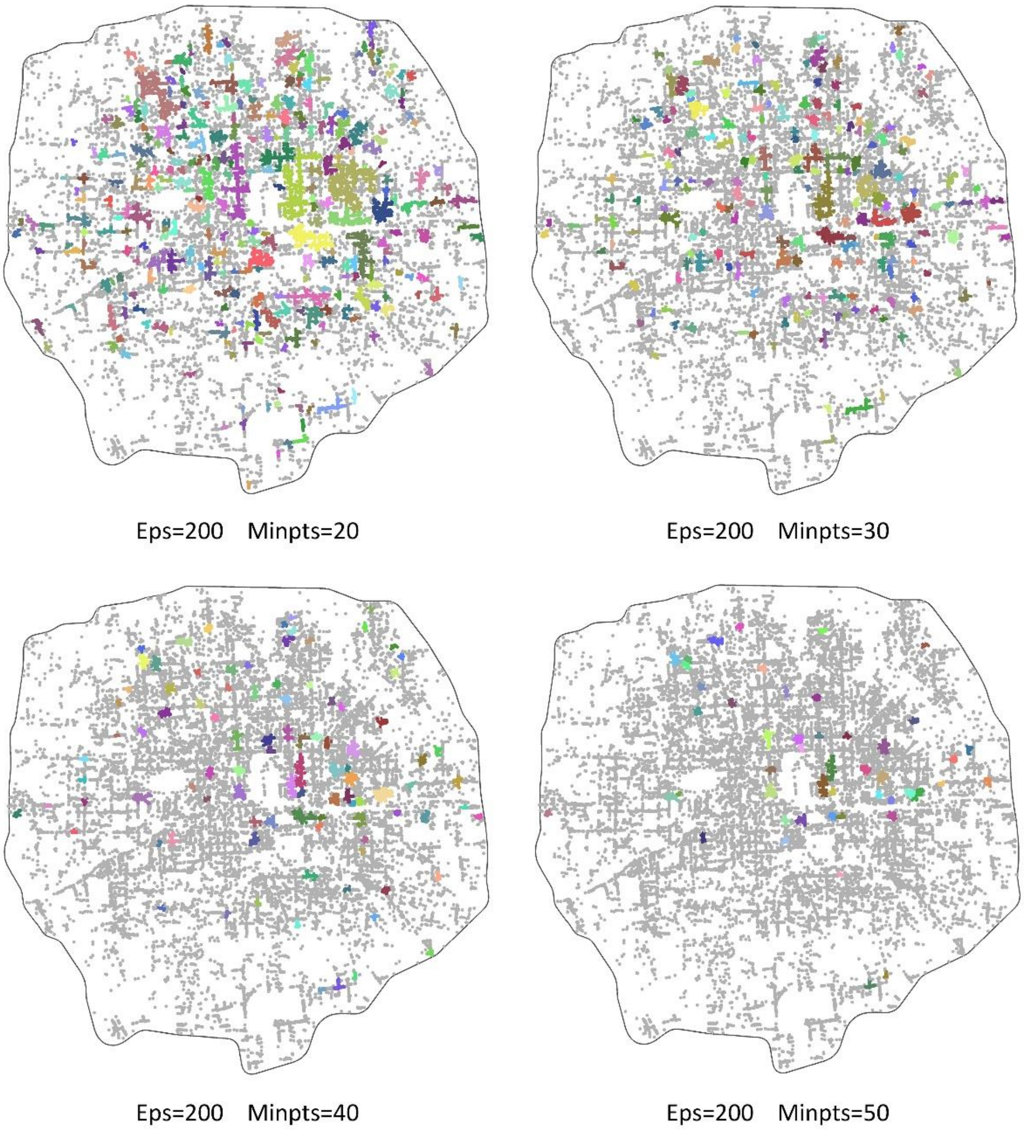

Figure 3-13 Clustering results using different parameters

Note: coloured points are clusters, grey points are those that do not belong to any cluster, colours may be reused

After identifying the commercial clusters, two measurements of the accessibility of a TAZ to these clusters are tested and compared. The two measurements are: (1) the distance from the centroid of a TAZ to the centroid of the nearest commercial cluster, (2) the average distance from a group of sample points in a TAZ, which are also selected using a 200m ∗ 200m grid, to the centroids of their nearest clusters. These two measurements turn out to be perfectly correlated (Pearson's correlation coefficient = 1).

As a result, the former measurement, which is simpler, is selected.

Figure 3-14 Distance to the nearest commercial cluster

3.4.4 Measuring (road network) design

As mentioned before, the 'design' factors in existing research are usually actually road network design (Ewing & Cervero, 2010). Measures include road density, intersection density, proportion of four-way intersections, and occasionally also sidewalk coverage, average street widths, street trees, or other physical variables that differentiate pedestrian-oriented environments from auto-oriented ones (Ewing & Cervero, 2010). However, the features related to the pedestrian environment are usually collected through expensive field audits, which take a lot manpower and are difficult to be implemented in the entire study area in this research. This problem is compensated by differentiating different types of roads when measuring road density. Three types of roads are considered based on the road hierarchy labelled in the OSM: primary roads, secondary roads and tertiary roads, which correspond to different levels of traffic volume and traffic speed. Lower-level roads are usually related to a better pedestrian environment. The OSM data include thirty labels in total. The mapping between the most commonly used labels and the road type in this research is shown in **Table 3-6**.

Table 3-6 Road classification

Label in OSM	**Road type in this research**
Tertiary& tertiary link	Secondary
Residential	Tertiary
Service	Tertiary
Footway[a]	-
Platform[b]	-
Secondary	Secondary
Motorway & motorway link	Primary
Trunk & trunk link	Primary
Unclassified	-

a Footways are not included in the measurement of road density because most of them are foot paths in parks and tourist sites, which are not actually part of the road system.
b Platforms are not included because they are not real roads.

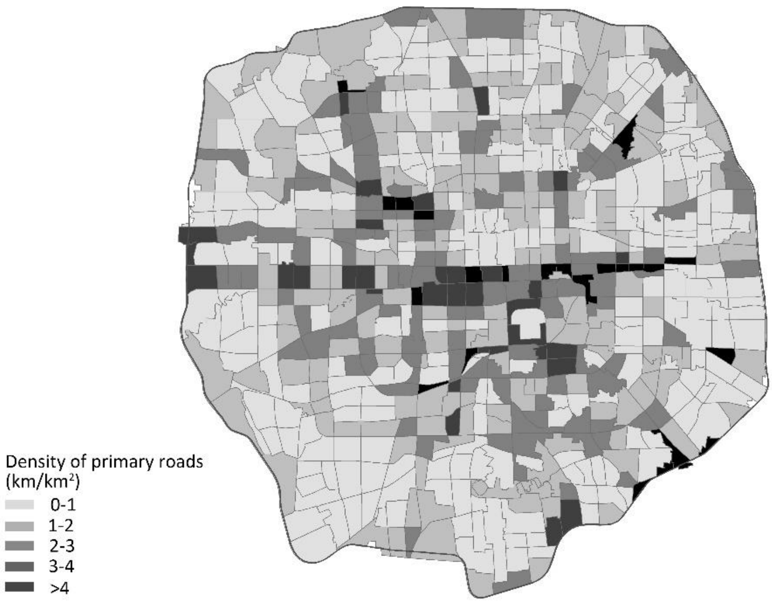

Figure 3-15 Density of primary roads

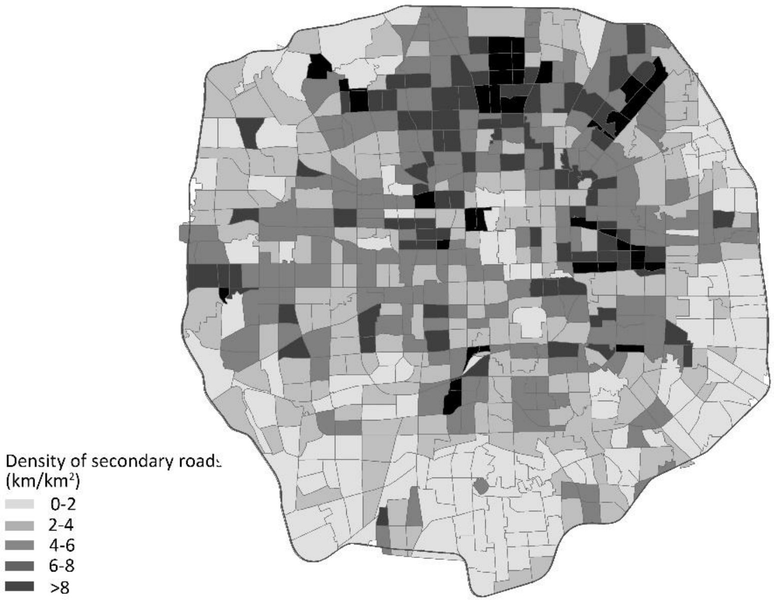

Figure 3-16 Density of secondary roads

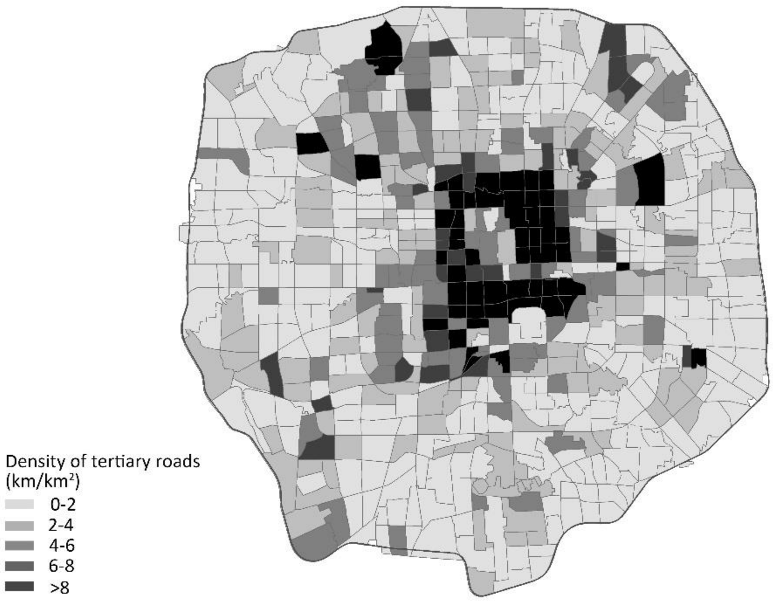

Figure 3-17 Density of tertiary roads

3.4.5 Measuring distance to transit

Two main types of public transit in Beijing are considered here: bus and subway. The locations of bus stops and subway stations are also from the POI data set. For subway, the distance to transit is measured as the distance from the centroid of a TAZ to the nearest subway station. The distance from the TAZ centroid is directly used since it is already proved that the direct measure from the centroid of a TAZ and the average measure from sample points in the TAZ are highly correlated. In terms of bus service, the index of bus coverage is more commonly used, which is calculated as the ratio between the area in a TAZ that is covered by a certain size of buffer zones of bus stops and the total area of the TAZ. Different buffer distances are tested and compared (see **Figure 3-18**). It turns out that the bus coverages calculated using 100, 200 and 300 metre buffer zones are highly correlated. The Pearson's correlation coefficients are 0.96 (100 metre and 200 metre), 0.87 (100 metre and 300 metre) and 0.96 (200 metre and 300 metre). The median number—200 metre buffer zone is selected.

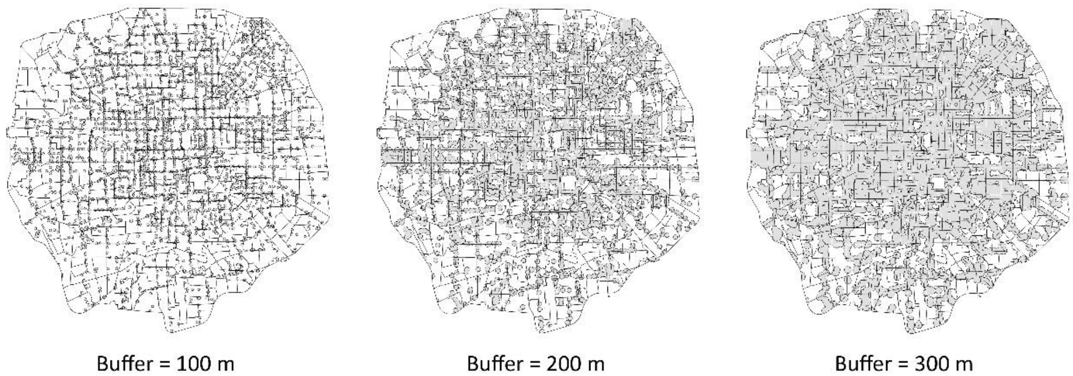

Figure 3-18 Areas of bus coverage using different buffer distances

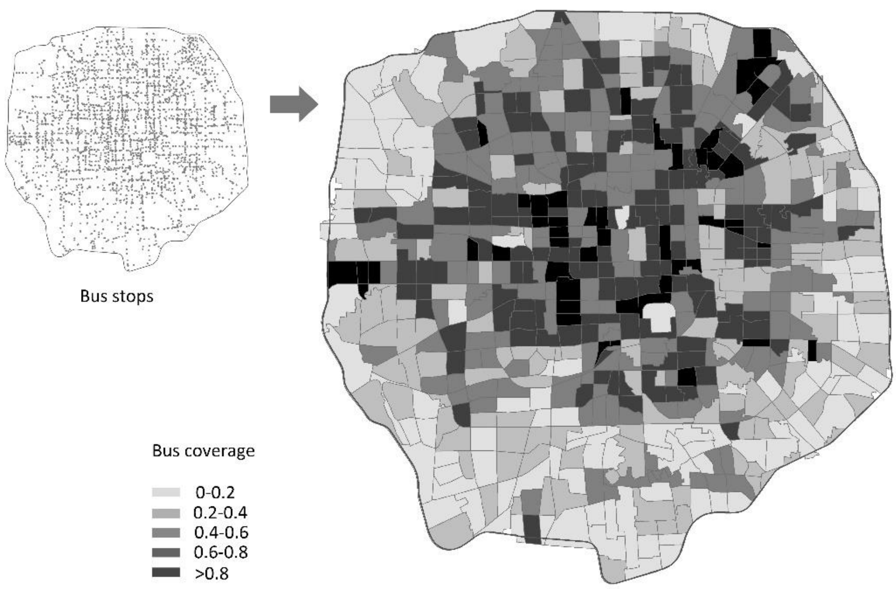

Figure 3-19 Bus coverage

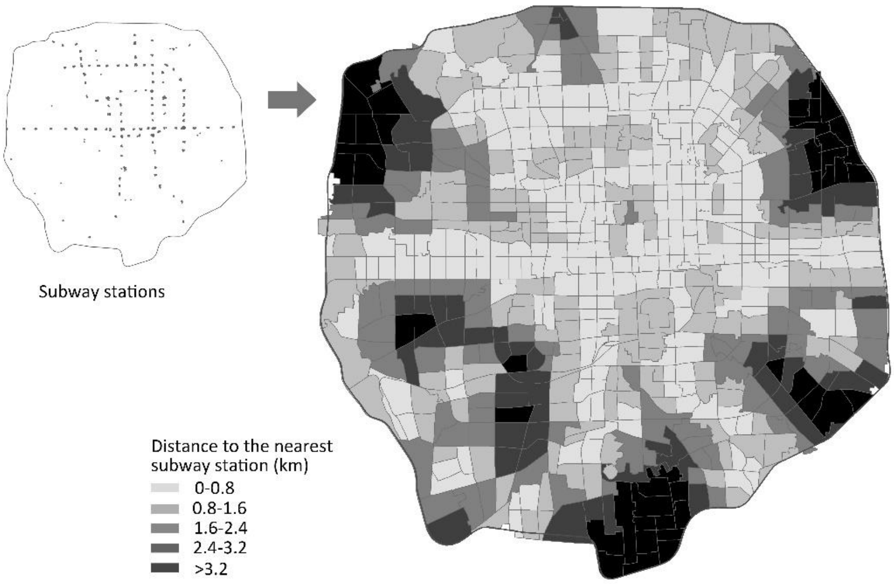

Figure 3-20 Distance to the nearest subway station

3.4.6 Measuring parking supply (demand management)

The measurement of parking supply is straightforward, which is calculated as the number of parking spaces in a TAZ divided by the area of the TAZ. Particularly, the number of parking spaces is used instead of parking lots, which provides more accurate estimation of the parking capacity of an urban area.

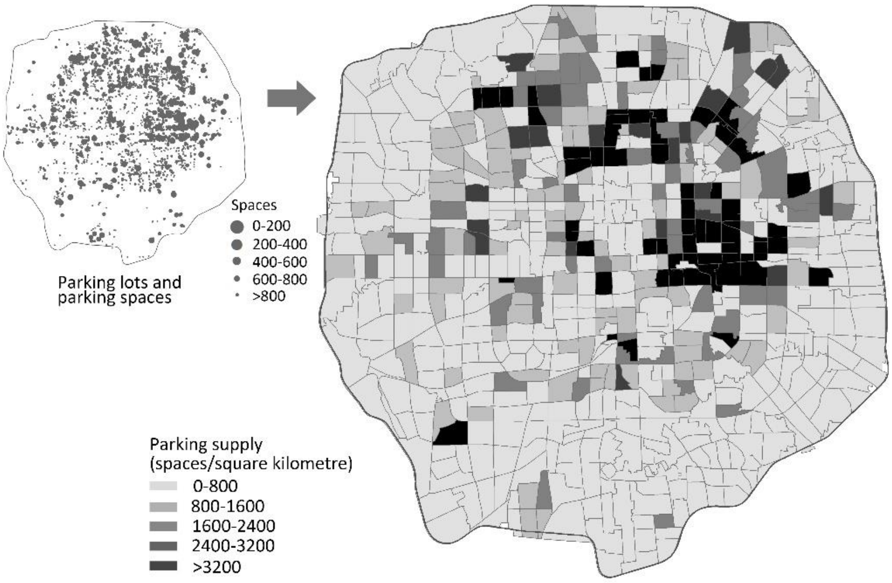

Figure 3-21 Density of parking

3.5 Chapter summary

This chapter introduces the study area, the data sources and the pre-processing of socioeconomic data and land use-related built environment features in the research. The study area is set to be the area within the 5th ring road of Beijing, which covers most of the built-up area in the city. Multiple data sources are used in this research. The information on travel behaviour is drawn from both a large travel diary survey and a small survey on travel decision making. The information on the built environment is drawn from both conventional census data and government documents (e.g. the Sixth National Population Census, the Beijing Master Plan), and the so-called 'big data' (e.g. POI data, OSM data, online parking data and street view images).

The pre-processing of the socioeconomic information in the travel survey involves creating an indicator of the overall well-being of the household. It is based on the notion that individual socioeconomic features may not be explicitly linked with daily travel behaviour, however, the overall well-being contributed by these factors may exert a significant influence on the pattern of daily travel. Latent class analysis is applied to stratify the sample households into three levels of well-being, namely the best-off, the 'middle class' and the least well-off.

The pre-processing of the built environment information includes the measurement of two-dimensional, land use-related features (the six 'Ds'), as well as the features related to the third dimension—the street facade (the seventh 'D'). The latter task is more complicated and described separately in the next chapter. Only the former task is described in this chapter. Thirteen features that belong to these six 'Ds' are measured, which are population density, employment density, retail density, entertainment density, land use diversity, distance to the city centre, distance to the nearest commercial cluster, primary road density, secondary road density, tertiary road density, distance to the nearest subway station, bus coverage and density of parking space.

Chapter 04 Advanced data pre-processing: measuring street facade features using machine learning algorithms

4.1 Why include street facade features

As mentioned in the introduction, the built environment features that have been examined in existing research are mostly two-dimensional, land use-related ones. The features related to the ‘third dimension’—the street facade, have received much less attention. However, the street facade, as an important component of the built environment, can also exert an influence on daily travel. The potential mechanisms of such influence can be both psychic and functional. Psychically, certain qualities of the street facade may foster positive or negative feelings, which connect or disconnect individuals to places (Sarkar et al., 2015). Positive feelings to the urban space are supposed to create the so-called ‘walkable’ places and induce more physical activity and active travel (Witten et al., 2012). Functionally, certain design and layout of the street facade may also relate to higher convenience and utility of travel, e.g. facade that provides plenty of space for street shops.

For the psychical impact, the fields of architecture and urban design have made many efforts in identifying the key qualities that contribute to people’s subjective experience. For instance, Moughtin (2003, p. 59) wrote that ‘order, unity, balance, symmetry, scale, proportion, rhythm, contrast and harmony are among the important tools used to define good architecture’. In urban design, rules of enclosure, coherence, variety and so on are widely acknowledged and discussed in many design handbooks as well as governments’ design codes (see for instance, Ewing et al., p. 8; American Planning Association, 2006, p. 165; Parolek, Parolek & Crawford, 2008, p. 41 for the narratives on enclosure).

Among the many street facade features, two are selected to be included in this research: one building-level feature—the construction and maintenance quality of building facade, and one street-level feature—the continuity of street wall. These two features are not necessarily more influential on travel behaviour than the other features mentioned above. They are just used as the starting points for the analysis on the street facade, which can be extended to include other features in the future. The definition and justification for the two selected features are discussed below.

4.1.1 Building level: Construction and maintenance quality of the building facade (facade quality)

The term 'construction and maintenance quality' is more commonly used in the context of engineering (Atkinson, 2003, p. 4; Brandt & Rasmussen, 2002). In this analysis, I shift the focus of this term away from the engineering domain and emphasise the specific elements that would affect the final appearance of the street facade. The construction- and maintenance-related elements that contribute to the appearance of street facade include

- **Building material:** whether the materials used are of high quality and fine textured;
- **Industrial precision and craftsmanship:** whether the facade is carefully constructed with high level of industrial precision and craftsmanship;
- **Maintenance:** whether the facade is free from cracks, bulges, broken components, deterioration, corrosion, dirt and stain, hanging wires, messy add-ons, etc.

Although this quality seems to be technical oriented, its impacts are not limited to the technical realm. In the book 'Sense of Beauty', Santayana (1955, p. 51) wrote highly of the aesthetic importance of material, saying that 'the beauty of material is thus the ground work of all higher beauty'. Leading modern architects such as Walter Gropius, Le Corbusier and Mies van der Rohe were inspired by what they saw as the great beauty of technical perfection (Voordt & Wegen, 2005). The famous saying of 'God is in the details' is also a reminder of the importance of technical perfection on the overall architectural quality. Dilapidation in the environment has been found to be related to negative affect frequently, and there is no compelling reason to expect different results (Nasar, 1983).

Furthermore, the facade quality also has obvious social effects. According to the famous theory of 'broken window' on urban appearance and social effects, neighbourhood appearances drive the reality of neighbourhood safety: one broken window leads to another broken window and, in turn, to future crimes (Quercia, O'Hare, & Cramer, 2014). In less extreme situations, deterioration in the physical environment may not necessarily lead to crime but may very possibly affect the image and identity of a place (Said, Zubir, & Rahmat, 2014) and the economic development potential available to it. Therefore, modelling results can not only help understand the physical conditions of the urban space but also help identify areas vulnerable to social disorder and economic deprivation.

4.1.2 Street level: Continuity of the street wall (facade continuity)

The street wall refers to the interface formed by street facade along a street. A continuous street wall is formed when buildings stand directly on the edges of their parcels (Lehnerer, 2009, p. 28). To be more specific, a continuous street wall requires the following:

- No 'dead spaces' between buildings, which include vacant lots, parking lots, drive ways or setbacks of a large building (Ewing & Handy, 2009)
- No solid and blank wall blocking the sight and activities from the street to the buildings, specifically in the context of China where most residences and work compounds are gated and surrounded by walls. However, if the wall itself is

carefully designed and visually attractive, it may also be perceived as a continuous flow of the street interface.

Psychically, a continuous street wall offers ‘a sense of enclosure’ (Ewing et al., 2013) and ‘majesty and controlled uniformity’ (Lyon, 1978). It positively affects the experience of urban space by ‘giving a psychical security’ (Lang, 1994, p. 324), ‘instilling a sense of position, of identity with the surroundings’ and ‘embodying the idea of hereness’ (Cullen, 1961, p. 29). Behaviourally, it draws pedestrians and activities and ‘sustains a vital urban district’ (Marcus & Francis, 1997, p. 19), and hence, it is considered to be one of the key rules for place making (Bain, Gray, & Rodgers, 2012, p. 7), which is the development of a built environment in which people want to live and relates to larger goals of creating sustainable communities (Rogerson, Sadler, Wong, & Green, 2010).

As early as the 15th century, relevant rules had appeared in street design codes in Nuremberg, Germany, which required buildings to be lined up to create an ‘undeviating building line’ (Kostof, 1999). Presently, it is addressed in numerous planning codes and guidelines, e.g. the American Planning Association (APA) Planning and Urban Design Standards requires infill projects to ‘maintain ground floor facade to define a consistent street edge’ (American Planning Association, 2006).

4.2 Data and methodology

4.2.1 Data and framework

Street view images are used as the data source on the appearance of street facade, which are provided by Baidu Map, the Chinese equivalent of Google Map. The images were requested at an interval of 200 m along all the streets in the city in February 2016, resulting in 360,796 images (800*500 pixels). Different from most existing studies that focused on the entire streetscape and used images taken with the camera facing the street, I emphasised more on the street facade and set the camera facing the buildings so that the buildings cover a larger proportion of the image (see **Figure 4-1**). However, approximately 30% of the images are still streetscape images, which are taken around street corners or entrances. Therefore, a machine learning model was developed to discern streetscape images from building images to screen out unqualified images.

I followed a two-step approach to develop the machine learning models and three models were developed in total. In the first step, I randomly sampled 3,500 images from the database and manually labelled them as ‘building images’ (2575) and ‘street images’ (925) as shown in **Figure 4-1**. These images were then used to train a ‘qualification’ model to decide whether the content of an image is appropriate to be included in the analysis. In the next step, the qualified ‘building images’ were labelled through expert rating on the two qualities. The two scores were then fed to develop the models of facade quality and continuity. I then applied the two models on all the qualified images from the entire study area.

Figure 4-1 Camera facing the street (left) and facing the buildings (right)

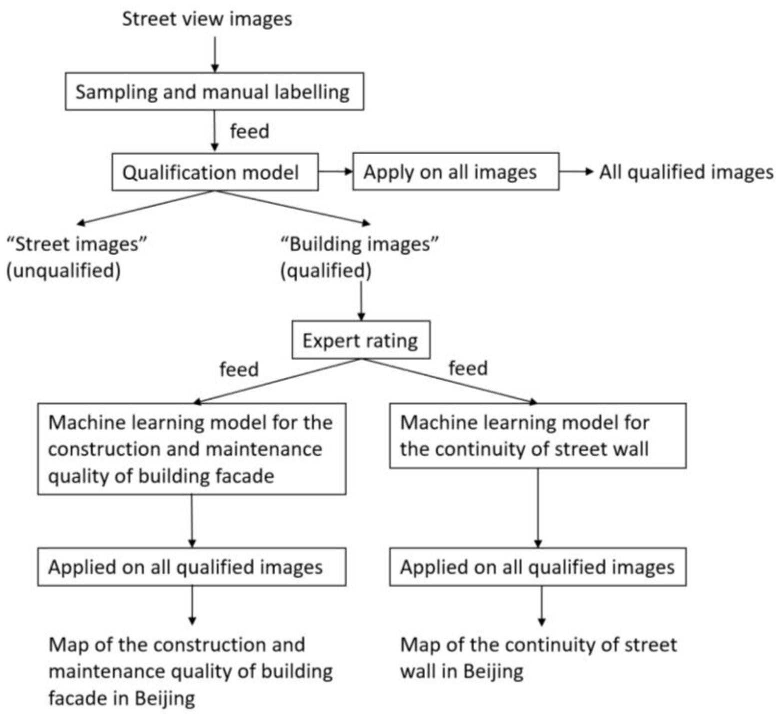

Figure 4-2 Work flow diagram

4.2.2 Expert rating

Expert ratings have been frequently employed in research that involves the measurement of qualities of the urban environment (Nasar, 1983; Wohlwill, 1976, p. 61). The judgemental approach is considered a simple way of measuring the qualities here (Nasar, 1983). Despite the element of subjectivity in the rating scale and the categorisation methods that are relied upon in this approach, the reliability of the resulting values has generally been found to be acceptable and, in some cases, quite high (Wohlwill, 1976, p. 63).

Ideally, experienced experts in the field should be invited to make judgements. However, given the size of the task in this research (each expert needs to rate several hundreds of images), it was difficult to invite experienced architects, urban designers or scholars to do this job. Therefore, I chose to recruit eight graduate students who have received architectural training for more than five years to accomplish this task. Although the validity of expert rating is supported by the virtue of their specialised expertise (Ewing

& Handy, 2009) and there is usually little reason to expect that their assessments would differ systematically from other such professionals (Nasar, 1983), I also took extra measures to reduce potential bias as much as possible. First, I held a training and discussion session with the recruited students to make an agreement on the rating standard for each quality (**Table 4-1** and **Table 4-2**), which linked the judgement of the qualities with more concrete features. Second, I held a practice session in which all students rated a same sample group of images until in most cases they made same judgments.

Table 4-1 Rating standard for facade quality

Ratings	**Rating standard**
Three points	Built with high quality, fine-textured materials; Built with high industrial precision or fine craftsmanship, e.g. building components and material pieces are well aligned, small gaps between material pieces unless they seem to be designed wide, etc.; Well maintained without obvious cracks, breakage, corrosion, dirt and stain or messy add-ons such as rusty iron rails on windows, hanging/loose wires
Two points	Built with lower quality, not very fine-textured materials; Do not show high level of industrial precision or craftsmanship, e.g. material pieces may not be well aligned and may have wide gaps in between; May have a few obvious cracks, breakage, corrosion, dirt and stain or messy add-ons but generally present a neat and clean look
One point	Built with low quality, not very fine-textured materials; Built with low-level industrial precision or craftsmanship; Show a lot of cracks, breakage, corrosion, dirt and stain or messy add-ons
Zero point	Built with low-quality materials, in many cases, bare cement and colour plate[4]; Built with low-level industrial precision or craftsmanship, sometimes seem unfinished; Seriously deteriorated with a lot of cracks, breakage, corrosion, dirt and stain or messy add-ons

Table 4-2 Rating standard for facade continuity

[4] I do not mean that these two materials are in themselves of low quality, but they are often used in low-quality buildings in Beijing.

Ratings	Rating standard
Continuous	Street facades progress through the image without any interruption, blockage or significant setback, at least at the eye height.
Discontinuous	There is a wide gap between two adjacent buildings. There is a significant setback of a wide building. There is a solid wall blocking the building from the street; however, if the wall is carefully designed and visually attractive, it can be considered continuous.

Figure 4-3 Rating examples

4.2.3 Machine learning

In the field of computer vision, there are many approaches for image representation. For this work, I evaluated three features: the conventional SIFT histogram (Lowe, 1999) and two state-of-the-art deep convolutional networks, namely AlexNet (Krizhevsky, Sutskever, & Hinton, 2012) and GoogLeNet (Szegedy et al., 2015). AlexNet and GoogLeNet outperformed all other features in the 2012 and 2014 ImageNet Large Scale Visual Recognition Competition, respectively. Compared with conventional image techniques, which are dominated by low-level features such as edges and corners, the deep convolutional networks can capture both local- and high-level image characteristics. I used the output of the last hidden layer of the two pre-trained neural networks and trained a SVR (Support Vector Regression) classifier for each of the scene attributes.

The labelled data set was randomly sampled into three subsets: the training set, the development set and the test set. For each task, the development set and the test set were equally and randomly sampled in each labelled class, and the rest of the images were used as the training set. For example, for the visual quality task, forty images were randomly sampled in each of the four scoring groups for the development set and sixty images each for the test set. The hyper parameters of SVM, namely the regularisation constant and the regression epsilon width, were optimised through grid searching on the development set. In terms of the evaluation of model performance, I used F1 score for the classification models (the qualification model and the continuity model) and MSE for the facade quality model, which were calculated using the following equations:

$$Recall = \frac{TP}{P}$$

$$Precision = \frac{TP}{TP + FP}$$

$$F1 = \frac{2TP}{2TP + FN + FP} = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

$$MSE = \frac{1}{n} \sum (y_i - t_i)^2$$

where P (positive), TP (true positive), FP (false positive) and FN (false negative) denote the number of the images that are qualified/continuous, both labelled and predicted to be qualified/continuous, labelled unqualified/discontinuous but predicted to be qualified/continuous and labelled true but predicted to be false, respectively, and y_i and t_i denote the machine rating and expert rating for each image, respectively.

The models with the best performance for the three tasks were chosen to be applied to the entire image database of the research area. I then calculated the average scores for each street segment and aggregated the results in the spatial unit of TAZs.

4.3 Results

4.3.1 Results of expert rating

In terms of the facade quality, the expert rating returned 485 three-point images (18.8%), 1079 two-point images (41.9%), 809 one-point images (31.4%), and 202 zero-point images (7.8%). In terms of facade continuity, the expert rating identified 1069 ‘continuous’ images (41.5%) and 1506 ‘discontinuous’ images (58.5%).

Table 4-3 Distribution of expert rating

Rating criteria	Proportion%
Qualification	
Qualified	73.6
Unqualified	26.4
Total	*100*
Facade quality	
3 points	18.8
2 points	41.9
1 point	31.4
0 point	7.8
Total	*100*
Facade continuity	
Continuous	41.5
Discontinuous	58.5
Total	*100*

4.3.2 Machine learning performance

Table 4-4 shows the performance of the SIFT, AlexNet and GoogLeNet features on the test set of the qualification task. The deep convolutional networks, AlexNet and GoogLeNet, perform better than the traditional SIFT features. GoogLeNet achieves a slightly higher F1 score than AlexNet, which indicates a more balanced performance between recall and precision. **Table 4-5** and **Table 4-6** show the performance on the other two tasks. Similar to the task of qualification, deep features outperform the SIFT features. GoogLeNet shows the best capability of generalisation with the lowest MSE on the development set on the task of facade quality. GoogLeNet and AlexNet show almost the same levels of capability on the task of continuity. Based on these results, I chose the GoogLeNet model for large-scale application.

Table 4-4 Performance of the qualification model

	Accuracy (%)	**Precision (%)**	**Recall (%)**	**F1 (%)**
SIFTHist + SVR	79.2	45.1	71.3	55.2
AlexNet + SVR	89.3	**48.2**	85.9	**61.8**
GoogLeNet + SVR	**90.0**	48.1	**86.3**	**61.8**

Table 4-5 Performance of the model on the facade quality (MSE)

	Training set	**Development set**	**Test set**
SIFTHist + SVR	0.36	0.84	0.84
AlexNet + SVR	**0.22**	0.64	**0.62**
GoogLeNet + SVR	0.28	**0.61**	0.64

Table 4-6 Performance of the model on facade continuity

	Accuracy%	**Precision%**	**Recall%**	**F1%**
SIFTHist + SVR	72.0	45.0	72.0	55.4
AlexNet + SVR	**75.0**	**48.0**	**72.0**	**57.6**
GoogLeNet + SVR	**75.0**	**48.0**	**72.0**	**57.6**

To better estimate the capability of the models, I took a closer look at the machine rating results on the test sets and compared with the expert rating scores. **Figure 4-4** shows that the machine scores generally fall into a narrower range than the expert rating scores (average score of ‘zero-point’ images = 2.0, of ‘one-point’ images = 2.3, of ‘two-point’ images = 2.9, and of ‘three-point’ images = 3.4). There are a number of overlaps between the machine scores within the groups of low-quality street facades (the ‘zero-point’ and ‘one-point’ images) and high-quality facades (the ‘two-point’ and ‘three-points’ images). However, there is little overlap between these two big groups (the lower quartile of the machine scores for ‘two-point’ images is higher than the higher quartile of the machine scores for the ‘one-point’ images). The results indicate that the model performs well in discriminating high quality from low quality but that it tends to produce more errors in identifying the nuanced differences within the two big groups.

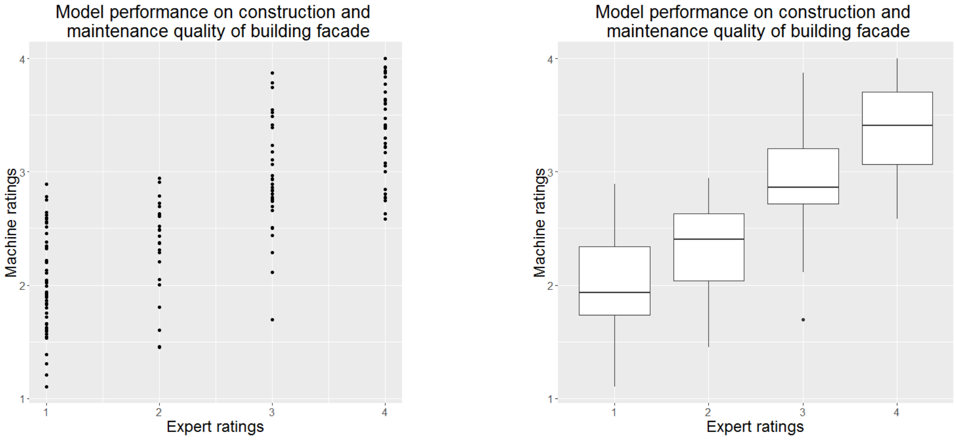

Figure 4-4 Comparison between machine scores and expert rating scores on the task of facade quality

Regarding the task of continuity, I manually analysed sixty images that were wrongly classified (false positive or false negative). Two major types of errors for false positive and three major types of errors for false negative were identified. For false positive, the major types of error are failing to identify an unattractive wall that disrupts the continuity (12%) and failing to identify the gap between buildings because of perspective (76%). For false negative, the major types of error are failing to identify dilapidated buildings as a building (20%), failing to identify a continuous street wall because of blockage by trees and cars (30%) and failing to identify a continuous street wall when the picture is taken from a distance (35%), usually from the opposite side of a wide street. These errors are mainly because of the lack of labelled data to train the model to be aware of relevant situations. Although the total number of labelled images is more than two thousand, when it comes to a very specific type of situation, the relevant sample size could be less than fifty. Therefore, the model performance may be further enhanced by collecting more labelled data.

FP Error Type 1: failing to identify an unattractive wall which blocks the sight

FP Error Type 2: failing to identify the gap between buildings because of perspective

FN Error Type 1: failing to identify dilapidated buildings as a building

FN Error Type 2: failing to identify continuous street wall because of the blockage of trees and cars

FN Error Type 3: failing to identify continuous street wall when the picture was taken from a distance

Figure 4-5 Examples of errors in the task of facade continuity

4.3.3 Evaluation results

By calculating the average score for each street segment, I developed the scoring maps for the two street facade features (**Figure 4-6** and **Figure 4-7**). It should be noted that although the indicators of the model fit indicates a generally good performance, the scores should not be considered absolutely accurate but as estimations with errors. For instance, the red coloured street segments may not always be of higher quality than the orange ones, but these are in most cases of higher quality than the blue ones. The scores are then aggregated to the spatial unit of TAZs.

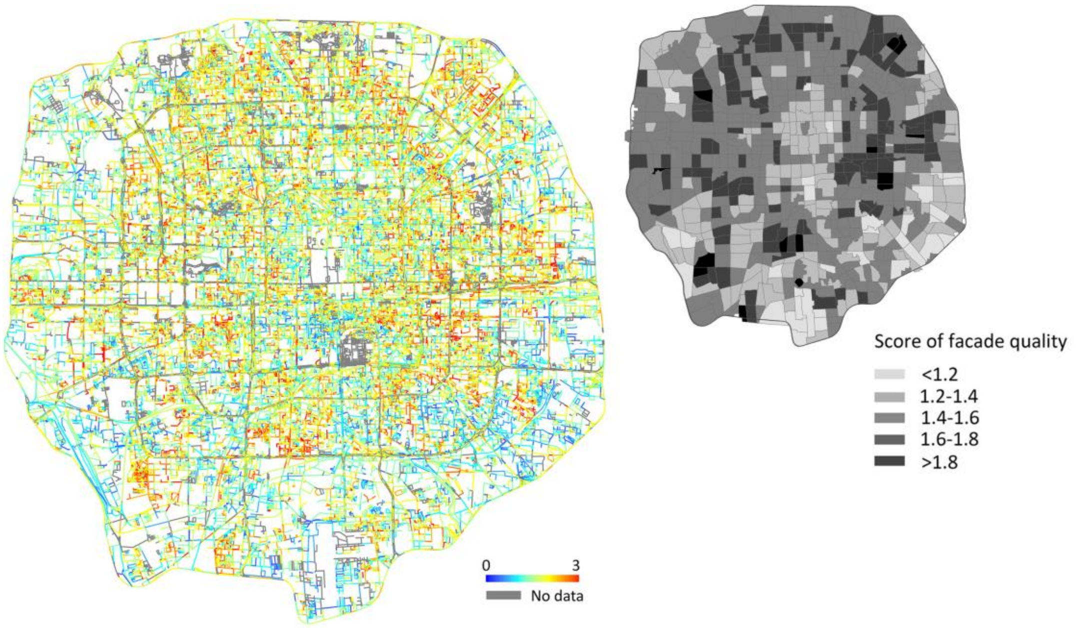

Figure 4-6 Scores on facade quality

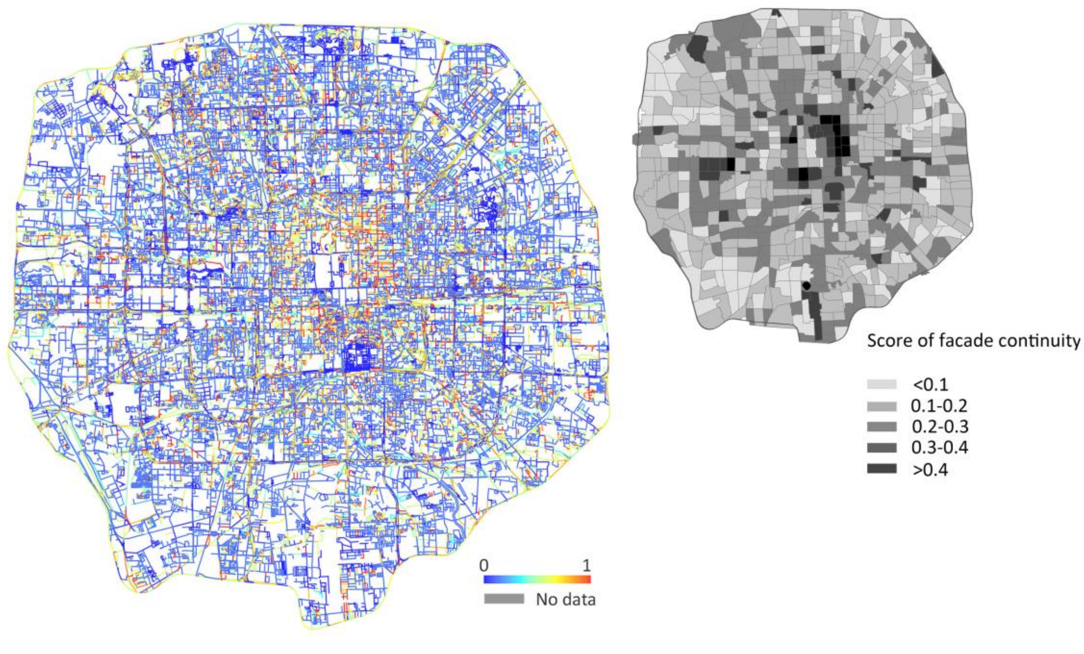

Figure 4-7 Scores on facade continuity

4.4 Chapter summary

This chapter sets out to develop and test a machine learning method to automatically evaluate two street facade features in a large scale, which creates two input variables for the activity-based model developed in the next section. This work aims to fill in the gap that the built environment features that have been examined in existing research are constrained to two-dimensional, land use-related ones, while the features related to the 'third dimension'—the street facade, can also exert an influence on daily travel. Two features are selected as the starting point of the analysis on the street facade: the construction and maintenance quality of building facade and the continuity of street wall. The method can be further extended to evaluate other street facade features that could be influential to the experience in the urban space, such as the building scale, the relationship between adjacent buildings, etc. By applying the state-of-the-art deep convolutional networks, a satisfying performance of the machine learning models can be achieved on the expert-rated data sets. The MSE for the task of facade quality is 0.61 on a rating scale of zero to three, and the accuracy for the continuity task is 75%.

Chapter 05 Activity-based modelling on the impacts of the built environment on travel behaviour

5.1 Overview of the model

As mentioned in the introduction chapter, this research develops an activity-based travel model that comprehensively incorporates the influence of the built environment conditions in the decision making process of daily activity-travel (named as Built Environment Activity-Travel Integrated Model, BEATIM). The development of BEATIM model will on one hand, help address the gap that there lacks understanding on the composition of the built environment's influence on travel behavior (e.g. the influences on travel frequency, distance of travel for a specific purpose, etc.), and on the other hand, help fill in the gap that built environment conditions are usually not sufficiently accounted for in existing activity-based models. The travel behaviour and built environment data developed in Chapter 03 and 04 will be fed into the model as inputs. The model will then be applied in the next chapter to simulate the impacts of various scenarios of built environment changes on people's travel behaviour, from which analysis on the composition of the built environment's influence can be made and various policy implications can be drawn.

5.1.1 The modelling paradigm

Basically, activity-based models predict which activities are conducted when, where, for how long, with whom, and the transport mode involved (Arentze & Timmermans, 2004; Castiglione et al., 2015; Ma et al., 2012). The strength of the BEATIM model lies in the intensive incorporation of the built environment conditions in the decision making process of daily activity-travel. The advantage of this incorporation is, on one hand, to enhance the comprehensiveness and behavioural realism of activity-travel modelling, and on the other hand, to enable a more detailed and behavior-oriented analysis of the influence of the built environment on travel.

In order to realise the promised benefits, a proper modeling paradigm needs to be selected. As mentioned in the literature review (Section 1.3), activity-based models typically fall into one of two categories: utility-maximising econometric models and computational process models. The former is implemented with a system of equations based on the assumption of utility maximisation in travel decision making (Bhat et al., 2004; Yasmin et al., 2015). The latter, on the other hand, is a production system that involves a series of condition-action rules (Gärling et al., 1994; Shabanpour et al., 2017). The computational process models have special strength in incorporating the idea of incomplete information and imperfect rationality, and modelling the learning process, which are considered to be cognitively more realistic (Arentze & Timmermans, 2004; Auld & Mohammadian, 2012). However, it should be noted again that this categorisation is neither exclusive nor exhaustive. Many utility maximisation-based models can also be considered as incorporating weak computational process features in the form of some sequential or partially sequential decision making process (Arentze & Timmermans, 2004).

The BEATIM model developed in this research generally falls in the category of utility maximisation models, and with weak computational process features that are reflected in the decision sequence and a few action rules. The utility maximisation paradigm is chosen for its strength in allowing the examination of alternative hypotheses regarding the causal (or correlational) relationships between activity-travel patterns, the built environment and socio-demographic characteristics of individuals (Bhat et al., 2004). Besides, the simulation results can be directly linked with the assumptions in Section 2.1.3 which are based on the same notion of utility maximisation.

The two main criticisms of the utility-maximising approach are that: (1) individuals are not necessarily fully rational utility maximisers, and (2) the approach does not explicitly model the underlying decision processes and the behavioral mechanisms that lead to the observed activity-travel decisions (Pinjari & Bhat, 2011). Nonetheless, some counterarguments can be made. First, most of the behaviours being modelled happen repeatedly in daily life (such as commuting, shopping, but activities like meeting a friend, going to hospital can be occasional). People may have already searched and compared many alternatives and optimised their choices before the observation/survey was conducted (though this may not be the case for occasional activities). For instance, Auld found that the mode and location choices tend to be quite routine, though the start time and especially the duration decisions tend to be impulsive (Auld & Mohammadian, 2012). Besides, the proliferation of information and communication technologies enhances the availability of the information in real time traffic, the locations and qualities of facilities, etc. (Kitchin, 2014), so the assumption of complete information is more likely to be approached. Although the above statements are subject to further proof, they are logically reasonable and suggest that utility maximisation can be taken as a feasible assumption for activity-based travel modelling.

The computational process elements of the model are reflected in a sequence of decision making and action rules. Like most existing models, BEATIM makes simplified assumptions about the scheduling process by using a fixed planning order for specifying the activity attributes (time, location, mode of travel, etc.) (Auld & Mohammadian, 2012). Actually, there has been a trend in activity-based modelling to introduce higher level of flexibility to the model framework to account for short-term adjustments and rescheduling processes since some activities can be opportunistically planned (e.g. the AURORA model, ADAPTS model) (Auld & Mohammadian, 2012). Such features are not incorporated in the current model due to the constraints of time and resource. Nonetheless, it could be a direction of subsequent research to incorporate a dynamic activity planning framework to the model when more data and resources are available. Besides, the incorporation of decision mechanisms under uncertainty and with imperfect rationality could also be a direction of model extension (Rasouli & Timmermans, 2014b).

5.1.2 The focus

"Entities should not be multiplied unnecessarily."

The quote above corresponds to the principle of Occam's razor by William of Occam in the 14^{th} century (Rasmussen & Ghahramani, 2001) and is nowadays often referred to as the law of parsimony that implies when two theories or models are equal, the burden of proof rests with the more complicated model to show it is able to make better predictions (Kelly, 2013). Although Occam's razor is not considered as an irrefutable principle and may mistakenly lead to oversimplification (Kelly, 2013), it is useful to guide the modelling process.

The activity-travel and related processes generally fall into the land use-transport interaction (LUTI) system (Acheampong & Silva, 2015). This system, by nature, involves complicated dynamics between system components, ranging from land use and building stock changes that are slow and hardly reversible, to activity and travel decisions that are updated on a daily basis, to route choices that are subject to instant adjustment according to the conditions of traffic flow (Simmonds, Waddell, & Wegener, 2013) (**Table 5-1**). When building models within the LUTI system, one could be tempted to include more and more components of the system to fully account for various mechanisms of interactions. For instance, one can easily make an argument that there needs to be a module for home location choice since people may consider move their homes if the travel distances for fulfilling their daily needs are longer than certain thresholds. However, the pursuit of comprehensiveness can distract the modelling efforts away from the key research questions. Therefore, in building this model, the focus on the daily travel behaviour, the ‘very fast’ type of process in the LUTI system (see **Table 5-1**), is carefully kept—other loosely-coupled, slower or faster processes, such as the location choices of home and work places, are taken as exogenous and constant in the simulation period. These modules can be added to the model system in future extensions.

Table 5-1 Components of the LUTI system

Speed	Change process	Stock affected	Response time (years)	Response duration (years)	Reversibility
Very slow	Transport construction	Transport networks	5-10	>100	Hardly reversible
	Land use change	Land use pattern	5-10	>100	Hardly reversible
Slow	Industrial construction	Industrial buildings	3-5	50-100	Very low
	Residential construction	Residential buildings	2-3	60-80	Low
Medium speed	Economic change	Employment/firms	2-5	10-20	Reversible
	Demographic change	Population/ households	0-70	0-70	Partly reversible
Fast	Firm relocation	Workplace occupancy	<1	5-10	Reversible
	Residential mobility	Housing occupancy	<1	5-10	Reversible
Very fast	Change in demand	Goods transport	<1	<5	Reversible
	Change in mobility	**Person travel**	**<1**	**<1**	**Reversible**

Source: (Simmonds et al., 2013)
Note: Bold indicates the focus of the BEATIM model.

The focus of the model is also reflected in the treatment of various choice facets. Since the main purpose of the BEATIM model lies in the examination of the built environment's impacts on activity-travel behaviour, the choice facets that are prominently influenced by the built environment are simulated through statistical models (e.g. number of activities, locations of activities, mode of travel), while the other choice facets are simulated in a simplified manner with observed probability distributions (e.g. the types of non-commute activities conducted in the day, the arrangement of activities, see **Table C-2** to **Table C-4** in Appendix C).

5.1.3 The decision makers in the model

The BEATIM model is a microsimulation model in which persons are represented explicitly and are capable of perceiving their environment, making decisions and acting into their environment. The representative individuals in the model possess some features of autonomous agents, but are not the same as typical agents in an agent-based model. Below is a comparison of the decision makers in the BEATIM model and typical agents, based on a summary by Castle and Crooks (2006).

- **Autonomy:** yes, the individuals in the model are autonomous units without any centralised control.
- **Heterogeneity:** yes, the individuals are highly heterogeneous in terms of the socioeconomic profile. They also face highly heterogeneous built environment conditions when making travel decisions. However, the model does not account for heterogeneous behavioural rules at the current stage.
- **Pro-active/goal-directed:** yes, the individuals are utility maximisers.
- **Reactive/perceptive:** weakly, the individuals are assumed to have a limited ‘awareness’ of their environment and choose locations of activities only from a random subset of all available alternatives (see Section 5.3 and Section 5.5).
- **Bounded rationality:** weakly, the limited ‘awareness’ of the individuals on their environment leads to bounded rationality when choosing activity locations.
- **Interactive/communicative:** no, the individuals do not directly interact or change information with each other. It is also based on the consideration that the focus of the model lies in the impacts of the built environment, so the interpersonal interactions are not specifically modelled, such as the task allocation among household members.

- **Mobility:** no, although the model deals with travel behaviour, the home location, where the individuals are affiliated to, does not change. In this sense, the individuals are immobile in the model space.
- **Adaptation/learning**: weakly, the adaptation/learning is reflected in the mechanisms that individuals would expand their knowledge about the environment if the highest utility derived from their origin knowledge is below certain a threshold, and that they can also adjust the time of activities if any mode choice in the originally selected time period does not bring a satisfying utility. However, these mechanisms do not turn out to be influential to the modelling results (see Section 5.3 to Section 5.5).

5.1.4 The modelling structure

Activity-based models predict activity-travel behaviour as a 'full day pattern' that involves an entire chain of trips made between the first time of leaving home in the day and finally arriving back at home (Cambridge Systematics, 2002; Liu et al., 2014). The full day pattern is composed of primary activities and intermediate stops, and the tours and trips that link them up. The definitions of these components, which will be used throughout this chapter, are explained below:

- Primary activities are defined as those with the longest duration among all activities conducted in a tour;
- Intermediate stops refer to the rest of activities with shorter durations;
- Trips are one-way movements from one location of staying to the next;
- Tours are chains of linked trips that start from and end at home, the use of which is considered to be a key feature of activity-based model systems in contrast to the trip-based four-step models (Cambridge Systematics, 2002).

As mentioned in Section 5.1.1, various choice facets of the full day pattern are organised in a decision sequence, which is derived from both the results of the small questionnaire survey and operational considerations. The survey results show that 44%

people chose ‘what shall I do today’ as the first consideration in making activity-travel plans on weekdays, 24% chose ‘where shall I go’ or ‘how far shall I go’, 16% chose ‘shall I go by car/metro/bus’ and 15% chose ‘when shall I go’. The results on weekends are similar with those on weekdays, except that the proportion of people choosing ‘Where shall I go?’ or ‘How far shall I go?’ is higher. This makes sense since people are usually faced with less time constraints on weekends, so that they have more freedom to choose activity locations that best suits their needs. Regarding to the planning of primary destinations and intermediate stops, 60% people chose that they ‘first decide long-stay/primary destinations and then short-stay/intermediate stops’ and 40% chose ‘decide all destinations together’. For weekends, the former option was more chosen (69%). The survey results suggest that: (1) the priorities that people give to these choice facets can be ranked as: the number and type of activities > the location of activities > the time of activities/the mode of travel, and (2) the location choice of primary activities and intermediate stops tend to be separate decisions, the latter dependent on the former.

Table 5-2 Results of the small questionnaire survey

What is your consideration when you make plans about your activities (except work) on weekdays?		
	First consideration	Second consideration
‘What shall I do today?’	44%	14%
‘When shall I go?’	15%	22%
‘Shall I go by car/metro/bus/walk..?’	16%	24%
‘Where shall I go?’ or ‘How far shall I go?’	24%	40%
What is your consideration when you make plans about your activities (except work) on weekends?		
	First consideration	Second consideration
‘What shall I do today?’	45%	13%
‘When shall I go?’	12%	20%
‘Shall I go by car/metro/bus/walk..?’	14%	17%
‘Where shall I go?’ or ‘How far shall I go?’	29%	50%
When deciding about the activity destinations on weekdays, which do you prefer?		
‘Decide all destinations together’	60%	
‘First decide long-stay/primary destinations and then short-stay/intermediate stops’	40%	

When deciding about the activity destinations on weekends, which do you prefer?	
‘Decide all destinations together’	69%
‘First decide long-stay/primary destinations and then short-stay/intermediate stops’	31%

Based on the survey results, the general structure of the model is designed as follows (**Figure 5-1**):

- First, the number and type of activities to conduct in the simulated day are decided and organised into primary destinations and intermediate stops;
- Second, the locations of primary destinations are selected;
- Third, the times of activities (in terms of time periods) and modes of tours are decided;
- Last, the locations of intermediate stops are selected.

These four decision components are simulated in the four sub-models described in Section 5.2 to Section 5.5. It should be noted that although a survey was conducted to inform the sequencing of decisions, this consequent model structure should not be taken as the only way of model construction. Actually, most activity-based models use more or less different sequences of decision making and there is no clear evidence that one sequence outperforms another (see for instance, the SACSIM ‘family’ of models mentioned in Section 2.4).

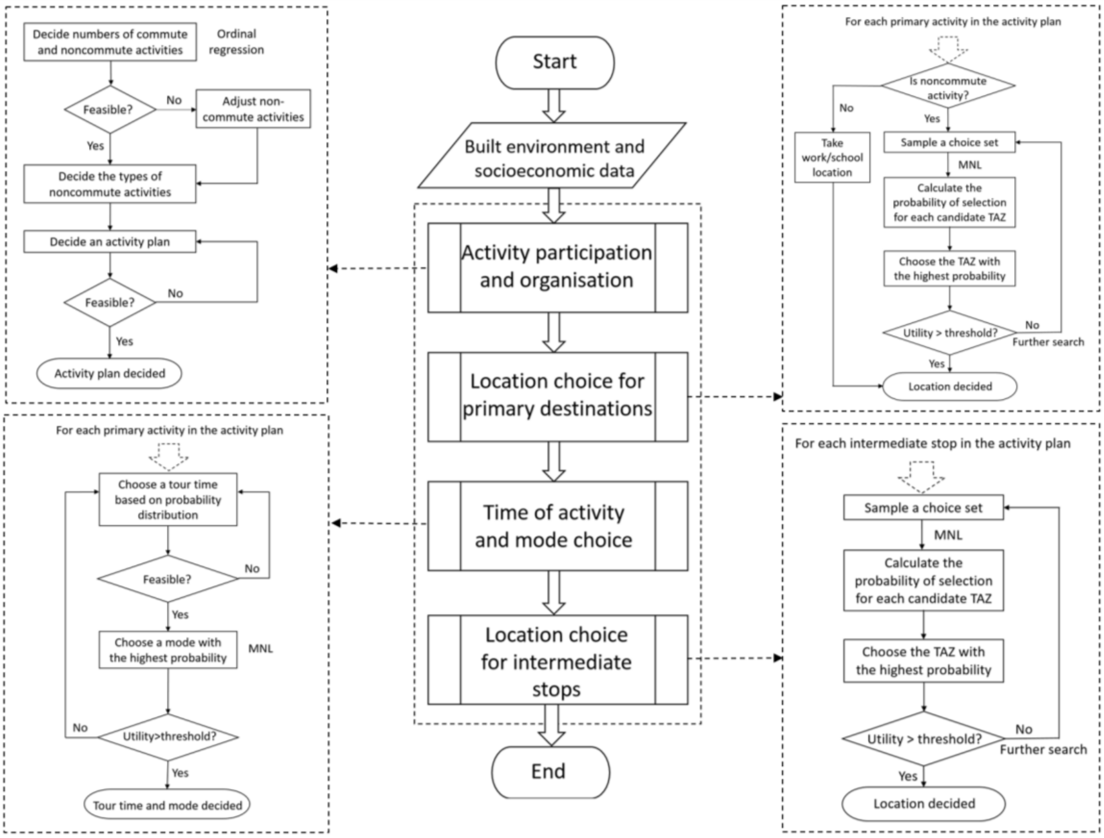

Figure 5-1 Model flow diagram

Note: The detailed flow diagram of each sub-model will be zoomed in in subsequent sections.

A total of 300,000 individual entities are simulated in the model, which is approximately 2% of the population in Beijing. The distributions of gender, age, employment status, household type and social status are kept in consistency with the travel survey. Other elements of the model include (see **Table 5-3**):

- Types of activities: two types of commute activities, which are work and go to school; six types of non-commute activities, which are shopping, entertainment, dining out, personal business, escorting/picking up/dropping off other people and others. The first five types of non-commute activities are modelled separately since they took up more than 2% of all activities recorded in the travel diary survey. The other activity types in the travel survey are categorised as 'others' (see **Table C-1** in Appendix C).
- Numbers of activities: upper limit three activities per day. Since according to the

travel survey, 95% people conducted no more than three activities in a day.

- Activity plans: eleven most common plans from the travel survey.
- Time of activity: six time slots.
- Locations: 652 TAZs in the study area.
- Modes of travel: four most popular travel modes identified from the survey, which together accounted for 96% of all trips.

Table 5-3 Elements of the model

Choice facets	Alternatives
Types of activities	Commute: work, go to school Non-commute: shopping, entertainment, dining out, personal business, escorting/picking up/dropping off other people and others
Numbers of activities	0-3
Activity plans [a]	h-d-h, h-d-h-d-h, h-s-d-h, h-d-s-h, h-d-h-d-h-d-h, h-d-s-s-h, h-s-d-s-h, h-s-s-d-h, h-s-d-h-d-h, h-d-s-h-d-h, h-d-h-d-s-h
Time of activity	early (3-7am), am peak (7-9 am), before noon (9 am-12 pm), afternoon (12-17 pm), pm peak (17-19 pm), evening (19 pm-3 am)
Locations	652 TAZs (for each activity, a distance-weighted subset of 10 TAZs)
Modes of travel	driving, public transit, cycling, walking

a 'h' denotes home, 'd' denotes a primary destination, 's' denotes an intermediate stop. 'h-d-h', for instance, means travelling from home to a primary destination and then back to home.

Table 5-4 Information delivery among sub-models

	Sub-model 1	Sub-model 2	Sub-model 3	Sub-model 4
Numbers of activities	O	I	I	
Types of activities	O	I		
Activity plans	O	I	I	I
Locations of primary destinations		O	I	I
Time of activity			O	
Modes of travel			O	I
Locations of intermediate stops				O

Note: I = input, O = output

There are basically two types of parameters in the model. The first type are parameters that reflect the impacts or weights of socioeconomic, built environment and travel-related factors, which are estimated from statistical regressions (i.e. ordered regression, multinomial logit model) or derived directly from the statistical distribution of the observed data. The second type are constants or threshold values which are calibrated in the model through parameter sweep (Castle & Crooks, 2006). The parameters are listed in **Table 5-5**. Eighty percent of the samples in the travel diary survey are used as the training set and the rest twenty percent are used as the test set.

Table 5-5 Parameters in the model

Sub-models	**Parameters estimated from regressions or statistical distributions**	**Parameters calibrated in the model**
Sub-model 1	Impacts of socioeconomic and built environment variables on the number of commute and non-commute activities in the day; Probabilities of choosing different types of non-commute activities; Probabilities of choosing an activity plan given the total number of activities	Cut-off values for the ordered regression models
Sub-model 2	Weights given to different distance bands when drawing a random sample of candidate activity locations; Impacts of socioeconomic, built environment and distance variables on the attractiveness of a location	Extra weights of different distance bands on the attractiveness of a location
Sub-model 3	Probability of choosing a certain time slot for an activity given the purpose and the day activity plan; Impacts of socioeconomic, built environment and travel variables on the attractiveness of a travel mode	Constants for different travel modes
Sub-model 4	Weights given to different detour distance bands when drawing a random sample of candidate stop locations; Impacts of socioeconomic, built environment and travel variables on the attractiveness of a location	Extra weights of different detour distance bands on the attractiveness of a location

5.2 Sub-model 1: Activity participation and organisation

5.2.1 Introduction to the sub-model

The making of an activity plan is a multifaceted decision, which involves deciding the number of commute and non-commute activities, the purposes of non-commute activities (if any) and the organisation of activities. Each of the choice facets can potentially be influenced by the built environment. As mentioned in the Chapter 2, enhanced accessibility could lead to more activities since people may use travel time saved from better accessibility of one activity on participating more (J. Lin & Yang, 2009; Maat & Timmermans, 2009; Maat et al., 2005; Sperry, Burris, & Dumbaugh, 2012). The influence can be further complicated by the possibility of trip chaining behaviour (i.e. organise an activity into an existing tour as an intermediate stop). On one hand, residents of low accessibility areas often compensate for the long distance by buying daily necessities along the route home from work (Naess, 2013). On the other hand, high accessibility may also encourage trip chaining by offering more activity opportunities in adjacent areas. Besides, the built environment can also exert an influence on the types of non-commute activities that one conducts. For instance, Chudyk et al. (2015) found that living in neighbourhoods with a greater prevalence of destinations was associated with making more trips (Chudyk et al., 2015)..

Therefore, statistical models were estimated for all these three choice facets. For the numbers of commute and non-commute activities, since the outcomes are ordinal, discrete data, two ordinal regression models were developed. In terms of the purposes of non-commute activities, six binary logistic models were estimated for whether a type of non-commute activity is conducted or not, when there is any non-commute activity in the day. For the organisation of activities, two multinomial logit models were built to estimate which activity plan would be chosen given the total number of activities in the day (2 or 3, if there is only one activity, then the activity plan can only be ‘h-d-h’).

It turns out that built environment features are only occasionally significantly correlated with the types of non-commute activities and the choice of activity plan. Besides, to include built environment features in the models of these two choice facets hardly improves the prediction performance (see **Table C-2** to **Table C-4** in Appendix C). Therefore, following the principle of parsimony, the influence of the built environment is only modelled at the level of activity numbers. The other two choice facets are instead simulated in a simplified manner based on the probability distributions observed from the survey data.

The computational process of this sub-model is designed as:

- First, an individual decides about the numbers of commute (if he/she is a worker or student) and non-commute activities based on ordinal regression models, which can be interpreted as an outcome of utility maximisation (Agyemang-Duah & Hall, 1997; Bhat & Pulugurta, 1998).
- Second, the individual checks whether the total number of activities is within the reasonable range (no more than three), as explained before. If not, some non-commute activities will be randomly drawn and removed from the plan.
- Third, the types of non-commute activities (if any) are allocated to the activity plan based on the frequency distribution of non-commute activities in the travel survey (see **Table C-5** in Appendix C).
- Fourth, an activity plan is allocated to the individual based on the probability distributions of activity plans given the total number of activities in the day (see **Table C-6** in Appendix C).
- Last, the individual fits his/her activities into the plan and checks the feasibility of the plan (criteria: commute activities cannot be intermediate stops). If not, this and the previous steps are repeated.

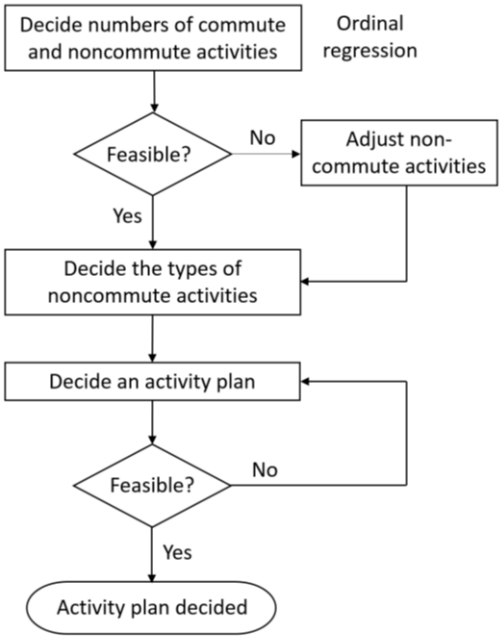

Figure 5-2 Flow diagram for Sub-model 1

5.2.2 Method for parameter estimation

As briefly mentioned before, since the numbers of activities are ordinal, discrete data, ordinal regression models (also known as cumulative link model) are applied to estimate the influencing weights of built environment and socioeconomic factors. The model is for an ordinal response variable, Y_i that can fall in j = 1,…,J categories. Then Y_i follows a multinomial distribution with parameter π where π_{ij} denotes the probability that the *i*th observation falls in the response category j. The cumulative probabilities are defined as (Christensen, 2010):

$$\gamma_{ij} = P(Y_i \le j) = \pi_{i1} + \cdots + \pi_{ij}$$

Use F(x) to denote the link function, then

$$\mathrm{F}(\gamma_{ij}) = \theta_j - x_i^T \beta$$

where x_i is a vector of explanatory variables for the *i*th observation and β is the corresponding set of regression parameters. Note that $x_i^T \beta$ does not contain an intercept, since the $\{\theta_j\}$ parameters provide each cumulative logit (for each j) with its own intercept (Christensen, 2010). Ordinal regression models can both be applied to discrete outcomes of a latent continuous measure and ordered choices, as an outcome of utility maximisation, such as car ownership (Bhat & Pulugurta, 1998) and numbers of trips (Agyemang-Duah & Hall, 1997). In these cases, the discrete outcomes can be interpreted as a reflection of the underlying preference intensity (Greene & Hensher, 2010).

The influencing factors considered in the model include:

- Individual's socioeconomic characteristics, including age, gender, household annual income, household type, household's social well-being, driving license, car ownership, motorcycle ownership, e-bicycle ownership;
- Built environment conditions at home, represented by the fifteen built environment features measured in Chapter 3 and 4;
- Commute distance if the person is a worker or student. This factor reflects the ideas of total time constraints and people's tendency to maximise the entire activity pattern in a day (Axhausen & Garling, 1992; Hägerstraand, 1970; Maat et al., 2005). It is supposed that long commute distance, and correspondingly long commute time, could discourage conducting more activities.

There are five typical link functions for ordered regression models (as shown in **Table C-7** in Appendix C). The link functions that produce the highest log-likelihoods are chosen for the final models. The model is estimated with the 'ordinal' package in R Studio v0.99.473. The threshold parameters are then calibrated to maximise the approximation between the simulated and observed total numbers of activities.

5.2.3 Results of parameter estimation

The log-likelihoods of models estimated with the five link functions are shown in **Table 5-6**. For commute activities, the "cauchit" function produces the highest log-likelihood. For non-commute activities, the "loglog" function produces the highest log-likelihood.

Table 5-6 Log-likelihoods with different link functions

Link function	**logit**	**probit**	**clog-log**	**log-log**	**cauchit**
Commute					
Log-likelihood	-4570.722	-4586.895	-4606.523	-4565.955	**-4502.836**
Non-commute					
Log-likelihood	-22073.33	22382.98	-24398.22	**-21155.41**	-22115.93

The model estimation results show that three socioeconomic variables, three built environment variables and the commute distance significantly affect the generation of commute activities. Older age and motorcycle ownership are positively associated with the number of commute activities. Car ownership is negatively associated the number of commute activities. It is a bit counter-intuitive that people living farther from the city centre tend to conduct more commute activities. However, it can be explained as follows: when the commute distance is controlled, people who have the time to go back home in the middle of work usually have less working pressure and thus earn less, and therefore more likely to live at the outer areas of the city. Other significant built environment variables are the density of secondary roads and the facade continuity, which are negatively and positively associated with the number of commute activities respectively. Besides, longer commute distance decreases the chance of more commuting, as one would expect.

There are more significant influencing factors for the generation of non-commute activities. Being older or having a household annual income of 100-150 thousand RMB and 250-300 thousand RMB are positively associated with the number of non-commute activities, comparing with the reference group of having a household annual income of less than 50 thousand RMB. Being retired, unemployed or in a status other than full-time employed increases the chance of conducting more non-commute activities, while being a student has the reversed effect. People from non-core-family households tend to conduct less non-commute activities, but the relationship is only significant for the household types of ‘couple’ and ‘others’. Holding a driving license and owning a car or an e-bicycle increase the participation in non-commute activities, as one would expect.

Regarding to the built environment characteristics at the home TAZ, higher levels of employment density, land use mix and the accessibility to city centres and commercial clusters are negatively associated with the number of non-commute activities. Some of the effects may be explained by that people can perform multitasks in one move when the accessibility is good, which, however, needs to be verified by further research. Other indicators of higher accessibility or better services are all positively associated with non-commute activities, including retail density, parking density and the accessibility to subway stations. These results are different from the finding of Handy (1993) that trip frequency is irrelevant with either local or regional accessibility. The quality and the continuity of street facade are also positively associated with the frequency of non-commute activities, which is plausible since they may have the effect of increasing the enjoyment of the urban environment. Longer commute distance reduces the chance of conducting more activities, which is not surprising.

The threshold coefficients are calibrated to 1.81 and 49.1 for the commute model, and 1.08, 2.94 and 5.04 for the non-commute model. After calibration, the total numbers of both types of activities are almost identical between the prediction and the observed data. The confusion matrices can be found in **Table C-8** and **Table C-9** in Appendix C.

Table 5-7 Ordinal regression results on activity numbers

	Commute		**Non-commute**	
Variables	**Coefficient**	**z-value**	**Coefficient**	**z-value**
Threshold: 0|1			1.5***	8.36
Threshold: 1|2	3.21**	3.11	3.44***	18.9
Threshold: 2|3	49.1***	8.97	5.04***	27.5
Socioeconomic				
Age	0.0326***	6.7	0.00426***	4.57
Gender (Ref=Female)				
Male	-0.0691	-0.642	-0.0326	-1.56
Annual income (Ref=<50 thousand RMB)				
50-100 thousand RMB	-0.219 Ψ	-1.84	0.0301	1.27
100-150 thousand RMB	-0.161	-0.77	0.0918*	1.99
150-200 thousand RMB	-0.267	-0.693	0.118	1.44
200-250 thousand RMB	0.435	0.983	-0.00813	-0.057

250-300 thousand RMB	-1.65	-0.718	0.432*	2.26
>300 thousand RMB	0.177	0.253	0.0957	0.585
Employment (Ref=Full-time worker)				
Student			-1.2***	-17.2
Unemployed			1.99***	48.1
Retired			2.12***	54.8
Others			0.478***	7.54
Household type (Ref=Core family)				
Single	-0.252	-1.32	-0.0405	-1.07
Couple	0.000412	0.00347	-0.0484 Ψ	-1.91
Multi-generation	-0.375	-1.13	-0.0363	-0.83
Others	-0.107	-0.654	-0.0771*	-2.39
Social well-being (Ref=Best-off)				
Middle	-0.0543	-0.32	0.0447	1.4
Least well-off	-0.102	-0.539	0.0306	0.853
Driving license (Ref=No)				
Yes	-0.205 Ψ	-1.66	0.19***	6.86
Car ownership (Ref=No)				
Yes	-0.489***	-3.43	0.117***	4.4
Motor cycle (Ref=No)				
Yes	0.62*	2.39	0.023	0.308
Electric bicycle (Ref=No)				
Yes	0.0304	0.212	0.0622*	2.08
Built environment				
Population density	3.64E-06	0.486	1.46E-06	1.07
Employment density	-1.07E-05	-1.55	-7.41E-06***	-5.68
Distance to the city centre	8.73E-05***	3.8	1.37E-05**	3.01
Distance to the nearest commercial cluster	-4.14E-05	-0.639	3.33E-05*	2.43
Retail density	-0.00175	-0.73	0.000973*	2.4
Entertainment density	-0.000499	-0.164	0.000148	0.269
Land use mix	0.331	1.49	-0.103*	-2.46
Primary road density	2.04E-05	0.401	-1.53E-05	-1.57
Secondary road density	-9.57E-05*	-2.48	3.77E-06	0.496
Tertiary road density	-1.60E-05	-0.736	2.81E-06	0.732
Parking density	5.66E-05 Ψ	1.7	1.52E-05*	2.26
Distance to the nearest subway station	0.000113	1.26	-6.23E-05**	-3.16
Bus coverage	0.0216	0.0569	-0.0408	-0.546
Facade quality	0.331	0.676	0.314***	3.5
Facade continuity	2.89***	3.83	0.59***	3.88
Travel characteristics				
Commute distance	-0.000483***	-13.9	-0.000106***	-24
	Log-likelihood: -4502.84		Log-likelihood: -	

	AIC: 9079.67	21155.41 AIC: 42394.82

5.2.4 Validation of the sub-model

The validation result shows that the sub-model is able to produce a good estimation of the total number of activities on the test set. The percentage of correct prediction (PCP) is 87.3% for commute activities and 62.4% for non-commute activities. The ratios between the predicted and simulated total numbers of activities are 97.2% on commute activities and 96.8% on non-commute activities. The model is most likely to make mistakes in identifying the people who conduct two commute activities or one non-commute activity in the day. It is tolerable since whether people conduct an activity or not on the day of the survey can contain substantial level of stochasticity (Kulkarni & McNally, 2000). For instance, a shopping tour might be conducted on the day before or after and thus not observed in the survey data. The accuracy of prediction may be improved by collecting multi-day travel data in the future.

Table 5-8 Confusion matrix on the number of commute activities

	1 (sim)	2 (sim)	3 (sim)
1 (obs)	2905	169	0
2 (obs)	244	61	0
3 (obs)	13	3	0

Table 5-9 Confusion matrix on the number of non-commute activities

	0 (sim)	1 (sim)	2 (sim)	3 (sim)
0 (obs)	3271	169	10	0
1 (obs)	592	256	1197	0
2 (obs)	146	76	499	0
3 (obs)	18	19	195	0

5.3 Sub-model 2: Location choice for primary destinations

5.3.1 Introduction to the sub-model

The explicit modelling of activity location choice is a process that substantially enhances the level of detail and behavioural realism comparing with the direct analysis of the synthesised travel outcomes (e.g. VMT) in existing built environment-travel research. Through this module, the processes of utility maximisation and the trade-offs between gains and costs underlying the observed outcomes can be better understood. Besides, a related advantage is the higher level of spatial detail in the examination of the built environment's influence, since one can explicitly model the impacts of built environment changes at any location of the modelled region on the travel of people at any location. In contrast, most existing research only deal with the built environment conditions around one's home or work places. The level of modelling detail is further enhanced by building separate models for different activity purposes, considering that the reaction of the travel behaviour to built environment conditions may vary with travel purposes (e.g. Song, Preston et al. 2013, Salon 2015).

As explained before, medium and long term conditions, such as the locations of jobs and schools, are taken as exogenous and constant in the model developed in this research. Therefore, it is the location choice of non-commute activities that needs to be simulated. However, this part can be computationally challenging since the universal choice set could be very large, consisting of all potential activity locations in the modelled region (in this case, 652 TAZs) (Auld & Mohammadian, 2011). Therefore, appropriate measures need to be designed to limit the number of alternatives when simulating choices (Bowman & Bradley, 2005). In BEATIM, distance is used to weight the probability that a TAZ is drawn into the candidate alternative set. For each choice situation, all TAZs are stratified into different distance bands using the tour origin (home) as the anchor point. A quota is allocated to each distance band to randomly draw a number of alternatives, based on the distance distribution of observed trips in the survey data. With the reduced choice set, the utility and probability of choosing each of the alternatives are estimated with a multinomial logit model, based on a combination of travel impedance (distance/time) and attractiveness variables.

The computational process of this sub-model is designed as:

- First, for each primary activity, a subset of ten TAZs are drawn as candidate locations using the distance weighted sampling method described above. Seven distance bands are considered: shorter than one kilometre from the tour origin, one to two kilometres, two to three kilometres, three to four kilometres, four to five kilometres, five to ten kilometres and farther than ten kilometres. The quota allocated to each distance band are shown in **Table C-10** in Appendix C.
- Second, the probability of each candidate TAZ being chosen is calculated. Since the choice is a discrete event, multinomial logit model (MNL) is employed to estimate the weights of impedance and attractiveness factors. This model is mathematically proved to reflect the utility maximising choice behaviour (McFadden, 1972, 1978).
- Third, the candidate TAZ with the highest selection probability is chosen as the activity location.
- Fourth, the systematic utility component of the chosen location is compared with a threshold value. If the utility is lower than the threshold, step one to three are repeated and a higher utility location is selected, if possible. As mentioned in Section 5.1.3, this mechanism reflects people’s learning and exploration of the urban environment when his/her original knowledge does not enable a satisfactory travel outcome, and can help avoid unrealistic predictions if all sampled locations happen to be very unsuitable for a certain activity.

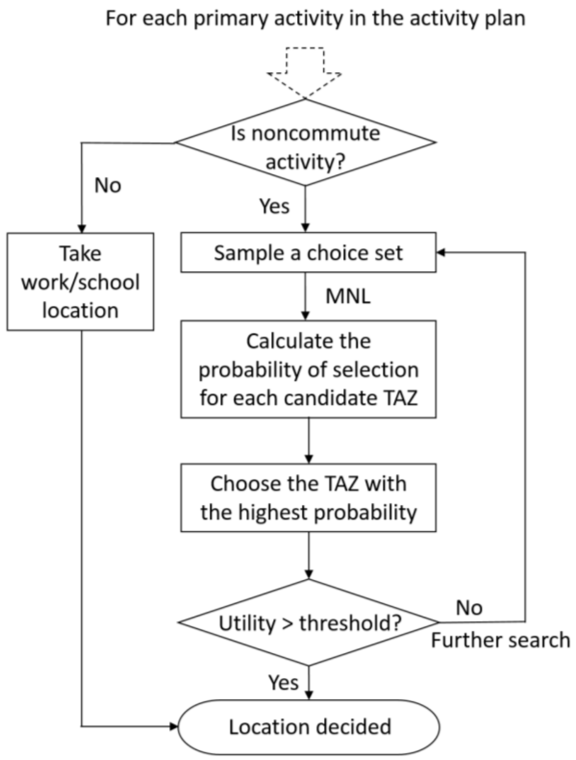

Figure 5-3 Flow diagram of Sub-model 2

5.3.2 Method for parameter estimation

The MNL model used to estimate the probability of choices is specified as:

$$U_{ji} = X_j\beta + \varepsilon_{ji}$$

where U_{ji} is the utility of destination alternative j for a given individual i; $X_j\beta$ represents the systematic component of utility, in which X_j is a vector of attributes for alternative j and β is a vector of coefficients; ε_{ji} represents the stochastic component of utility. Alternative j is chosen if U_{ji} is bigger than the utility of any other choice. If the independence of irrelevant alternatives (IIA) assumption is met, ε_{ji} is Gumbel-distributed and the probability that individual i chooses alternative j is (McFadden, 1978):

$$p(i) = \frac{\exp(\mu X_j\beta)}{\sum_{i\in L}\exp(\mu X_j\beta)}$$

where μ is a scale parameter; and L is the set of available alternatives.

There are basically three types of variables and coefficients in MNL (Croissant, 2012):

- alternative specific variables with a generic coefficient;
- individual specific variables with alternative specific coefficients;
- alternative specific variables with alternative specific coefficients.

Considering that the total number of alternatives under examination can be up to several hundreds, all variables and coefficients in the model are specified as the first type. The variables considered in the model include:

- Travel distance measured as the centroid-to-centroid distance from the tour origin to the candidate TAZ;
- Logarithmic travel distance to allow for a diminishing effect of longer distance;
- Distance band variables to account for non-continuous effect of distance;
- Distance interacting with gender, driving license ownership and car ownership to take into account varying sensitivity to travel distance by individuals with different capacity of mobility;
- Expected travel time by driving and transit;
- Built environment features of the candidate TAZ.

The model is estimated with the ‘mlogit’ package in R Studio v0.99.473. Six models are estimated separately for the six types of activities. Considering that the results of parameter estimation may be affected by the random sampling of alternatives, I repeated the sampling and model estimation several times. It is found that the average values of many coefficients would converge and fluctuate by less than 10% when the process is repeated for ten times. The variables whose coefficients do not converge are all insignificant at 0.1 level and therefore removed from the model. Wald test is then applied to make sure that the model fit is not significantly affected by removing the variables.

In the next step, the weights for different distance bands are calibrated to maximise the approximation between the simulated and the observed average travel distances. The calibration takes a basic parameter sweep approach (Malleson, 2014). The search starts from the values estimated by the MNL model and searches both upwards and downwards with an increment of 0.2, then the range of search is manually narrowed down based on the model fit and a second round of search is conducted with an increment of 0.1. The model fit is judged by both the root mean squared error (RMSE) and the ratio between the simulated and the actual travel distances.

Last, the threshold of utility is calibrated using a similar approach. However, the model prediction turns out to be insensitive to this parameter—in few cases would an individual find a new location that provides a higher utility than the original choice within five rounds of resampling. It suggests that the alternative sampling method is able to produce an effective representation of the activity opportunities in the modelled region. This parameter is therefore excluded from the final model.

5.3.3 Results of parameter estimation

The results of the location choice models for the six types of activities are shown in **Table 5-10**. Population density significantly increases the chance of choosing a TAZ as the destination for most activities, including shopping, personal business, escorting and others. In contrast, employment density is negatively associated with the destination choice for most activities (the effects are significant for shopping, entertainment and others). Higher retail density encourages the destination choice for shopping and personal business, but discourages that for entertainment. Similarly, higher entertainment density is positively associated with shopping, dining out and other activities, but negatively associated with entertainment and personal business. The effects of these two variables on entertainment seem a bit counter-intuitive, but can be because many of the entertainment activities reported by the interviewees actually refer to physical exercises such as strolling and jogging, which tend to be conducted in a more natural environment. For the similar reason, land use mix, which is supposed to enhance the activeness of the urban environment and the level of convenience, also shows a positive association with most activities except for entertainment.

In terms of the transport infrastructure, the density of primary roads does not turn out as a very influencing factor on destination choice, which is only negatively associated with entertainment. The density of secondary roads is positively associated with choosing a destination for shopping, but negatively associated with entertainment and escorting. Possible explanation could be that people may prefer high connectivity for shopping but tend to avoid traffic for the latter two activities. Similarly, higher density of tertiary roads increases the chance of choosing a location for shopping and other activities, but again discourages the location choice for entertainment. Besides, lower accessibility to subway station decreases the chance of choosing a location for most activities, while higher bus coverage generally increases the chance of location choice, as one would expect.

In terms of the street facade features, the facade quality is negatively associated with the destination choice for most activities while the facade continuity has the reversed effects. Besides, the interaction terms show that being male, owning a car and holding a driving license all increase the chance of choosing a farther destination, among which holding a driving license has the largest effect. The calibration of the weights of distance bands are shown in **Table 5-11**.

Table 5-10 MNL model results on the location choice for primary destinations

Variables	Shopping		Entertainment		Dining out		Personal business		Escorting		Others	
	B	S.E.	B	S.E.	B	S.E.	B	S.E.	B	S.E.	B	S.E.
Distance	-	-	-	-	8.93E-05	8.17E-05	5.82E-05	3.71E-05	-	-	5.90E-05	2.95E-05
In(distance)	-1.84	0.0838	-2.09	0.0889	-2.1	0.387	-1.54	0.213	-1.29	0.215	-1.77	0.244
Distance band (ref=<1000m)												
1000-2000m	1.16	0.0642	1.19	0.0704	1.68	0.296	1.17	0.168	0.714	0.17	1.21	0.204
2000-3000m	1.8	0.109	2.21	0.115	2.6	0.467	1.81	0.255	1.6	0.267	1.9	0.304
3000-4000m	2.77	0.14	3.02	0.151	2.88	0.602	2.41	0.308	2.32	0.34	3.25	0.364
4000-5000m	2.54	0.167	2.79	0.177	3.25	0.664	2.4	0.342	1.87	0.396	3.3	0.404
5000-10000m	3.81	0.179	3.95	0.196	5.4	0.741	3.16	0.39	2.84	0.452	3.9	0.454
>10000m	4.91	0.25	4.96	0.261	6.28	1.03	3.54	0.504	3.42	0.569	4.42	0.516
Expected travel time by public transit	-0.000491	2.85E-05	-0.000392	2.86E-05	-0.00062	0.000105	-0.000209	4.95E-05	-0.00038	6.01E-05	-0.000171	4.66E-05
Expected travel time by driving	0.000105	6.66E-05	0.000149	6.55E-05	-	-	-	-	-	-	-	-
Population density	1.45E-05	2.29E-06	-	-	1.21E-05	8.37E-06	1.37E-05	4.29E-06	1.36E-05	5.16E-06	2.01E-05	3.93E-06
Employment density	-2.13E-05	2.27E-06	-1.54E-05	2.48E-06	-	-	-	-	-6.77E-06	4.58E-06	-1.32E-05	3.49E-06
Retail density	0.00356	0.000494	-0.00731	0.000779	-	-	0.00325	0.00085	0.000977	0.00114	-	-
Entertainment density	0.00587	0.000784	-0.00337	0.000999	0.00724	0.00265	-0.0031	0.00154	0.0022	0.0019	0.00582	0.00123
Land use mix	-	-	-0.255	0.0596	0.541	0.241	0.679	0.125	0.385	0.146	0.448	0.11
Distance to the nearest commercial cluster	-	-	-	-	-0.000242	0.000112	-8.87E-05	5.51E-05	-0.000139	6.58E-05	-	-
Primary road density	-	-	-7.52E-05	1.60E-05	-5.08E-05	5.55E-05	-	-	-	-	-	-
Secondary road density	5.29E-05	1.00E-05	-3.47E-05	1.10E-05	-	-	-	-	-7.40E-05	2.43E-05	-	-
Tertiary road density	1.40E-05	5.85E-06	-0.00015	6.44E-06	-	-	1.92E-05	1.07E-05	1.28E-05	1.30E-05	2.52E-05	8.86E-06
Parking density	-5.20E-05	1.06E-05	2.15E-05	1.11E-05	-3.42E-05	3.42E-05	-	-	4.58E-05	2.14E-05	-	-
Distance to the nearest	-5.08E-05	3.13E-05	-0.000311	3.33E-05	-0.000451	0.00013	-0.000145	6.01E-05	-0.000171	7.18E-05	-3.65E-05	4.77E-05

subway station												
Bus coverage	0.397	0.114	0.31	0.126	0.484	0.451	1.7	0.231	1.74	0.284	0.425	0.195
Facade quality	-0.645	0.136	-0.702	0.0259	-0.213	0.0878	-1.05	0.267	-0.241	0.0881	-	-
Facade continuity	1.29	0.217	5.75	0.211	1.7	0.699	0.656	0.415	1.88	0.453	-	-
In(distance): Gender (male)	0.197	0.0379	-	-	-	-	-0.014	0.062	0.0874	0.0812	-	-
In(distance): Car ownership (yes)	-	-	-	-	0.234	0.13	0.131	0.0798	0.0754	0.0916	0.15	0.0707
In(distance): Driving license (yes)	0.247	0.0455	0.137	0.0552	0.201	0.127	0.381	0.0825	0.514	0.0956	0.0813	0.0707

Table 5-11 Calibrated weights of distance bands

Distance band (m)	Shopping	Entertainment	Dining out	Personal business	Escorting	Others
1000-2000	1.94	1.23	2.51	1.60	1.72	2.17
2000-3000	2.92	2.69	3.63	2.72	1.86	3.05
3000-4000	3.84	3.43	4.68	3.19	2.04	4.34
4000-5000	3.65	3.36	5.04	3.29	2.77	3.65
5000-10000	4.97	4.79	5.72	3.64	3.23	5.18
>10000	6.49	5.56	6.91	4.38	3.79	5.64

5.3.4 Validation of the sub-model

The results on the test set show that the model is able to produce good estimation of the average travel distances for the six types of activities. The largest gap between observation and simulation happens on the dining out activities, which can be at least partly explained by the small sample size of the test data. When the results are split by ring roads, the model is able to capture the trend that residents in outer areas of the city generally travel longer distances. In some cases, the model provides quite approximate estimations to the observed data, e.g. the average shopping distance of the residents living between the 2nd and 3rd ring roads, and that of the residents living between the 4th and 5th ring roads. The model performs the worst on entertainment activities, which suggests that the model may not sufficiently account for the factors related to entertainment location choices.

Table 5-12 Simulated and observed travel distances on the test set

		Simulated		**Observed**		**Mean (sim)/Mean (obs)**
	N	**Mean**	**S.D.**	**Mean**	**S.D.**	
Shopping	1021	1639	1220	1542	1989	1.06
Entertainment	968	1649	930	1860	2252	0.89
Dining out	64	2564	1448	3019	3779	0.85
Personal business	217	3852	4068	3670	3587	1.05
Escorting	160	2771	1507	2895	3177	0.96
Others	224	5586	1427	5493	4538	1.02

Table 5-13 Simulated and observed travel distances by ring roads on the test set

		Within 2nd ring	**2nd-3rd ring**	**3rd-4th ring**	**4th-5th ring**
	N	262	318	211	230
Shopping	obs	1317	1565	1427	2048
	sim	1227	1513	1912	2045
	N	317	321	184	146
Entertainment	obs	1625	1586	2020	2791
	sim	1789	1675	1500	1470
Dining out	N	21	17	9	17

	obs	2290	2605	4486	3336
	sim	1905	2519	2568	3262
	N	58	68	49	42
Personal business	obs	2949	3294	4548	4177
	sim	2500	3016	4878	5660
	N	38	36	50	36
Escorting	obs	2326	2974	2859	3451
	sim	1916	2423	3057	3589
	N	60	74	44	46
Others	obs	5696	4700	5579	6474
	sim	4377	5030	6645	7153

Note: Underlined numbers do not follow the general trend that residents at outer rings travel longer distances, which might be systematic and worth attention, or simply results of large random errors due to the small sample sizes.

5.4 Sub-model 3: Time of activity and mode choice

5.4.1 Introduction to the sub-model

The time of activity and the travel mode are modelled jointly since the questionnaire survey suggests that people tend to give similar priority to these two choice facets. Moreover, the two choices can be mutually influential. For instance, one may take subway instead of driving if he/she has to travel in rush hours, or may try to avoid rush hour if he/she prefers driving. The time of activity is estimated through the probability distribution from the observed data, given the activity type and the position in the activity plan. The mode choice is estimated through multinomial logit models, which is straightforward. Besides, a mechanism is designed to allow the adjustment of activity scheduling if none of the travel modes could provide a satisfactory utility under the original schedule.

The 24 hours of a day are divided into six periods in the model: before am peak (03:00-07:00 am), am peak (07:00-09:00 am), before noon (9:00-12:00 am), afternoon (12:00-17:00 pm), pm peak (17:00-19:00 pm) and after pm peak (19:00 pm-03:00 am). The peak hours are identified from the 24-hour congestion index (annual average) in the

Transportation Report of Beijing 2011 (**Figure 5-4**). For the ease of computation, if more than half of the travel time of a tour falls in a certain time period, the tour is considered to be conducted in that period. The key point of travel time prediction is to differentiate between peak hour tours and non-peak hour tours, which could influence the mode choice.

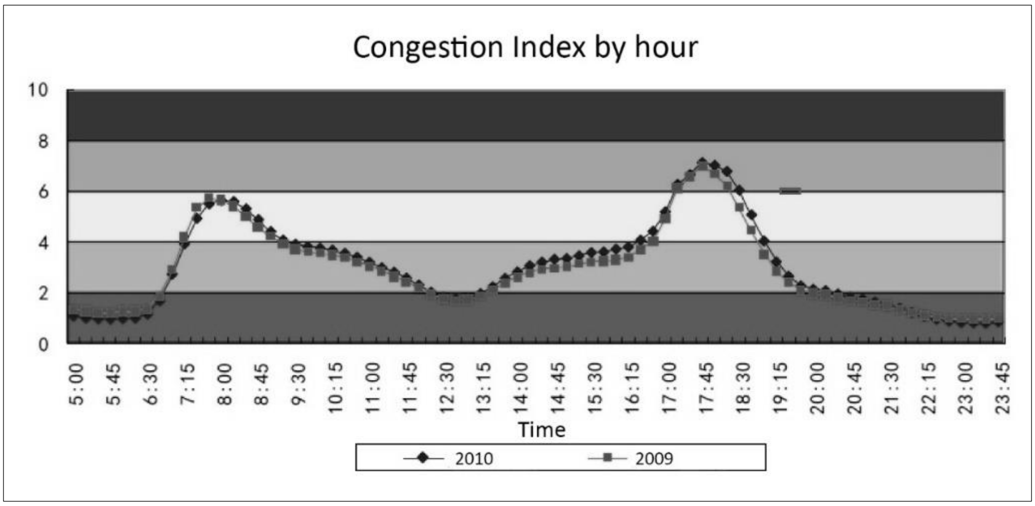

Figure 5-4 Congestion levels by hour

Source: Transportation Report of Beijing 2011, pp.52.

The computational process of this sub-model is designed as:

- First, the individual arranges a time slot for each of the primary activities in his/her activity plan based on the observed probabilities of activity time choice (see **Table C-11** in Appendix C).
- Second, the individual checks whether the time arrangement is reasonable: (1) activities that come later in the activity plan should not take an earlier time than previous activities, (2) non-commute activities that take place after commute activities should not take a time slot earlier than afternoon. If there is any conflict, the first step is repeated until a reasonable time plan is produced.
- Third, the individual calculates the probabilities of choosing each of the four travel modes using multinomial logit models and chooses the mode with the highest probability.

- Last, the individual compares the systematic utility component of the chosen mode with a threshold value. If the utility is lower than the threshold, he/she reschedules the activities and repeats the mode choice process.

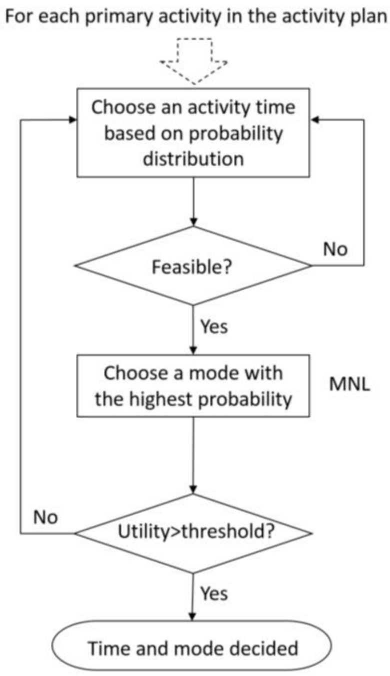

Figure 5-5 Flow diagram for Sub-model 3

5.4.2 Method for parameter estimation

Considering that the mode choice for different types of activities may be subject to different considerations, three separate MNL models are estimated for work, school and non-commute tours. The influencing factors considered in the models include:

- Individual's socioeconomic characteristics;
- Built environment conditions at the tour origin (home);
- Built environment conditions at the tour destination;
- Travel-related variables including the travel distance, expected travel time by each mode, whether the tour contains any intermediate stop and whether the tour is conducted during peak hours.

All the variables are estimated with alternative specific coefficients to account for varying effects of these variables on different modes. For instance, the same time spent in a car and walking may be valued differently (Bowman & Bradley, 2005).

In the next step, the constants of different modes are further calibrated to maximise the approximation between the simulated and the observed mode shares. The parameter sweep starts from the values estimated by the MNL models and searches both upwards and downwards with an increment of 0.1. The range of search is then manually narrowed down based on the model fit and a second round of search is conducted with an increment of 0.05. The model fit is judged by both the percentage of correct predictions and the mode shares.

Last, the threshold of minimum utility is calibrated. However, similar to the location choice, the model results turn out to be insensitive to this parameter—in few cases would an individual find a combination of travel time and mode that induces a higher utility if the original choice does not meet the threshold. This parameter is therefore also excluded from the final model.

5.4.3 Results of parameter estimation

The results of MNL models are presented in **Table 5-14**. For work tours, all socioeconomic variables have significant impacts on mode choice. Older age is associated with higher chance of cycling, but lower chance of taking public transit, which could be a result of the habits of the elder generation. Men are more likely to drive than women. Higher income and social status are generally related to higher chance of driving than taking public transit and cycling. Both couples and single persons are more likely to drive comparing with people from a core family. Holding a driving license and owning a private car are positively associated with driving, while owning a motorcycle or an e-bicycle are positively associated with cycling, as one would expect.

In terms of the built environment characteristics at the tour origin (home), higher population density marginally increases the chance of cycling, which could be because higher density can lead to a safer and more active environment for walking/cycling. Longer distance to commercial clusters decreases the chance of cycling and taking public transit (although insignificant, the p-values are not much above 0.1), possibly by reducing the activeness of the environment and the convenience of shopping along the way. For similar reasons, retail density and land use mix are positively associated with all non-driving modes, although the former is only significant for walking and the latter is only significant for cycling. When comes to the transport infrastructure, the density of secondary road has a significant effect in discouraging cycling, probably because the volume of traffic could pose safety risks for cycling (though the traffic volume is the largest on primary roads, cyclers usually ride on secondary and tertiary roads). Besides, longer distance to subway stations discourages transit use, as one would expect, but the effect is marginal. In terms of the street facade features, the only significant relationship is the positive correlation between the facade continuity and public transit use.

In terms of the built environment characteristics at the tour destination (work place), higher population density is positively associated with cycling and walking (marginal), which is similar to its effects at the tour origin. Longer distance to the city centre decreases the chance of public transit use and cycling, which is also reasonable. Similar to the results at the tour origin, retail density, entertainment density and land use mix at the tour destination also show a generally positive relationship with all non-driving modes, although only the effect of entertainment density on walking is significant. Regarding to the transport infrastructure, the density of tertiary roads encourages cycling, since they are usually more cycling and walking friendly. Besides, higher bus coverage decreases the chance of walking and cycling, probably because people may take bus as a substitute for walking and cycling.

Regarding to the tour characteristics, longer travel distance decreases the chance of cycling and walking, as one would expect. When the travel distance is controlled, travel time is positively associated with taking public transit and walking but negatively associated with cycling. Traveling during peak hours increases the chance of taking public transit and cycling. Making an intermediate stop along the tour discourages the use of all non-driving modes since the automobile is more convenient for making stops (Bowman & Bradley, 2005).

School tours are different from work tours in many aspects since most of the travellers are young students and are usually escorted and chauffeured by adults. Therefore, the influencing mechanism of the mode choice for school tours can be different from adults' commute tours. In terms of socioeconomic variables, older age is associated with higher chance of using non-driving modes, which can be explained by the increasing independency of older children so there is less need for chauffeuring. For similar reasons, male students are more likely to cycle. Unlike in work tours, holding a driving license does not show a significant impact on mode choice. Instead, household car ownership is positively associated with driving. These results can also be explained by the fact that students usually do not drive by themselves but are chauffeured by parents. Besides, some of the relationships are similar to those in work tours, which include: students from high income families (annual income=250-300 thousand RMB) are less likely to walk; couples are more likely to drive; lower social status increases the use of all non-driving modes; and owning a motorcycle or e-bicycle increases the chance of cycling and walking.

In terms of built environment variables, many of the results are similar with work tours. Higher retail density at the tour origin increases the chance of choosing non-driving modes. Higher density of secondary roads at the tour origin is negatively associated with cycling. Higher density of tertiary roads at the tour destination is positively associated walking. Entertainment density and the use of non-driving modes are negatively associated at the tour origin but positively associated at the tour destination, which is also observed in work tours but not statistically significant. Higher density of parking spaces decreases the chance of choosing non-driving modes, especially the public transit, due to the increased convenience of driving.

In terms of tour characteristics, longer travel distance is negatively associated with walking and cycling, which is the same as in work tours. When the travel distance is controlled, the expected travel time is positively associated with public transit use and negatively associated driving. Making an intermediate stop along the tour also reduces the possibility of using non-driving modes.

For non-commute tours, older people are more likely to use non-driving mode. Men are less likely to take public transit and walk. People from non-core family households are all more likely to drive. Lower social status is positively associated with all non-driving modes and the relationships are all significant at 0.1 level. Both holding a driving license and owning a car significantly increase the chance of driving. Owning a motorcycle or an e-bicycle increases the chance of cycling.

In terms of the built environment variables, higher employment density at the tour origin and being closer to the city centre are positively associated with walking and cycling, which is consistent with the assumptions in Section 2.1.4. The density of tertiary roads at the tour origin is negatively associated with the use of all non-driving modes, which indicates that driving tends to benefit the most from the enhanced connectivity of road network. Higher bus coverage at the tour origin decreases the chance of cycling, possibly because people tend to substitute cycling with taking bus when the latter is more convenient. When the tour destination is farther from the city centre, people are less likely to take public transit. Higher land use mix at the tour destination is negatively associated with all non-driving modes and the effect is significant for walking. This result is a bit counter-intuitive but could be because that the higher level of functional mix makes it more possible to perform multi-tasks, which is more convenient by driving. Besides, higher level of facade continuity at the tour destination marginally encourages public transit use and cycling, which, as assumed in Section 2.1.4, could be related to the enhanced attractiveness of the urban environment. Last, the effects of tour characteristics are generally similar with those in work and school tours.

Table 5-14 MNL model results on mode choices

	Work			School			Non-commute		
	Transit	**Cycle**	**Walk**	**Transit**	**Cycle**	**Walk**	**Transit**	**Cycle**	**Walk**
Variable	**B**	**B**	**B**	**B**	**B**	**B**	**B**	**B**	**B**
Constant	4.03***	5.64***	8***	1.66	2.42	4.31*	2.88*	4.48***	7.78***
Socioeconomic									
Age	-0.0191***	0.0224***	2.20E-05	0.0782***	0.0635**	0.0725***	0.0288***	0.0105**	0.024***
Gender (Ref=Female)									
Male	-0.819***	-0.188*	-0.383***	0.168	0.727***	0.351 Ψ	-0.649***	0.0119	-0.438***
Annual income (Ref=<50 thousand RMB)									
50-100	0.00388	-0.247**	0.0947	-0.293	0.133	-0.107	-0.0675	-0.24 Ψ	-0.286*
100-150	-0.192	-0.547***	-0.297 Ψ	-0.154	0.346	0.324	-0.077	-0.372 Ψ	-0.52**
150-200	-0.423*	-0.326	0.0653	0.0423	0.299	0.0856	1.07**	0.424	0.503
200-250	-0.765*	-1.46**	-0.658	-0.141	-0.634	0.352	1.68*	0.808	0.631
250-300	-0.525	-0.235	0.408	-0.96	-0.861	-2.69*	0.228	-0.945	-0.858
>300	-0.98**	-1*	-0.238	0.0668	-1.55	-0.226	-0.427	-0.959	0.00221
Life cycle (Ref=Core family)									
Single	-1.02***	-1.38***	-0.787***	-2.23 Ψ	-0.558	0.619	-0.161	-0.622***	-0.316*
Couple	-0.25**	-0.513***	-0.118	-2.48**	-2.36**	-3.55***	-0.101	-0.556*	-0.0201
Multi-generation	0.141	0.03	-0.102	0.332	0.441	0.387	-0.316 Ψ	-0.799***	-0.425*
Others	-0.253*	-0.758***	0.103	-0.259	-0.31	-0.192	-0.141	-0.494*	-0.278
Social well-being (Ref=Best-off)									
Middle	0.21*	0.198 Ψ	0.0242	0.291	0.378	0.335	0.548***	0.594***	0.379**
Least well-off	0.323**	0.351**	0.159	0.715*	0.675*	0.899**	0.614***	0.542**	0.324 Ψ
Driving license (Ref=No)									
Yes	-1.38***	-1.49***	-1.35***	0.429	0.48	1.04	-1.2***	-1.2***	-1.16***
Car ownership (Ref=No)									
Yes	-3.81***	-3.98***	-3.62***	-3.97***	-3.96***	-4.07***	-3.07***	-2.9***	-2.92***

Variable	Work Transit B	Work Cycle B	Work Walk B	School Transit B	School Cycle B	School Walk B	Non-commute Transit B	Non-commute Cycle B	Non-commute Walk B
Motor cycle (Ref=No)									
Yes	0.536 Ψ	1.09***	0.553	-0.444	1.59 Ψ	1.99*	0.275	0.947*	0.348
Electric bicycle (Ref=No)									
Yes	-0.328**	1.4***	0.253 Ψ	-0.333	0.903***	0.0868	-0.0726	1.23***	0.329*
Built environment at tour origin									
Population density	-3.20E-07	1.02E-05 Ψ	5.70E-07	-9.78E-06	1.39E-05	-1.69E-06	-6.46E-06	-7.93E-06	-1.42E-05 Ψ
Employment density	-3.32E-06	-1.83E-06	-4.71E-06	1.77E-05	2.33E-05 Ψ	6.91E-06	1.02E-05	1.82E-05*	1.81E-05*
Distance to city center	7.17E-06	1.50E-05	-3.79E-05	1.33E-05	-0.000114 Ψ	-0.000264**	3.16E-05	-0.000124***	-0.00012***
Distance to the nearest commercial cluster	-7.26E-05	-0.00015**	-1.88E-05	-1.19E-05	4.74E-05	7.21E-05	-0.000122 Ψ	-6.00E-05	-3.01E-05
Retail density	0.00113	0.00112	0.00513**	0.00859*	0.0055	0.0127**	0.00431 Ψ	0.000631	0.000729
Entertainment density	-0.000536	-0.00257	-0.00102	-0.0156**	-0.0081	-0.0177**	-0.00189	9.45E-05	-0.00112
Land use mix	0.192	0.368*	0.354 Ψ	-0.337	0.083	0.35	-0.292	-0.321	-0.16
Primary road density	2.13E-05	5.75E-05	5.48E-05	5.21E-05	2.22E-05	0.000173 Ψ	2.86E-06	-3.76E-05	-5.15E-05
Secondary road density	2.81E-05	-6.49E-05*	3.08E-05	-7.43E-05	-0.000138*	-0.000127 Ψ	-3.68E-05	-1.48E-05	-4.38E-05
Tertiary road density	-2.39E-05 Ψ	-1.05E-05	-3.30E-05 Ψ	-3.92E-06	-3.02E-06	-5.83E-05	-4.12E-05*	-4.34E-05*	-8.25E-05***
Parking density	-2.09E-05	1.33E-05	5.48E-05 Ψ	-1.66E-05	-4.83E-05	9.45E-06	-1.26E-05	-2.76E-05	4.67E-06
Distance to the nearest subway station	-0.000112 Ψ	-5.81E-05	-1.45E-05	-0.000264	-0.000112	-0.000268	4.31E-05	-1.20E-05	6.09E-05
Bus coverage	0.104	-0.167	-0.109	0.716	-0.842	0.557	-0.171	-0.888*	-0.371
Facade quality	0.512	-0.437	-0.509	0.491	0.443	1.2	0.288	0.401	0.186
Facade continuity	1.97***	0.463	0.776	1.37	2.41 Ψ	0.045	0.822	0.844	1.06
Built environment at tour destination									
Population density	-2.13E-07	1.15E-05*	1.13E-05 Ψ	-2.51E-06	-2.53E-05 Ψ	-1.46E-05	-1.23E-05 Ψ	-1.03E-05	-8.05E-06
Employment density	3.56E-06	-5.77E-07	-5.25E-06	-5.45E-06	-4.22E-06	7.42E-06	-5.71E-06	-1.15E-05 Ψ	-9.92E-06
Distance to city center	-3.74E-05*	-4.59E-05*	3.20E-05	-8.73E-06	6.06E-05	0.000264***	-9.21E-	9.57E-07	5.89E-06

	Work			School			Non-commute		
	Transit	**Cycle**	**Walk**	**Transit**	**Cycle**	**Walk**	**Transit**	**Cycle**	**Walk**
Variable	**B**	**B**	**B**	**B**	**B**	**B**	**B**	**B**	**B**
							05***		
Distance to the nearest commercial cluster	-2.45E-05	-7.76E-05	-6.07E-05	6.53E-05	-7.63E-05	-4.31E-05	1.71E-05	0.000144	0.000191*
Retail density	0.000525	0.000343	-0.000371	0.00401	-0.00173	-0.00475	0.00198	0.000169	0.000614
Entertainment density	0.000957	0.00173	0.00435*	0.01*	0.00904 Ψ	0.0112*	-0.0014	0.004	-0.000345
Land use mix	0.146	0.198	0.191	0.408	-0.00203	-0.267	-0.302 Ψ	-0.057	-0.399*
Primary road density	1.21E-05	-4.98E-05	2.84E-05	2.65E-05	-4.41E-05	-9.89E-05	-8.59E-06	1.24E-05	8.08E-05
Secondary road density	2.38E-05	4.03E-05 Ψ	1.04E-05	-7.35E-05	-3.60E-05	-7.13E-05	-5.78E-05 Ψ	-5.65E-05	-1.45E-05
Tertiary road density	1.56E-05	2.97E-05*	2.69E-05	-3.78E-05	2.75E-05	0.000105**	-1.62E-06	3.16E-05	1.91E-05
Parking density	-3.46E-06	-1.51E-05	3.96E-06	-0.000134*	-6.23E-05	-7.04E-05	-7.44E-06	-3.08E-05	-1.40E-05
Distance to the nearest subway station	-0.000198**	-9.17E-05	-8.44E-05	1.37E-05	1.57E-05	0.000263	-0.000139	-4.19E-05	-1.82E-05
Bus coverage	-0.122	-0.572*	-0.629*	-0.865	-0.349	-1.09	0.457	0.227	-0.414
Facade quality	-0.431	-0.418	-0.537	0.335	0.892	0.253	-0.00858	0.528	0.638
Facade continuity	0.358	0.664	0.504	1.41	0.974	1.45	1.34 Ψ	1.42 Ψ	0.775
Tour characteristics									
Travel distance	2.54E-05 Ψ	-0.000295***	-0.00163***	7.07E-05	-0.00041***	-0.00195***	-2.98E-05	-0.000514***	-0.00184***
Expected travel time	8.33E-05*	-0.000243***	3.82E-05*	0.000213*	0.000236	4.84E-05	0.000203***	0.000114	2.07E-05
Travel during peak hours: yes	0.465***	0.235*	-0.148	-0.301	0.166	0.369 Ψ	0.176	0.183	0.174
Include intermediate stop(s): yes	-0.954***	-0.928***	-1.14***	-1.2**	-0.904*	-1.24**	-0.74***	-0.663***	-0.893***
	Expected travel time (drive): B=-6.3952e-05			Expected travel time (drive): B=5.5896e-04			Expected travel time (drive): B=1.5519e-04		
	Log-Likelihood: -9548.8			Log-Likelihood: -1858			Log-Likelihood: -9481		
	McFadden R2: 0.4612			McFadden R2: 0.37326			McFadden R2: 0.3987		
	Likelihood ratio test: chisq = 16347			Likelihood ratio test: chisq = 2213.1			Likelihood ratio test : chisq = 12573		

	Work			School			Non-commute		
	Transit	**Cycle**	**Walk**	**Transit**	**Cycle**	**Walk**	**Transit**	**Cycle**	**Walk**
Variable	**B**	**B**	**B**	**B**	**B**	**B**	**B**	**B**	**B**
	(p-value < 2.22e-16)			(p.value < 2.22e-16)			(p.value < 2.22e-16)		

By performing the parameter sweep, the constants in the three models are calibrated to the values shown in **Table 5-15**. The percentages of correct prediction (PCP) are 69%, 66% and 74% after calibration. In a similar work by P. Zhao (2011), the PCP of the MNL model for worker's mode choice is 71.3%, in which only three types of travel modes are considered (car, public transport, and foot, bicycle or other modes). If walking and cycling are merged as one mode in my model, the PCP can increase to 78% (without calibration). Therefore, the model fit can be considered as fairly good.

Table 5-15 Calibrated constants for mode choice

Work	**School**	**Non-commute**
Public transit: 4.53	Public transit: 1.51	Public transit: 2.88
Cycle: 6.14	Cycle: 2.62	Cycle: 4.62
Walk: 7.60	Walk: 4.06	Walk: 6.69
PCP = 69%	PCP =66%	PCP =74%

5.4.4 Validation of the sub-model

The simulation performed on the test set shows that the model is able to provide a good estimation of mode choice. The shares of the four modes are close between the simulation and the observation for all three types of activities. The percentages of correct prediction are 71% and 73% for work and non-commute tours and 58% for school tours.

Table 5-16 Confusion matrix of the mode choice for work tours on the test set

	Cycling (sim)	**Driving (sim)**	**Transit (sim)**	**Walking (sim)**	**Total**
Cycling (obs)	226	26	103	83	20.0%
Driving (obs)	17	242	56	19	15.2%
Transit (obs)	98	72	664	26	39.3%
Walking (obs)	84	19	31	425	25.5%
Total	19.4%	16.4%	39.0%	25.2%	**PCP=71%**

Table 5-17 Confusion matrix of mode choice for school tours on the test set

	Cycling	**Driving**	**Transit**	**Walking**	**Total**

	(sim)	(sim)	(sim)	(sim)	
Cycling (obs)	52	7	27	22	27%
Driving (obs)	2	16	16	7	10%
Transit (obs)	25	12	66	10	28%
Walking (obs)	32	4	7	102	35%
Total	27%	10%	29%	35%	**PCP=58%**

Table 5-18 Confusion matrix of the mode choice for non-commute tours on the test set

	Cycling (sim)	**Driving (sim)**	**Transit (sim)**	**Walking (sim)**	**Total**
Cycling (obs)	125	11	83	152	14%
Driving (obs)	12	84	31	18	6%
Transit (obs)	63	27	396	58	21%
Walking (obs)	131	24	84	1295	59%
Total	13%	5%	23%	59%	**PCP=73%**

5.5 Sub-model 4: Location choice for intermediate stops

5.5.1 Introduction to the sub-model

The choice of intermediate stops is modelled in a similar approach as the choice of primary destinations. The major difference lies in that two anchor points are considered in measuring the impedance of travel, which is calculated as the detour distance from the tour origin to the candidate stop and then to the primary destination minus the direct distance from the tour origin to the destination. Besides, the travel mode is already decided and therefore taken into account in the choice of intermediate stops, which is supposed to enhance the model performance since existing research suggest that long detour trips are more likely to happen with certain modes than others (Ho & Mulley, 2013). The computational process of this sub-model is designed as:

- First, for each intermediate stop in the activity plan, a subset of ten TAZs are drawn as candidate locations using the detour distance as the weighting factor. The quota given to each detour distance band is derived from the distribution of detour distance for each type of activity in the observed data (see **Table C-12** in Appendix C).

- Second, the probability of each candidate TAZ being chosen is calculated, also estimated using multinomial logit models.
- Third, the candidate TAZ with the highest selection probability is chosen as the location of the intermediate stop.
- Last, the systematic utility component of the chosen location is compared with a threshold value. If the utility is lower than the threshold, step one to three are repeated and a higher utility location is selected, if possible.

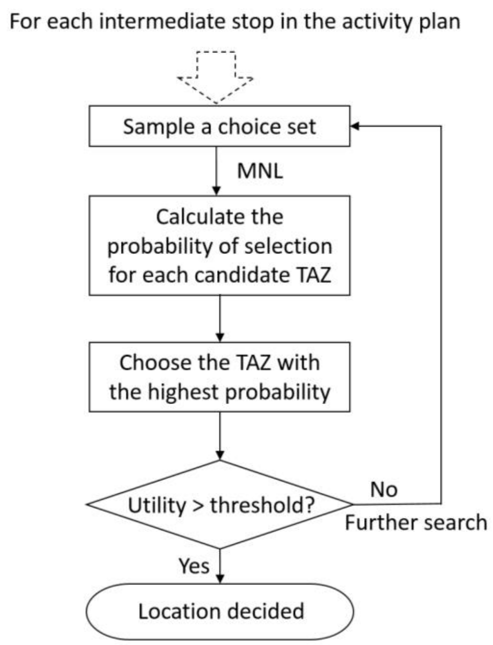

Figure 5-6 Flow diagram of Sub-model 4

5.5.2 Method for parameter estimation

As mentioned before, the method used in this sub-model is very similar to the second sub-model. MNL model is applied to estimate the weights of the influencing factors, which include:

- Detour distance measured as the total distance from the tour origin to the candidate TAZ and then to the tour destination, minus the direct distance from the tour origin to the tour destination;
- Logarithmic detour distance to allow for a diminishing effect of distance;

- Detour distance interacting with gender and travel mode to take into account varying sensitivities to the detour distance by different genders and different modes;
- Detour distance band variables to account for non-continuous effect of distance
- Built environment features of the candidate TAZ.

The candidate TAZs are also re-sampled and the model parameters are re-estimated for ten times. The variables whose coefficients do not converge (fluctuate by less than 10%) are removed (all these variables are not significant at 0.1 level). Wald test is applied to make sure that the model fit is not significantly affected by removing the variables. After estimating the MNL model, the weights of different detour distance bands are calibrated using parameter sweep to maximise the approximation between the simulated and the observed average detour travel distances. The threshold of utility also turns out to be uninfluential to the modelling results and therefore is removed.

5.5.3 Results of parameter estimation

The results of the stop location choice models for the six types of activities are shown in **Table 5-19**. Many of the relationships are similar to those in the location choice of primary destinations. Population density is positively associated with the chance of choosing a location for most activity types, except for entertainment and ‘other’ activities (significant for shopping, dining out and ‘others’). On the contrary, employment density has the reversed effects for most activities (significant for shopping). Higher retail density and entertainment density both increase the chance of choosing a location for all activity purposes except for entertainment, which is also observed in the location choice for primary destinations. The effects of retail density are significant for shopping, entertainment, dining out and personal business, while the effects of entertainment density are significant for shopping, dining out and escorting. Higher level of land use mix is positively associated with the location choice for most activities, since it could contribute to the availability of activity opportunities as well as the activeness of the urban environment.

Regarding to the transport infrastructure, the density of primary roads does not play a significant role in the stop location choice for any activity purpose, which is similar to the results on primary destinations. The density of secondary roads only show a significantly positive relationship with the stop location choice for shopping. Higher density of tertiary roads significantly increases the chance of the location choice for shopping but decreases that for entertainment, which indicates that people prefer high connectivity for shopping but tend to avoid traffic for entertainment. Besides, both better accessibility to subway stations and higher bus coverage enhance the attractiveness of a location for almost all types of activities, which is reasonable.

The two street facade variables do not exert a significant influence in most cases, except for the positive relationship between the facade continuity and the location choice for entertainment. Regarding to the interaction terms, gender does not show a significant relationship with the detour distance. Driving and taking public transit increases the chance of choosing a stop that needs a longer detour distance for almost all purposes, and walking shows the opposite effects, as one would expect.

Table 5-19 MNL model results on the location choice for intermediate stops

	Shopping		Entertainment		Dining out		Personal business		Escorting		Other	
Variables	**B**	**S.E.**	**B**	**S.E.**	**B**	**S.E.**	**B**	**S.E.**	**B**	**S.E.**	**B**	**S.E.**
Detour distance	-0.000204	4.25E-05	-0.000199	8.48E-05	-0.000206	4.80E-05	-0.000101	3.63E-05	-0.000195	3.46E-05	-0.000128	3.47E-05
In(detour distance)	0.038	0.0539	0.345	0.173	-	-	-0.289	0.127	-0.0481	0.0646	0.111	0.194
Detour distance band (ref=<1000m)												
1000-2000m	-0.372	0.0894	-0.736	0.214	-0.441	0.139	0.222	0.233	-0.317	0.124	-0.578	0.247
2000-3000m	0.273	0.145	-0.719	0.306	-1.42	0.213	-0.384	0.283	0.00516	0.159	-0.462	0.312
3000-4000m	-0.486	0.199	-0.762	0.458	-1.4	0.279	-0.231	0.385	0.179	0.215	-0.66	0.448
4000-5000m	-0.902	0.248	-0.638	0.494	-2.05	0.372	-0.987	0.52	0.15	0.238	-0.441	0.44
5000-10000m	0.938	0.325	1	0.7	0.632	0.393	0.285	0.384	0.77	0.274	0.792	0.425
>10000m	0.887	0.614	0.354	1.38	0.994	0.717	1.43	0.681	2.1	0.507	0.544	0.641
Population density	2.19E-05	4.24E-06	-6.91E-06	1.12E-05	1.23E-05	6.18E-06	2.48E-06	9.34E-06	6.90E-06	5.21E-06	-2.45E-05	1.02E-05
Employment density	-1.62E-05	3.91E-06	-1.27E-05	1.01E-05	-	-	-1.07E-05	8.20E-06	-3.26E-06	4.08E-06	-1.52E-05	9.37E-06
Distance to the nearest commercial cluster	-	-	-0.000105	0.000145	-3.41E-05	9.42E-05	-	-	-	-	-	-
Retail density	0.0033	0.000888	-0.00593	0.0025	0.00355	0.00106	0.00395	0.00178	0.000494	0.00109	0.00196	0.00224
Entertainment density	0.00811	0.00136	-	-	0.00417	0.00177	0.00566	0.00341	0.00744	0.00181	0.00194	0.00367
Land use mix	-0.0521	0.107	0.069	0.253	0.361	0.157	1.12	0.275	0.377	0.147	0.811	0.282
Primary road density	-	-	-0.000123	7.02E-05	-5.55E-05	3.69E-05	-4.71E-05	6.21E-05	-	-	-	-
Secondary road density	-6.65E-05	1.86E-05	6.51E-05	4.74E-05	-1.43E-05	2.24E-05	-	-	-	-	-3.90E-05	4.99E-05
Tertiary road density	4.48E-05	1.76E-05	-3.84E-05	4.74E-05	1.85E-05	2.28E-05	3.47E-05	3.92E-05	-3.78E-05	2.34E-05	-	-
Parking density	2.85E-05	9.67E-06	-0.000108	2.62E-05	-	-	-	-	-	-	-1.64E-05	2.51E-05
Distance to the nearest subway station	-0.000157	6.31E-05	-0.000124	0.00015	-9.47E-05	9.40E-05	-0.000236	0.000133	-0.000162	7.28E-05	-0.000143	0.000125
Bus coverage	-	-	0.243	0.536	0.599	0.303	-	-	1.12	0.28	1.61	0.532
Facade quality	-	-	-	-	-	-	-	-	-0.279	0.315	-0.29	0.456
Facade continuity	-	-	1.42	0.321	-	-	0.207	0.32	0.83	0.484	1.13	0.943

	Shopping		Entertainment		Dining out		Personal business		Escorting		Other	
Variables	**B**	**S.E.**	**B**	**S.E.**	**B**	**S.E.**	**B**	**S.E.**	**B**	**S.E.**	**B**	**S.E.**
In(detour distance): male	-	-	-	-	0.0113	0.109	-	-	-	-	-	-
In(detour distance): drive	0.473	0.096	0.184	0.23	0.878	0.139	0.584	0.15	0.319	0.0637	0.486	0.181
In(detour distance): transit	0.335	0.0629	-0.0825	0.178	0.548	0.14	0.56	0.139	0.303	0.0911	0.261	0.175
In(detour distance): walk	-0.122	0.0561	-0.536	0.171	-0.157	0.108	-0.0722	0.152	-0.261	0.0824	-0.214	0.184

Table 5-20 Calibrated weights of detour distance bands

Distance band (m)	Shopping	Entertainment	Dining out	Personal business	Escorting	Other
1000-2000	0.20	-0.43	-0.11	0.28	-0.02	0.23
2000-3000	0.45	-0.8	-0.1	0.1	0.12	0
3000-4000	0.30	-0.37	0.3	-0.05	0.2	-0.8
4000-5000	-0.26	-0.46	0.1	-0.49	0	0
5000-10000	0.43	0.95	0.86	0.18	0.73	1.03
>10000	0.90	0.64	0.9	1.27	2.02	0.6

5.5.4 Validation of the sub-model

The results on the test set show that the model is able to produce good estimation of the average detour distance for the six types of activities (see **Table 5-21**). The differences between the simulation and the observation are smaller than 10% for five of the six activity types. The largest gap happens on the activity type of escorting, which require further research to improve the modelling performance in the future. Due to the small sample size of the test set, the comparison is not done at the ring road level.

Table 5-21 Simulated and observed detour distances on the test set

		Simulated		**Observed**		**Mean (sim)/Mean (obs)**
	N	**Mean**	**S.D.**	**Mean**	**S.D.**	
Shopping	224	1367	1275	1260	2048	1.08
Entertainment	41	2033	2340	2142	3365	0.95
Dining out	185	1379	2136	1513	3268	0.91
Personal business	45	3078	4498	3432	4866	0.9
Escorting	146	2312	2966	3050	4059	0.76
Others	41	3606	3894	3813	4965	0.95

5.6 Validation of the whole model

Two indicators of synthesised travel outcomes are selected for the validation of the whole model: the total travel distance for non-commute purposes and the VMT. The former is the combined outcome of the participation, the organisation and the location choice of non-commute activities. The latter is the outcome of all choice facets in the activity-travel. Ideally, doing the validation at more disaggregate levels can enhance the strength of the results. However, the sample size at more disaggregate levels could be small and thus contain large random errors. For instance, the test data have only 21 samples per TAZ in average. Therefore, the comparison is also done at the ring road level.

The results show that the model is able to capture the general pattern of the amounts of travel—residents living at outer parts of the city tend to both travel longer distances for non-commute purposes and drive more. The R^2 values between the simulation results and the observations are fairly high. However, the model tends to underestimate the VMT, especially for residents living between the 4th and the 5th ring roads.

Table 5-22 Validation of the whole model

		Within 2nd ring road	2nd-3rd ring road	3rd-4th ring road	4th-5th ring road	R^2
N		1620	1817	1397	1702	
Non-commute distance	obs	2527	2687	2795	3182	0.98
	sim	2669	2901	3033	3387	
VMT	obs	1135	1428	1903	2406	0.83
	sim	1049	1055	1672	1780	

5.7 Chapter summary

This chapter proposes a disaggregate activity-travel model, named as BEATIM, which has a particular emphasis on the impacts of the built environment on the various aspects of daily travel, including activity frequency, activity location choice, mode choice, etc. To the best of the author's knowledge, it is the most comprehensive model that explicitly links the activity-based modelling approach, which is mainly developed in the field of transport simulation, with the analysis of the built environment-travel relationship, which is mainly conducted in the field of urban planning and design. It is built on the gap that, on one hand, existing activity-based models usually do not take sufficient account of the built environment conditions, and on the other hand, previous research on the built environment-travel relationship focus mainly on the synthesised outcomes of travel instead of the underlying behavioural processes.

The BEATIM model generally falls in the category of utility maximisation models, and takes weak computational process features. It should be noted that the LUTI system is by nature a complicated system with many interrelated components and one could be tempted to include more and more behavioural processes to fully account for various mechanisms of interactions, when developing models within this system. Therefore, special care is taken to keep close focus on the daily travel behaviour and the influence of the built environment when building the BEATIM model. The model system contains four major components: namely the sub-model for the activity participation and organisation, the sub-model for the location choice of primary activities, the sub-model for the time of activity and mode choice, and the sub-model for the location choice of intermediate stops.

The validation shows that the model is able to produce a good simulation of people's daily travel behaviour. It is acknowledged that there can be large prediction errors at the individual level due to the complex, stochastic nature of activity-travel behaviour (Kulkarni & McNally, 2000). However, the correlation between the simulation results and the observed behaviour at more aggregate levels, such as the ring road level, can be fairly high. Besides, the results of model parameter estimation do suggest that the built environment factors can play a significant role in many choice facets of activity-travel and should be accounted for in travel demand forecasting—something which currently does not happen a lot (Zegras, 2010).

A major shortcoming of the model lies in that only point estimates of model parameters are used. The estimates could be more or less different from the population parameter due to chance error, which can accumulate with each modelling step (Cao & Fan, 2012). Therefore, future development of the model should consider the confidence intervals of the model parameters.

In conclusion, the BEATIM model offers new opportunities for the analysis of the built environment-travel relationship. The high level of behavioural detail enables a closer examination of the underlying processes that give rise to the observed impacts of the built environment, which can help answer questions such as whether the reduction on VMT is caused by smaller share of driving, or shorter travel distance, and if the latter, the distance of which types of activities, etc. Such potential is further exploited in the next chapter.

Chapter 06 Model application and simulation results

6.1 Method

In this chapter, the BEATIM model developed in Chapter 5 is applied to simulate how would people's travel behaviour change in response to changes in the built environment. The simulation involves several scenarios, each assumes a certain amount of change in a certain aspect of the built environment and in a certain spatial extent. The simulation results will respond to the gap of lack of understanding on how the built environment influences the detailed behavioural processes of daily travel (Research Question 1&3 in Section 1.2). These results will be compared with those from existing research on American and European cities and help address the gap of lack of research on fast growing Asian cities (Research Question 4 in Section 1.2).

Two types of scenarios are designed for the simulation, namely 'local' scenarios and 'regional' scenarios. Local scenarios refer to changing the built environment conditions in one TAZ and examining the impacts on the travel behaviour of the residents in that TAZ. The results from local scenarios are comparable to many of the existing research that focus on the neighbourhood built environment, the spatial extent of which is similar to the TAZs. Regional scenarios refer to changing the built environment conditions in all TAZs within a buffer distance from a central TAZ (measured in centroid-centroid distance) and examining the impacts on the travel behaviour of the residents in that central TAZ. This type of scenario probes into the spatial extent of the built environment's influence and answers questions such as whether travel behaviour is influenced by the built environment in a narrow 1-kilometre radius or in as large as a 5-kilometre radius (Aditjandra, Cao, & Mulley, 2012; Pinjari & Bhat, 2011).

It should be noted these scenarios are relatively basic and simple. Given the high level of spatial and behavioural detail of the model, more complicated scenario designs can also be accommodated, such as the simultaneous changes of two or more built environment features or more realistic land development patterns. Therefore, the potential application of this model can be quite diverse.

More technically, the local scenarios are designed to be a 50% or 100% increase in one built environment aspect (except for the distance to the city centre, which is unrealistic to be changed), resulting in a total number of 2*14=28 scenarios. 50% and 100% increases are selected since the effect sizes of the built environment are usually small and these are relatively large changes, so that the consequent impacts on travel behaviour can be more prominent. Larger increases are not considered since those situations would imply a drastic change of the urban form (e.g. three times of the current density), which are not very realistic. The regional scenarios are designed to be a built environment change in five different buffer radius, which are 0 (equivalent to a local scenario), 500 metres, 1000 metres, 1500 metres and 2000 metres. Considering that the five buffer radius and the fourteen built environment features already result in 5*14=70 scenarios, only 100% increases in the built environment features are tested. Each of the local and regional scenarios is tested on all resided TAZs and the average effect sizes are calculated. Considering the small sample size of residents in each TAZ, the experiment on each TAZ is run for ten times and the average effect size is used.

As mentioned before, a major advantage of the BEATIM model lies in its ability to probe into the detailed processes of the built environment's influence. Therefore, when presenting the results of local built environment changes, I will first briefly show the results on the integrated indicator of daily travel, the VMT, and then decompose the results to the impacts on detailed behavioural aspects, including the activity frequencies, travel distances and the mode choices. In a later section, the results on VMT will also be compared against the meta-analysis results in Section 2.3. For regional scenarios, since the focus lies in analysing the spatial extent of the influences, only the results on synthesised travel indicators are presented, i.e. the total, commute and non-commute VMT and the total non-commute travel distance. The simulation results will also be compared against the assumptions made from theoretical deductions in Section 2.1.4. How the empirical findings support or provide alternative insights into the theoretical deductions will be examined. Last, policy suggestions will be drawn from the findings.

6.2 Results of 'local' scenarios

6.2.1 Overall influences

Figure 6-1 shows the simulated changes of per capita VMT when the numeric values of built environment features increase by 50% and 100%. The shapes of the lines indicate a generally linear relationship. The features that are related to large changes of VMT (larger than ±10%) are land use mix (-15%), bus coverage (-17%), facade quality (-36%) and facade continuity (-27%). The changes related to other features are all smaller than 10%. Particularly, the two newly added features (facade quality and facade continuity) demonstrate quite large effects. However, it should be noted that considering the mean values (0.47 and 0.18) and the standard deviations (0.16 and 0.08) of these two features, a 100% increase could mean substantial changes to the current condition and be difficult to achieve. The directions of most features' influences are consistent with the theoretical assumptions in Section 2.1.4, except for entertainment density and tertiary road density, which will be discussed in the next section by examining the detailed influencing process.

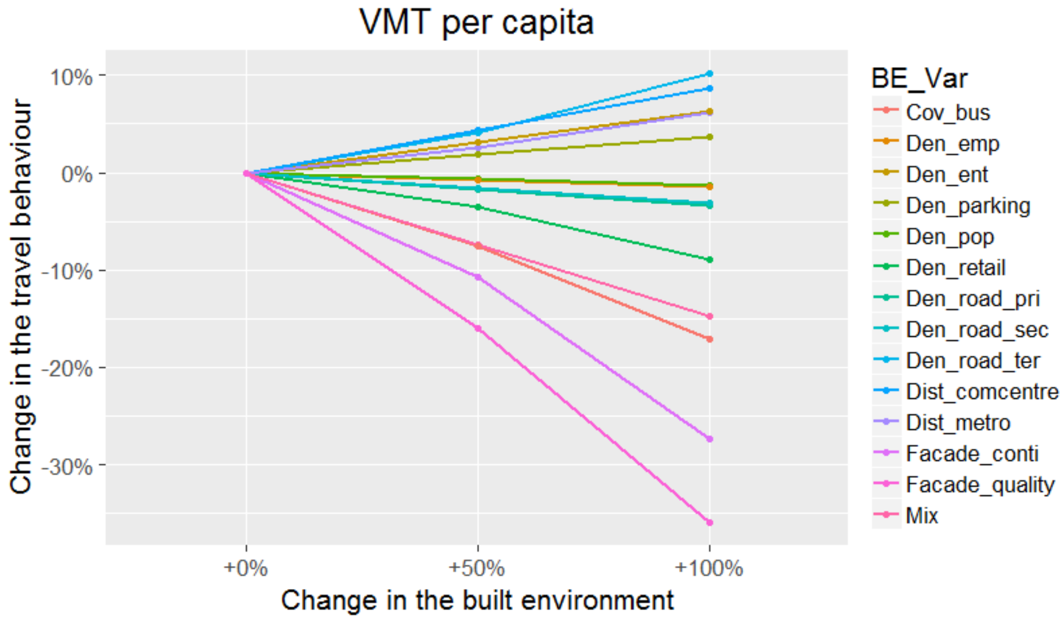

Figure 6-1 Impacts of built environment changes on VMT per capita

6.2.2 Detailed influences

According to the diagram on the 'components' of travel behaviour in Section 2.1.3, the overall impacts of the built environment on VMT can be traced back and subdivided as shown in **Figure 6-2** to **Figure 6-15**. Since the work place is taken as exogenous in the model, the results do not involve any change in the commute distance. A common finding of all scenarios is that the changes of VMT are numerically closer to the changes of commute VMT than those of non-commute VMT, which can be explained by the fact that commute VMT contributes to approximately 80% of total VMT according to the travel diary survey. This result indicates that the sole focus on the total VMT as a key indicator of daily travel may mask relationships that are numerically less dominant. Though the total VMT could be a more balanced representation of commute and non-commute VMT in cities with a generally shorter commute distance, it is still important to have a more comprehensive understanding on the influence of the built environment.

Density

A 100% increase in population density is related to no change in commute VMT and

6% decrease in non-commute VMT. The result on commute VMT is a combined outcome of 1% increase in activity frequency and 3% decrease in driving mode share. The behavioural process that gives rise to the result of non-commute VMT is more complicated. It is assumed in Section 2.1.4 that higher population density could lead to shorter non-commute travel distance and smaller share of driving, but these effects could be compensated by more activity participation. These assumptions are supported by the simulation results. Nonetheless, the changes in non-commute travel distance vary with different activity purposes: the distance of shopping decreases the most, while on the contrary, the distance of entertainment activities increases. The explanation for the latter could be that entertainment turns out to be more about physical activities like strolling and outdoor physical exercises, which tend to seek natural surroundings and avoid crowdedness.

Employment density does not show a prominent effect on commute travel, but is associated with 14% less non-commute VMT when it doubles. However, different from the effects of population density, the decrease in non-commute VMT is mainly accounted for by a decrease in driving mode share, while the travel distance actually increases by 7%. More detailed results show that higher employment density is associated with longer travel distances for four types of non-commute activities, especially shopping. Possible explanation could be that a large proportion of the retail facilities at business areas are higher-end and not suitable for the needs of everyday shopping. Besides, the compensation mechanism is also observed here—the frequency of non-commute activities reduces by 7% in response to the increased travel distance.

When retail density doubles, the commute VMT shows an 8% decrease and the non-commute VMT shows a 14% decrease. The reductions are mainly contributed by a decrease in driving mode share in both the cases of commute and non-commute travel. The travel distances for shopping, personal business and escorting all decrease as expected, while that of entertainment activities increases significantly, which is not surprising given the explanation mentioned before.

Last, a 100% increase in entertainment density is related with 8% increase in commute VMT and a marginal increase in non-commute VMT. The increase in commute VMT is mainly contributed by an increase in the share of driving, which is contrary to my assumption. Besides, although the combined effect on the total non-commute distance is marginal (-2%), there exist some large effects on the travel distances of specific purposes, including 15% decrease in shopping distance, 9% increase in entertainment distance and 14% decrease in dining out.

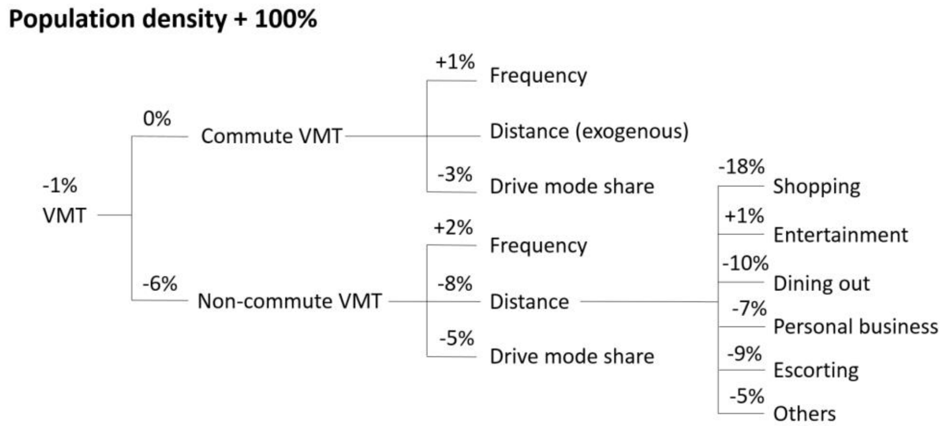

Figure 6-2 Decomposition of the influence of population density

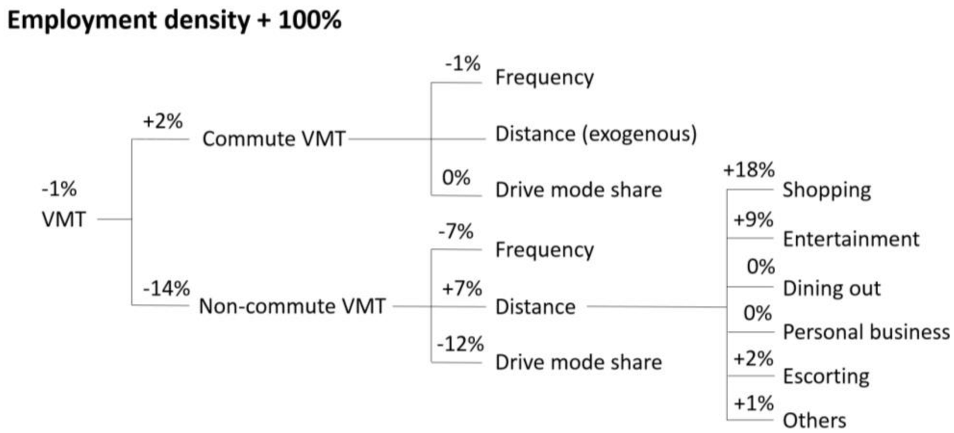

Figure 6-3 Decomposition of the influence of employment density

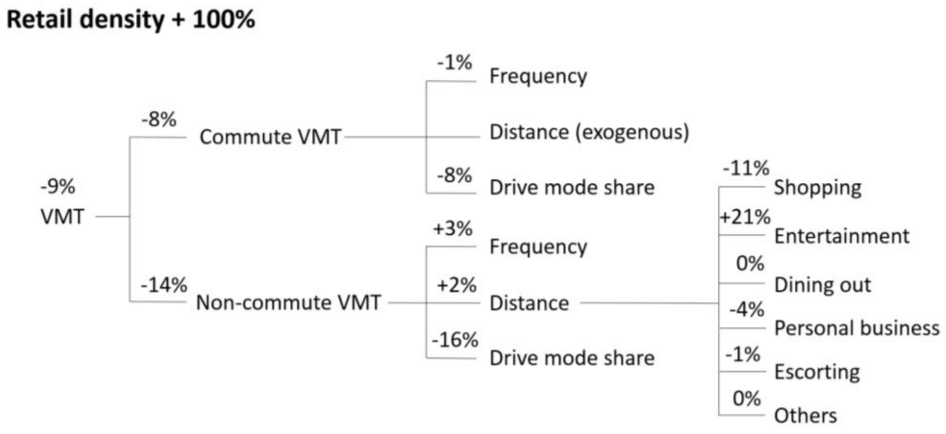

Figure 6-4 Decomposition of the influence of retail density

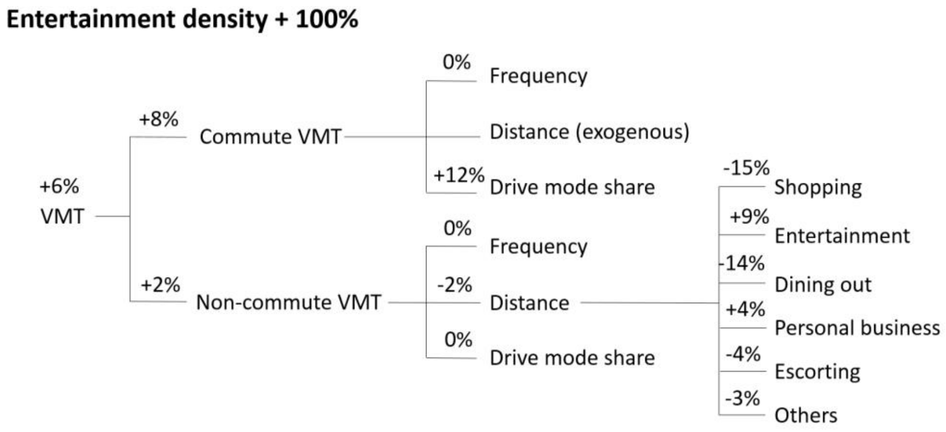

Figure 6-5 Decomposition of the influence of entertainment density

Diversity

A 100% increase in land use mix is related to 21% decrease in commute VMT and 8% increase in non-commute VMT. The decrease in commute VMT is mainly contributed by a 29% decrease in driving mode share, which is larger than the effects of all density features. The increase in non-commute VMT is a combined outcome of 11% decrease in activity frequency, 3% decrease in average travel distance and 26% increase in driving mode share. The large increase in non-commute driving mode share contradicts the hypothesis that higher diversity may discourage driving. A possible explanation could be that higher diversity may induce more trip chaining behaviour since a larger variety of activity opportunities are available, and driving could be a more convenient travel mode when there are multitasks. This assumption is supported by the decrease in travel frequency. In terms of the travel distance, although again the combined effect on the total non-commute distance is marginal (-3%), there exist some large effects on the travel distances of specific purposes, including 25% decrease in dining out, 22% decrease in personal business, 19% decrease in escorting and 18% increase in entertainment.

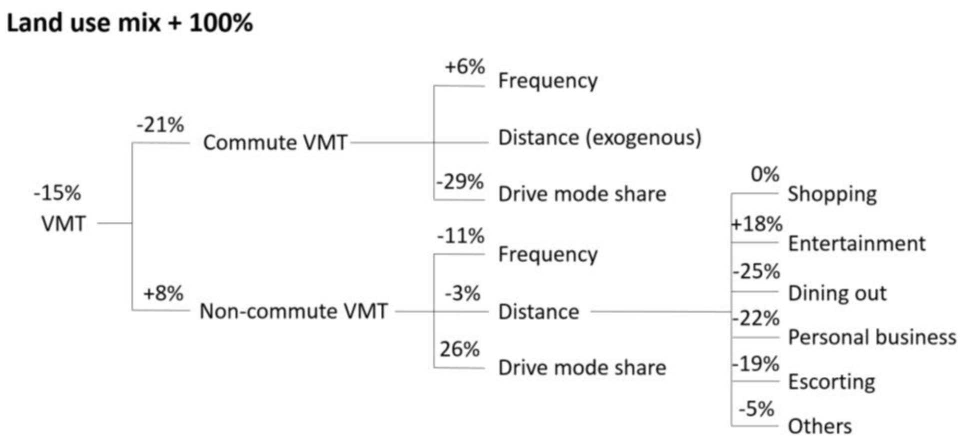

Figure 6-6 Decomposition of the influence of land use mix

Destination accessibility (to sub-centres)

A 100% increase in the distance to the nearest commercial cluster is related to 7% increase in commute VMT and 15% increase in non-commute VMT. The increases mainly come from a higher share of driving. The impacts of this change on non-commute travel distances are generally smaller than the previous scenarios, which is plausible since people may only occasionally need to visit these commercial clusters to fulfil the daily needs.

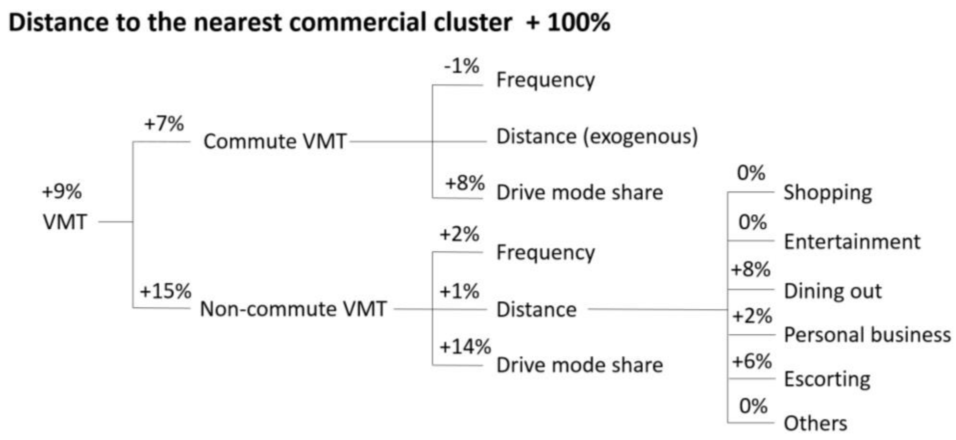

Figure 6-7 Decomposition of the influence of the distance to commercial clusters

Design of the road network

Primary road density shows small impacts on both commute and non-commute VMT and the other detailed aspects of travel. It is probably because that primary roads are mainly expressways and therefore play a smaller role in daily travel. The densities of secondary and tertiary roads show stronger impacts, which are mainly reflected in the increases in the mode share of driving in both commute and non-commute travel. A 100% increase in secondary road density is associated with 5% increase in the share of driving for commute purposes and 14% increase in the share of driving for non-commute purposes. A 100% increase in tertiary road density is associated with 9% increase in the share of commute driving and 22% increase in the share of non-commute driving. Besides, in terms of travel distances, secondary road density is associated with longer distances for entertainment and escorting, and a shorter distance for shopping. Tertiary road density is only largely associated with longer distance for entertainment.

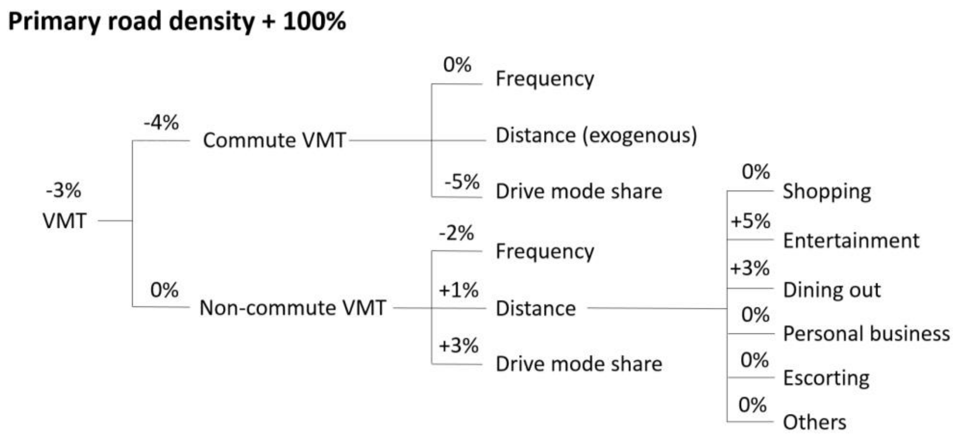

Figure 6-8 Decomposition of the influence of primary road density

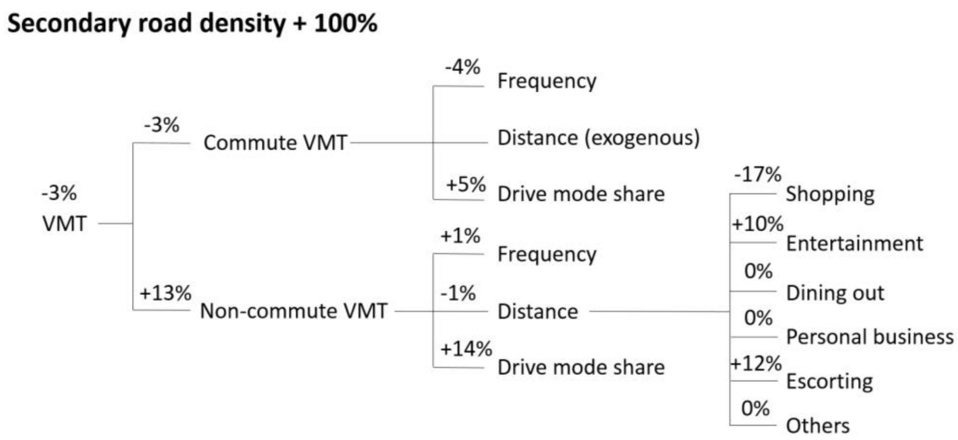

Figure 6-9 Decomposition of the influence of secondary road density

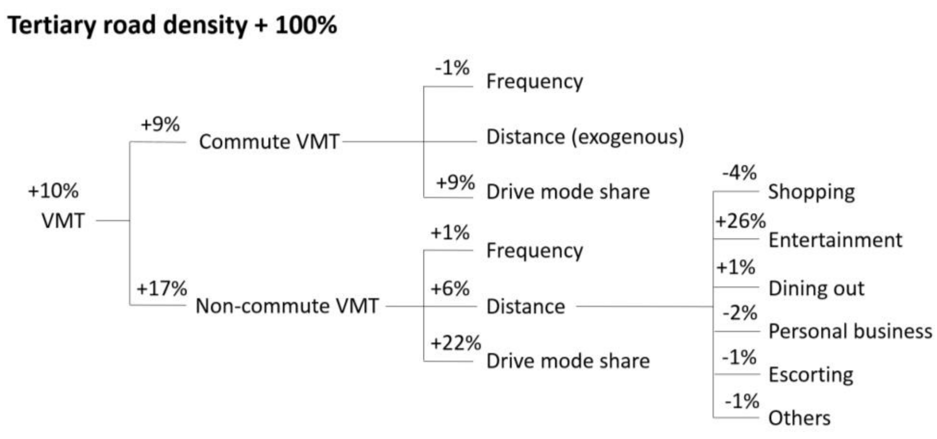

Figure 6-10 Decomposition of the influence of tertiary road density

Distance to transit

Both higher bus coverage and shorter distance to subway stations show significant impacts on reducing commute VMT, mainly through reducing the share of driving. The effect of bus coverage is comparatively larger. Bus coverage also shows a larger impact on non-commute travel. When bus coverage doubles, the non-commute VMT would decrease by 11%, which is a combined outcome of 2% increase in travel frequency, 13% decrease in travel distance and 10% increase in the share of driving. When the distance to the nearest metro station doubles, the non-commute VMT would decrease by 5%, which is a combined outcome of 4% decrease in travel frequency, 5% increase in travel distance and 2% decrease in the share of driving. Besides, more detailed observation on the travel distances shows that all types of non-commute activities tend to be conducted at nearer places when the accessibility to public transit is enhanced.

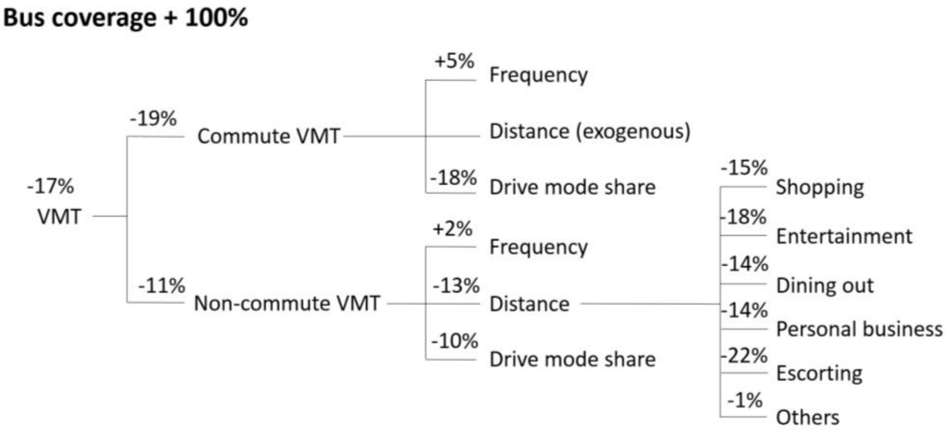

Figure 6-11 Decomposition of the influence of bus coverage

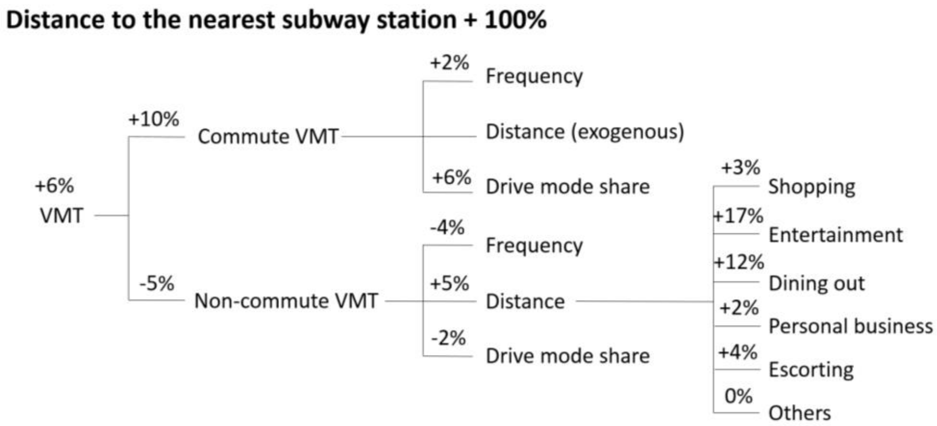

Figure 6-12 Decomposition of the influence of bus coverage

Demand management (parking supply)

A 100% increase in parking density is related to 3% increase in commute VMT and 6% increase in non-commute VMT. The effect on non-commute VMT mainly comes from a 5% increase in the mode share of driving. The result indicates that the hypothesised effect of parking density in encouraging driving is more obvious on non-commute travel than on commute travel.

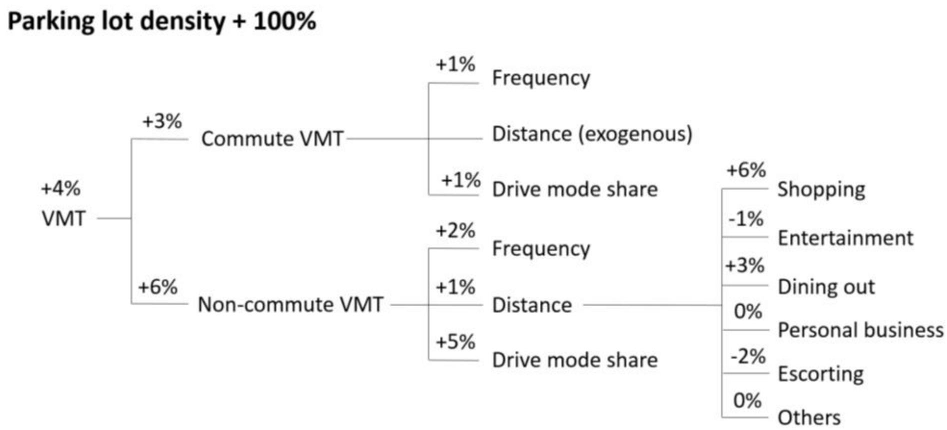

Figure 6-13 Decomposition of the influence of parking density

Design of street facade

As found in the last section, both of the two newly added features demonstrate large effects on travel behaviour. When the score of facade quality doubles, there would be 50% decrease in commute VMT, induced from 28% decrease in the share of driving and 9% increase in the travel frequency; and 12% increase in non-commute VMT, induced from 27% increase in activity frequency, 37% increase in average travel distance and 10% decrease in the share of driving. While the directions of changes in travel frequency and mode choice are consistent with the theoretical assumptions, the changes in travel distances are quite reversed. These changes can be traced back to the results of parameter estimation in Section 5.3.3 that higher quality of street facade is negatively correlated with the attractiveness of destinations for all types of non-commute activities except ‘others’. Therefore, an increase in the facade quality score in the home TAZ would result in more residents choosing farther locations. This negative correlation counters the assumption that high quality street facades can arouse positive feelings and contribute to the attractiveness of an area. A possible explanation is that, in the context of Beijing, high quality buildings (residences, offices, etc.) are usually more private and gated and have stricter control on accommodating small businesses, in order to avoid messiness. As a result, though good facade quality may bring an extra psychic gain for travellers, the level of convenience (the utilitarian value) can be largely undermined in the context of Beijing.

In terms of the facade continuity, when the score doubles, there would be 32% decrease in commute VMT, which is a combined effect of 30% decrease in the share of driving and 10% increase in the travel frequency; and 13% decrease in non-commute VMT, as a combined effect of 6% increase in activity frequency, 13% decrease in average travel distance and 20% decrease in the share of driving. Besides, the travel distances of all types of non-commute activities decrease with enhanced continuity. These changes are all in consistency with the theoretical assumptions.

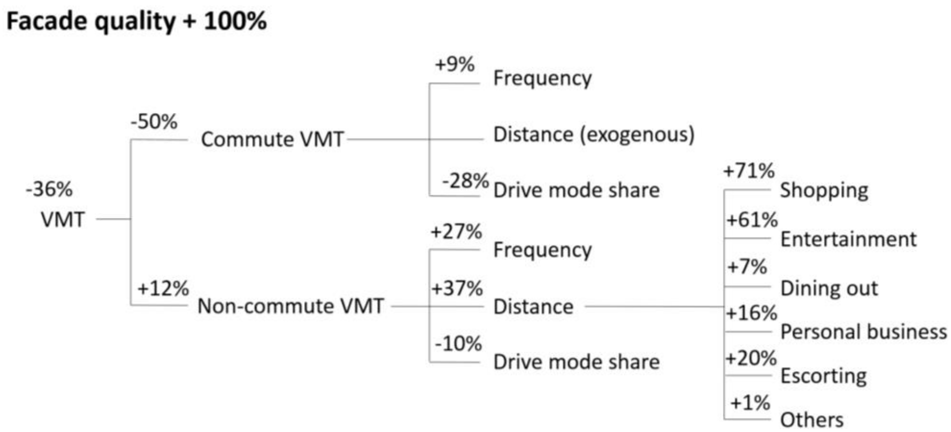

Figure 6-14 Decomposition of the influence of facade quality

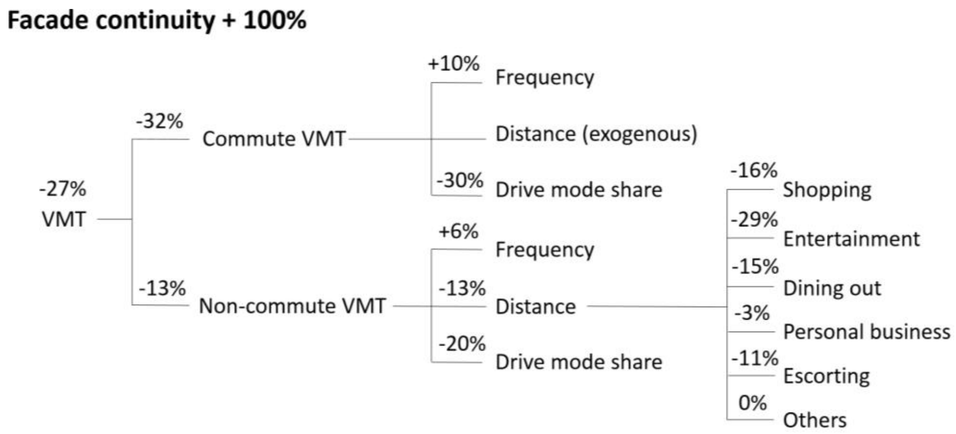

Figure 6-15 Decomposition of the influence of facade continuity

6.3 Results of 'regional' scenarios

Figure 6-16 to **Figure 6-19** show the changes in four major indicators of travel behaviour, i.e. total, commute and non-commute VMT and total non-commute travel distance, under regional scenarios with different buffer sizes. As described in Section 6.1, the regional scenarios are designed to be a built environment change in five different buffer radius around a central TAZ and examining the impacts on the travel behaviour of the residents in the central TAZ. The buffer radius are 0 (equivalent to a local scenario), 500 metres, 1000 metres, 1500 metres and 2000 metres.

The results on the total and the commute VMT are similar, since as explained before, approximately 80% of the total VMT is accounted for by the commute VMT. Generally speaking, there are significant changes in these two types of VMT when the scenario switches from BAU to 'built environment change only in the central TAZ' (Scenario 1) and then to 'built environment change in the 500 metre buffer zone' (Scenario 2). The latter (from Scenario 1 to Scenario 2) induces more drastic changes in VMT than the former (from BAU to Scenario 1). The changes are then much gentler afterwards (from Scenario 2 to Scenario 5).

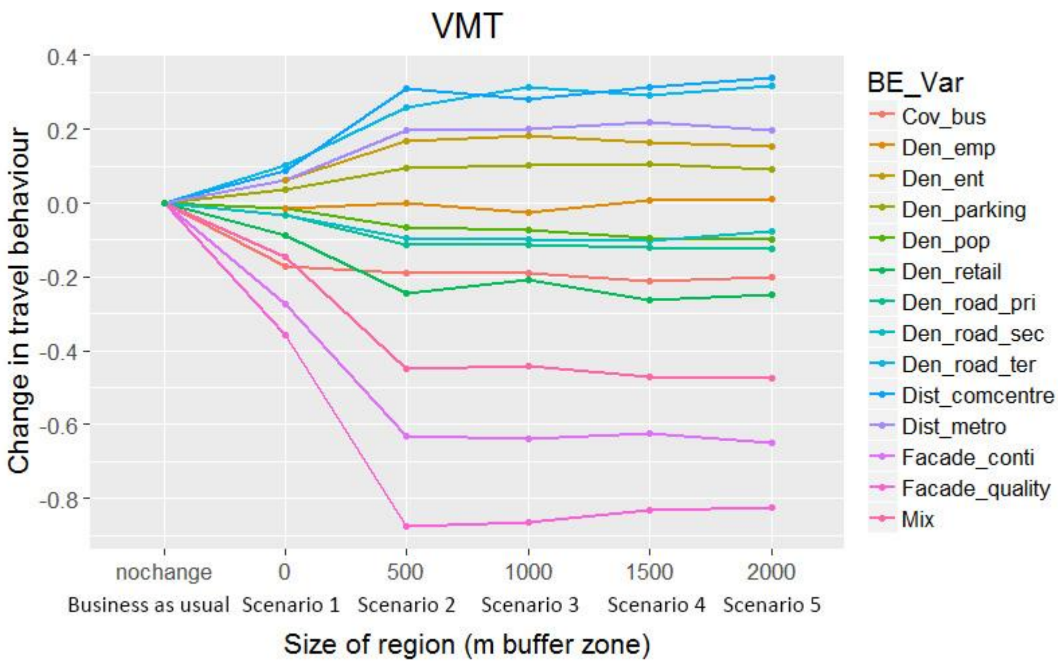

Figure 6-16 Changes in VMT under different regional scenarios

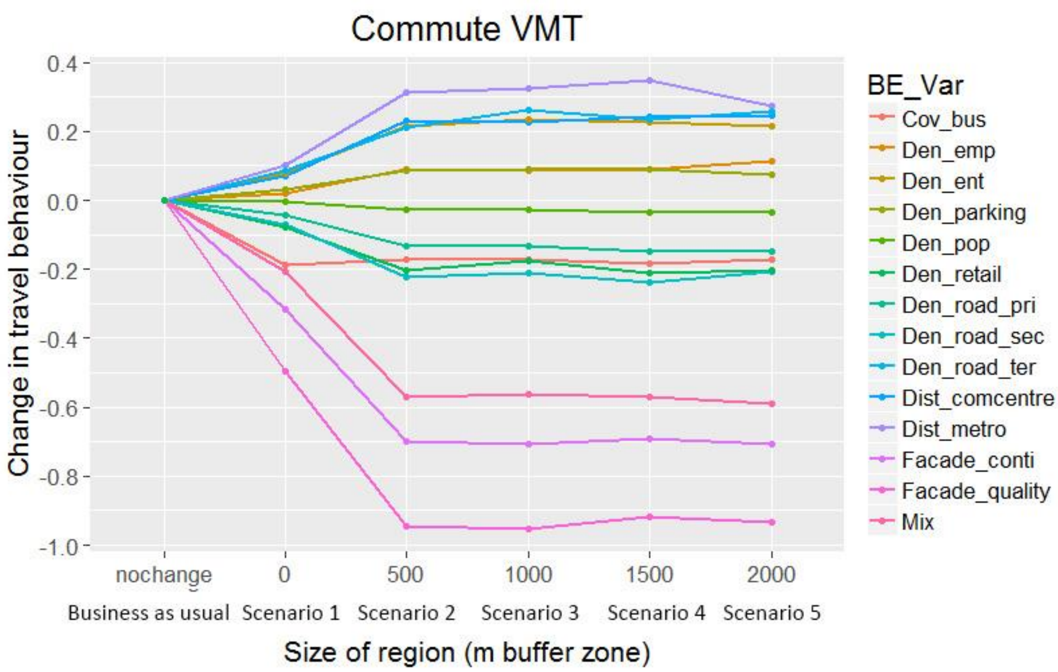

Figure 6-17 Changes in commute VMT under different regional scenarios

The patterns of results are more complicated for non-commute VMT and non-commute distance, since the location and mode choice of non-commute activities are more likely to be influenced by the built environment in the proximate areas from home. Comparing with the total and the commute VMT, these two indicators show larger fluctuations when the built environment changes extend beyond the 500 metre buffer zone (Scenario 3 to 5). Nonetheless, the travel behaviour changes from BAU to Scenario 2 are also generally more drastic than those between Scenario 2 and Scenario 5.

In terms of the non-commute distance, some built environment features show monotonous impacts of increasing or decreasing the distance, including population density, retail density, land use mix, accessibility to commercial clusters, tertiary road density and bus coverage. The other built environment features demonstrate fluctuating impacts as the buffer zone expands. For instance, as the distance to the nearest subway station increases across the buffer zones, the non-commute distance first increases from BAU to Scenario 1 and decreases from Scenario 1 to Scenario 2, then increases again from Scenario 3 to Scenario 5. These fluctuations can be understood as outcomes of the changing trade-offs between the attractiveness and the travel impedance of near and far destinations. Decreases in the travel distance indicate that nearer destinations become generally more attractive under certain built environment conditions, and vice versa. The results on non-commute VMT involve even more fluctuations, since it is a combined outcome of non-commute distance (as mentioned before, influenced by the relative attractiveness of near and far destinations) and mode choice (influenced by the built environment at trip origins and destinations). However, it is still the case that the trends of changes become much gentler after Scenario 2.

The main conclusions from these results are

- When the work place and the related commute distance are taken as exogenous, people’s travel behaviour is mainly influenced by the built environment in the near neighbourhood (in my experiment, within the 500 metre buffer zone).
- Built environment measures will not be very effective in reducing the VMT or travel distance of the residents at a place if the measures are not taken in the near

neighbourhood. In other words, there are diminishing returns in terms of the spatial extent of built environment measures.

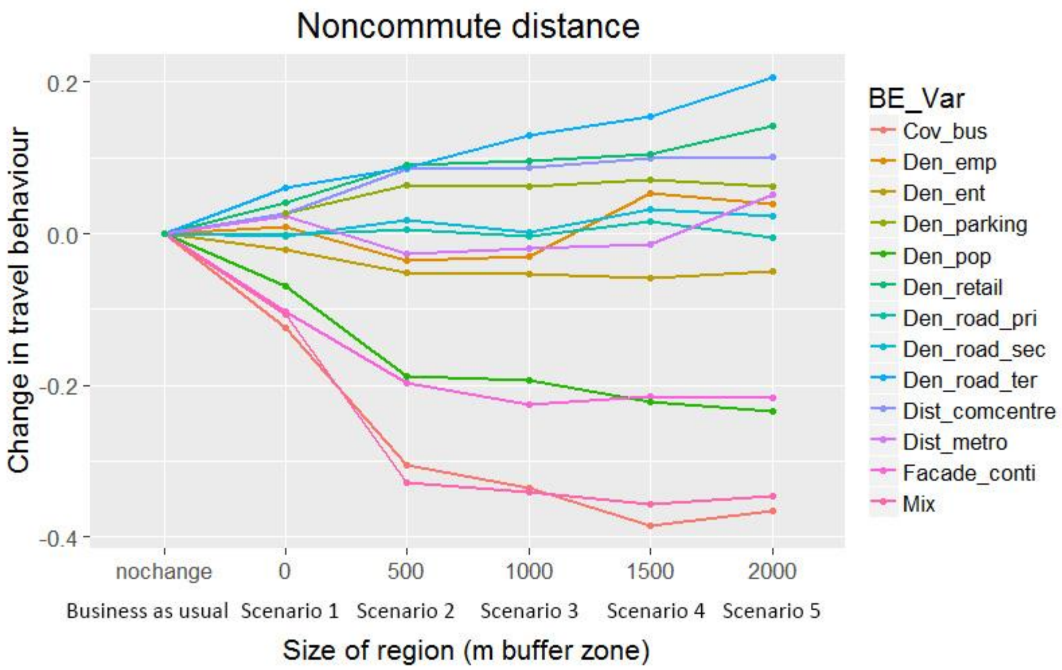

Figure 6-18 Changes in non-commute distance under different regional scenarios

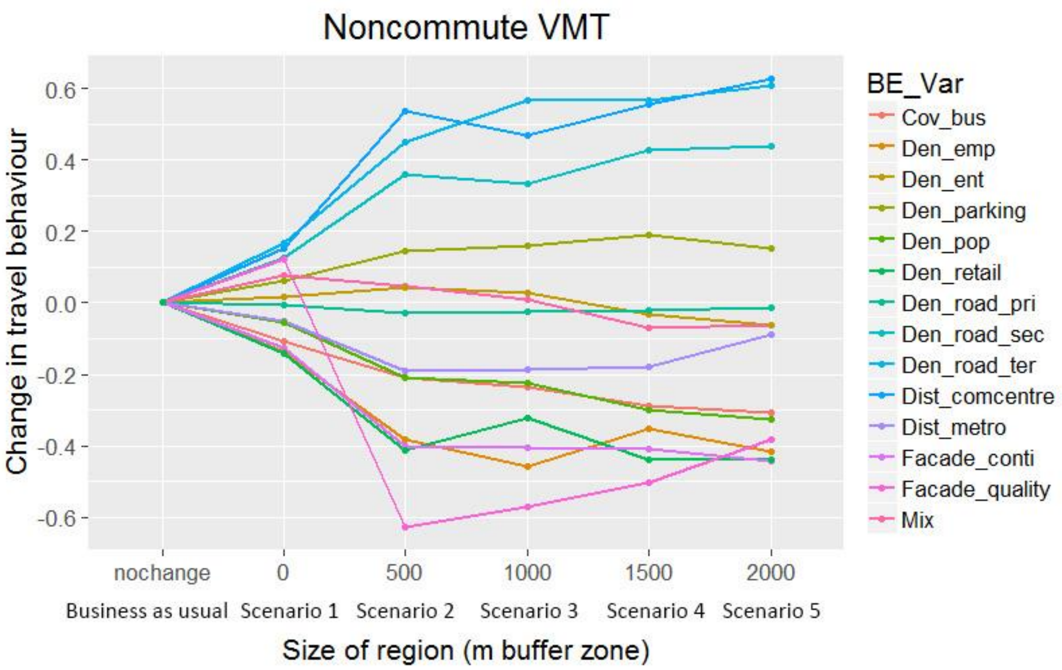

Figure 6-19 Changes in non-commute VMT under different regional scenarios

6.4 Comparing with theoretical assumptions

A particular feature of this research is that it involves a large bundle of assumptions on how daily travel would respond to the built environment conditions, since both daily travel behaviour and the built environment are defined and measured in multiple aspects. Therefore, this section deals with the comparison of the simulation results against the assumptions made from the theoretical deductions (see **Table 2-1** in Section 2.1.4) and examines how the results support or provide alternative insights into the theoretical assumptions. Since the directions of the influences of regional changes are generally consistent with local changes (though the effect sizes might vary), the comparison is mainly based on the results of local scenarios in Section 6.2. Considering that the responses of commute and non-commute travel to the built environment conditions can be quite different, the comparison is made separately for these two types of travel and also separately for the VMT and the travel distance (non-commute only, since commute distance is taken as exogenous). The simulation results that are inconsistent with the assumptions are underlined and highlighted in **Table 6-1**. These differences are discussed below, from which implications for the understanding on the built environment-travel relationship are made.

First, the results seem to indicate that whether higher density relates to enhanced travel gains and thus shorter travel distance could depend on the matchness between the types of density and people's needs. The densities of retail and entertainment facilities have merely marginal effects since they are only relevant to limited types of activities. Besides, the types of goods and services that come with the higher density may also matter. As explained before, higher employment density is related with longer shopping distance probably because a large proportion of the retail facilities at business areas are higher-end and not suitable for everyday shopping.

Second, tertiary roads, which are supposed to be low-speed and pedestrian friendly, turn out to be associated with more driving for both commute and non-commute purposes. A possible explanation lies in people's preference towards driving in China so that the increased road density is more likely to be taken advantage for car use. The preference for driving can be further linked to a social culture of 'car pride', which associates positive self-representation with driving beyond the functional purpose (Z. Zhao & Zhao, 2017). Such an emotional tendency can distort the travel (dis)utility curves and make driving more likely to be chosen.

Third, as discussed before, the facade quality shows largely reversed effect on non-commute travel distance and VMT. The likely explanation is that, in the context of Beijing, high quality buildings are usually more private and gated and have stricter control on accommodating small businesses, in order to avoid messiness. As a result, though good facade quality may bring an extra 'psychic' gain for travellers, as assumed in Section 2.1.4, the level of convenience (the utilitarian gain) can be largely undermined. To a certain extent, this result indicates that since travel is basically an activity for fulfilling needs, utilitarian considerations tend to overweigh psychic enjoyments.

Fourth, the 'compensation' mechanism between travel distance and frequency is observed in many cases. For instance, when the employment density doubles, 7% decrease in non-commute travel frequency is observed with 7% increase in travel distance; when the score of facade continuity doubles, 6% increase in non-commute travel frequency is observed with 13% decrease in travel distance. However, in all scenarios, the changes in the total travel in a day are in the same direction as in individual trips, which indicates that the 'compensation' mechanism is not likely to be stronger than the original effect. This finding is consistent with some previous research, for instance, Feng et al. (2013) mentioned that any rebound effects are considerably smaller than the reduced travel distances resulting from a compact urban structure (Feng et al., 2013).

Table 6-1 Assumptions and simulation results

Built environment feature	**Effect on distance travelled (non-commute)**	**Effect on VMT (commute)**	**Effect on VMT (non-commute)**
Higher density	**On individual trips** Assumption: decrease *Result: depending on the type of density, decrease with population density, increase with employment density, the effects of retail and entertainment density are marginal* **On total travel in a day** Assumption: ambiguous *Result: same as above*	**On individual trips** Assumption: decrease *Result: depending on the type of density, decrease with retail density, increase with entertainment density, the effects of population and employment density are marginal* **On total travel in a day** Assumption: ambiguous *Result: same as above*	**On individual trips** Assumption: decrease *Result: decrease except for entertainment density* **On total travel in a day** Assumption: ambiguous *Result: same as above*
Higher diversity	**On individual trips** Assumption: decrease *Result: marginally decrease* **On total travel in a day** Assumption: ambiguous *Result: decrease*	**On individual trips** Assumption: decrease *Result: decrease* **On total travel in a day** Assumption: ambiguous *Result: same as above*	**On individual trips** Assumption: decrease *Result: increase* **On total travel in a day** Assumption: ambiguous *Result: same as above*
Higher accessibility to district centres	**On individual trips** Assumption: decrease *Result: marginally decrease* **On total travel in a day** Assumption: ambiguous *Result: same as above*	**On individual trips** Assumption: decrease *Result: decrease* **On total travel in a day** Assumption: ambiguous *Result: same as above*	**On individual trips** Assumption: decrease *Result: decrease* **On total travel in a day** Assumption: ambiguous *Result: same as above*
Higher road density	**On individual trips** Assumption: likely to increase	**On individual trips** Assumption: ambiguous	**On individual trips** Assumption: ambiguous

Built environment feature	Effect on distance travelled (non-commute)	Effect on VMT (commute)	Effect on VMT (non-commute)
	Result: depending on the type of road, increase with tertiary road, primary and secondary roads have almost no effect **On total travel in a day** Assumption: likely to increase *Result: same as above*	*Result: increase with tertiary road, primary and secondary roads have marginal effects* **On total travel in a day** Assumption: ambiguous *Result: same as above*	*Result: increase with secondary and tertiary road, primary road has almost no effect* **On total travel in a day** Assumption: ambiguous *Result: same as above*
Higher transit accessibility	<u>**On individual trips**</u> <u>Assumption: slightly likely to increase</u> <u>*Result: both decrease, but the effect of metro accessibility is marginal*</u> <u>**On total travel in a day**</u> <u>Assumption: slightly likely to increase</u> <u>*Result: same as above*</u>	<u>**On individual trips**</u> Assumption: decrease *Result: decrease* <u>**On total travel in a day**</u> Assumption: decrease *Result: decrease*	**On individual trips** Assumption: decrease *Result: decrease with bus coverage, the effect of metro accessibility is marginal* **On total travel in a day** Assumption: decrease *Result: same as above*
More parking provision	**On individual trips** Assumption: slightly likely to increase *Result: almost no effect* **On total travel in a day** Assumption: slightly likely to increase *Result: marginally increase*	<u>**On individual trips**</u> <u>Assumption: increase</u> <u>*Result: almost no effect*</u> <u>**On total travel in a day**</u> <u>Assumption: increase</u> <u>*Result: almost no effect*</u>	**On individual trips** Assumption: increase *Result: marginally increase* **On total travel in a day** Assumption: increase *Result: increase*
Better building design	<u>**On individual trips**</u> <u>Assumption: slightly likely to decrease</u> <u>*Result: depending on the specific feature, decrease with facade*</u>	<u>**On individual trips**</u> Assumption: likely to decrease *Result: decrease* <u>**On total travel in a day**</u>	<u>**On individual trips**</u> Assumption: likely to decrease *Result: decrease* <u>**On total travel in a day**</u>

Built environment feature	**Effect on distance travelled (non-commute)**	**Effect on VMT (commute)**	**Effect on VMT (non-commute)**
	continuity, increase with facade quality **On total travel in a day** Assumption: slightly likely to decrease *Result: same as above*	Assumption: likely to decrease *Result: decrease*	Assumption: likely to decrease *Result: decrease with facade continuity, increase with facade quality*

a Travel behaviour changes ≦ 5% are considered to be 'marginal', ≦ 1% are considered to be 'almost no effect'
b Results that are inconsistent with assumptions are underlined.

6.5 Comparing with findings from American and European cities

In this section, the simulation results on VMT will be compared against the meta-analysis of existing findings on American and European cities. The differences in the findings will be discussed. It should be noted that this comparison is quite rough for at least two reasons. First, the meta-analysis itself should be used only as ballpark estimates, considering the sample size and the fact that dissimilar studies and variables were combined in producing the results. Second, the behavioural process under examination is slightly different. The commute distance is taken as exogenous in the BEATIM model, but most of the estimates in the meta-analysis are from direct regressions between VMT and the built environment features and therefore include the effects of the built environment on commute distance. Therefore, caution needs to be taken in interpreting and generalising the results. However, rather than omitting this comparison, I aim in this analysis to seed the cross-regional study of built environment and travel, expecting that others would expand and strengthen the results over time.

Besides the meta-analysis results on total VMT presented in Section 2.3, I also provide extra results on commute and non-commute VMT so that the comparison can be made in more detail. However it should be noted that many of the extra results are based on very small sample sizes, or even only one study. Therefore these extra results are just presented for the readers' reference.

Generally speaking, the effects of population density, job density and retail accessibility are similar between the meta-analysis on American and European cities and my study on Beijing (differences between the effect sizes <= 0.03). The effects of population density and job density in Beijing are slightly smaller than those in the meta-analysis. It corresponds to the finding of Eom and Cho (2015) in Seoul that the impact of higher density on reducing car use is greatly reduced when the gross density is already high, which is the situation in many Asian cities (Eom & Cho, 2015). Diversity shows a larger effect in Beijing, which is mainly accounted for by its effect in reducing commute car use. Transit accessibility also shows a larger effect, indicating that improving the level of service of public transit can induce more returns in Beijing. Street density, however, shows opposite effects in American and European cities and in Beijing. The direction of influence from the meta-analysis is consistent with the theoretical assumption while that in Beijing is reversed, as discussed in the last section. But since the results on American and European cities are from only one study, this finding cannot be taken as quite concrete. Possible explanation to this difference can also be associated with the culture of 'car pride' and the preference for driving in China, in which situation the increased streets are more likely to be taken advantage for driving instead of for walking or cycling.

In summary, this preliminary cross-regional comparison shows that the impacts of the built environment on VMT are neither perfectly consistent nor completely different in different urban contexts. The effects of density-related features are more similar than transport infrastructure-related features in the comparison between American and European cities and Beijing. Possible explanations for this difference could lie in people's preference towards different travel modes, which can be further associated with the social cultural contexts.

Table 6-2 Comparison between the impacts of built environment features on VMT

Built environment features	American and European cities			Beijing		
	Total	**Commute**	**Non-commute**	**Total**	**Commute**	**Non-commute**
Population density	-0.04	NA	-0.09[a]	-0.01	0	-0.06
Job density	-0.03	NA	-0.23[b]	-0.01	0.02	-0.14
Diversity	-0.07	-0.07[c]	-0.05[c]	-0.15	-0.21	0.08
Transit accessibility	-0.05	NA	NA	-0.18 (bus) -0.06 (sub way)	-0.19 (bus) -0.10 (subway)	-0.11 (bus) -0.05 (subway)
Retail accessibility	-0.01[d]	NA	-0.17[e]	-0.03	-0.01	-0.14
Street density	-0.04[f]	-0.06[g]	-0.12[g]	0.01	-0.01	0.10

a Based on the works of Boarnet et al., 2004, Chatman, 2003, Chatman, 2008, Salon, 2015.
b Based on the works of Boarnet et al., 2004, Chatman, 2003, Chatman, 2008.
c Only the work of Salon, 2015 is used.
d Only the work of Cervero & Duncan, 2006 is used.
e Only the work of Bhat & Eluru, 2009 is used.
f Only the work of Hedel & Vance, 2007 is used.
g Only the work of Salon, 2015 is used.

6.6 Policy implications

The making and reshaping of the built environment is closely related to the policies of urban planning and management. The results could help to inform policy debate and encourage more effective critical thinking about spatial processes and impacts, and alternative policy scenarios (Wong, Baker, Webb, Hincks, & Schulze-Baing, 2015). Many policy implications can be drawn from the simulation results. Basically, the results provide evidence for the effectiveness of planning measures in affecting the travel behaviour of city residents and the potential of various planning strategies in alleviating various transport-related urban problems. First of all, a toolkit of planning measures for various policy goals can be derived from the simulation results:

- For the goal of reducing total car use, effective measures include increasing the retail density, the mix of uses and the accessibility to sub-centres, enhancing the coverage of bus services and improving the quality and continuity of street facades (effective measures here refer to those with an effect size larger than 0.05).
- For the goal of reducing non-commute car use, effective measures include increasing the population density, employment density, retail density and the accessibility to sub-centres, enhancing the coverage of bus services, decreasing the parking space and improving the continuity of street facade.
- For the goal of reducing the total travel distance needed for non-commute purposes, effective measures include increasing the population density, enhancing the coverage of bus services and improving the continuity of street facade.
- For other policy goals, one can refer to **Figure 6-2** to **Figure 6-15**.

It should be noted that there can be substantial differences in the effects of built environment measures that belong to a same type (e.g. increasing density), therefore policies need to be specific enough to be effective. For instance, the four density features all have different effects on travel behaviour. For another instance, bus coverage shows a larger effect in both reducing the car use and travel distance than the distance to subway station, though they are both indicators of public transit accessibility. Therefore, it is not enough to simply state ‘high density’ or ‘good accessibility to public transit’ as a planning measure. Instead, policies should fully refer to the detailed findings and be effectively specific.

Moreover, policies could also aim at the mediating factors that intervene the relationship between the built environment and travel. For instance, the comparison between the simulation results with the theoretical assumptions and meta-analysis results suggest that the preference towards driving and the culture of ‘car pride’ might be a reason for the positive correlation between road density and car use in Beijing. For another example, the quality of street facade is positively correlated with the travel distances of non-commute activities, which is supposed to be mediated by the fact that high quality areas are usually more private and gated and thus less convenient for conducting activities. Besides, employment density is also found to be positively associated with the travel distances of most non-commute activities, especially shopping, possibly explained by the mismatch between the types of goods and services at business areas and the everyday needs. Policies that aim at these mediating conditions include:

- Alter people's preference towards driving and the thinking of 'car pride' by improving the pedestrian environment and improving the image of walking and cycling;
- Increase the space for street shops in the areas where the street facade is good in quality but does not provide many activity opportunities;
- Increase the number of facilities that serve everyday needs, probably in medium-to low-price, in business areas.

However, since the effect sizes are generally small, policy making should also consider a cost-benefit analysis to ascertain whether changes to the built environment are a cost-effective way to modify travel behaviour, given the opportunity costs of spending resources in another way (Mokhtarian & Cao, 2008). For instance, policy makers should consider the energy consumption or carbon emission in the process of the deconstruction and reconstruction of buildings and other structures in order to realise a built environment change, and compare with the amount of energy and carbon emission saving in a given period.

Last, the differences between the results from Beijing and other cities suggest that special care needs to be taken when transferring the above mentioned policies to elsewhere. One needs to closely scrutinise the urban and social contexts and evaluate whether there is any factor, such as the mediating factors mentioned above, that would distort the relationship between the built environment and travel in another city.

6.7 Chapter summary

This chapter sets out to simulate the changes of travel behaviour in response to various scenarios of built environment changes using the BEATIM model. Many of the conclusions drawn from the simulations can be identified through careful observation of the diagrams in **Figure 6-2** to **Figure 6-19**. A key note is that the effects of the built environment on VMT, which is the subject of analysis of many existing research, are to a large extent accounted for by the effects on commute travel. As a result, the relationship between the built environment and non-commute and other aspects of daily travel would be masked if only this synthesised indicator is used. For some built environment features that show similar impacts on VMT, their impacts on detailed behavioural aspects can be very different, e.g. on the mode choices for commute and non-commute activities, the travel distances for various non-commute purposes, etc. Besides, both commute and non-commute travel is shown to be more sensitive to the built environment in the near neighbourhood of one's home (in my experiment, 500 metre buffer zone), when the work place is taken as exogenous.

The simulation results are partly consistent with theoretical assumptions and partly not. The comparison with the meta-analysis also shows that the impacts of the built environment are neither perfectly consistent nor completely different in various urban contexts. Four major implications can be made from the inconsistent results: (1) whether higher density relates to enhanced travel gains and thus shorter travel distance could depend on the matchness between the types of density and people's needs; (2) social cultural factors (in the case of Beijing, the 'car pride') can play a non-negligible role in shaping the (dis)utility of travel choices and distort the relationship between the built environment and travel; (3) in the context of Beijing, high (construction and maintenance) quality of street facade can related to lower utilitarian values, when that happens, utilitarian considerations tend to overweigh the psychic enjoyments, thus making a location less attractive; (4) the ‘compensation’ mechanism between travel distance and frequency does exist, but is not likely to be stronger than the original effect.

Chapter 07 Conclusions and final remarks

7.1 Summary of findings

This work sets out to investigate how people's daily travel behaviour would be influenced by the built environment conditions. Travel utility maximisation is used as the theoretical base for the possible influences and a series of assumptions on the relationship between various built environment features and travel behaviour. It is argued in the first chapter that despite of a large number of studies on this topic, there still exist many research gaps. The gaps include:

- First of all, a major gap lies in that a large proportion of existing research focus on the synthesised outcomes of the complex process of travel decision making, while the behavioural processes that give rise to these outcomes have received much less attention. It is related to the gap in methodology that many existing research use regressions between the synthesised outcomes and a set of socioeconomic and built environment explanatory variables, which usually cannot probe into the detailed behavioural processes.
- Second, there is a lot of inconsistency in existing findings in terms of the directions and sizes of the influences, which undermines the reliability and generalisability of the findings.
- Third, the built environment features that have been studied are mainly two-dimensional and land use-related. The features related to the dimension of street facade have received much less attention.
- Last, most existing studies are based on American and European (plus a few Oceanian) cities, while evidences from Asian cities are relatively scarce.

In order to address the first gap, an overarching methodology is designed by linking the activity-based modelling approach, which is mainly developed in the field of transport simulation, with the analysis of the built environment-travel relationship. Activity-based models simulate the full process of decision making in daily activity participation and travel, including which activities are conducted when, where, for how long, and the transport mode involved. Although the development of activity-based models has progressed substantially since the 1990s, the built environment factors are seldom sufficiently account for in the model systems. Therefore, this research is novel in developing an activity-based model that fully takes into account the built environment contexts and using this model to scrutinise the impacts of the built environment on travel behaviour at much greater detail.

After the introduction, Chapter 2 provides a review of related theoretical and empirical works and partly addresses the second gap with a meta-analysis of existing works. A comprehensive conceptual framework is developed on the relationship between the built environment and the travel costs and gains. Based on this, a series of assumptions are made regarding to the influences of various built environment changes on the integrated outcomes of activity-travel (e.g. total travel distance, total car use) based on utility maximisation. The review of empirical studies and activity-based model developments provides evidences for the gaps mentioned above. Besides, the meta-analysis shows that the effect sizes of the built environment are more consistent across studies on VMT than on walking and transit use. Therefore, the results on VMT are used to compare with my own findings in Beijing, from which the second gap can be further addressed.

Chapter 3 and Chapter 4 describe the study area and the process of data collection and pre-processing. Particularly, Chapter 4 deals with the third gap that features related to the dimension of street facade are seldom studied. I proposed a novel method that automatically evaluate the street facade in a large-scale by leveraging state-of-the-art machine learning techniques and online street view images. Two specific features are selected based on architecture and urban design theories, which are the construction and maintenance quality of building facade and the continuity of street wall. The performance indicators show that the machine learning models are able to produce acceptably good approximation to the expert ratings.

The data pre-processed in Chapter 3 and 4 are fed to Chapter 5, which develops the BEATIM model. The model generally takes the paradigm of utility-maximising econometric models, coupled with weak features of computational process models. Special care is taken to keep close focus on the daily travel behaviour and the influence of the built environment when building the model. The model system contains four major components: namely the sub-models for the activity participation and organisation, the location choice for primary activities, the time of travel and mode choice, and the location choice for intermediate stops. The validation shows that the model is able to provide a reasonably good prediction of people's daily travel. It is acknowledged that there can be many prediction errors at the individual level due to the complex, stochastic nature of activity-travel behaviour (Kulkarni & McNally, 2000). However, the correlation between simulation results and observed travel behaviour at more aggregate levels, such as by ring roads, can be high (R^2=0.8-1). To the best of the author's knowledge, it is the most comprehensive model that explicitly links the activity-based modelling approach from the field of transport simulation with the analysis of the built environment-travel relationship in the field of urban planning and design.

Scenario analysis is conducted in Chapter 6 with the BEATIM model. Two types of scenarios are designed, namely local scenarios and regional scenarios. The former analyse the impacts of the built environment in the immediate neighbourhood of one's home (the home TAZ). The latter explore the impacts of the built environment in the 0-to-2000-metre buffer zones from one' home. The findings from the scenario analysis include

- In the case of Beijing, total VMT is prominently affected by land use mix, bus coverage, facade quality and facade continuity.
- The influences of the built environment on total VMT are to a large extent accounted for by the influences on commute VMT. Therefore, a sole focus on this indicator, which is the case in many existing research, could mask the understanding of the influence on many other aspects of daily travel.
- Both commute and non-commute travel are more sensitive to the built environment in proximity to home place (in my experiment, 500 metre buffer zone), if the work place is taken as exogenous.
- For a full description of the effects of the built environment on detailed aspects of activity-travel, please refer to **Figure 6-2** to **Figure 6-15**.

The simulation results are partly consistent with the theoretical assumptions put forward in Chapter 2 and partly not. The comparison with the meta-analysis also shows that the impacts of the built environment are neither perfectly consistent nor completely different between Beijing and European and American cities. The implications include

- Whether higher density relates to enhanced travel gains and thus shorter travel distance could depend on the matchness between the types of density and people's needs.
- Social cultural factors (in the case of Beijing, the 'car pride') can play a non-negligible role in shaping the (dis)utility of travel choices and distort the relationship between the built environment and travel.
- In the context of Beijing, high street facade quality can related to lower utilitarian value, when that happens, utilitarian considerations tend to overweigh the psychic enjoyments, thus making a location less attractive.
- The 'compensation' mechanism between travel distance and frequency does exist, but is not likely to be stronger than the original effect.

It should be noted that the first three implications are based on assumed explanations for the results that are inconsistent with theoretical assumptions and existing findings. Further empirical evidence is needed to testify these explanations, which is beyond the scope of this research but can be an important topic for future research.

7.2 Limitations and future research

The BEATIM model developed in this research can work as a helpful tool in answering a multitude of questions about the influence of the built environment on travel behaviour. However, there remain many ways to refine or extend the model so that the complex nature of the urban system and travel behaviour can be better understood.

First, this research started at a time when urban big data began to emerge and be applied in the field of urban studies. In this research, a few types of big data are applied in the measurement of several built environment features, including the street view images for the measurement of the street facade, the Point of Interest (POI) data for the measurement of the density of facilities, etc. However, the data source for people's travel behaviour is still the conventional travel diary survey. The survey data are limited in several ways: (1) they rely on the self-report of the interviewees, which can be subject to dishonesty, retrospective errors, misunderstanding and other types of errors (Hoskin, 2012); (2) the locations of activities are recorded in the spatial unit of TAZs instead of the exact coordinates; (3) the travel routes are not recorded; (4) constrained by the costs and willingness of the interviewees, only a one-day travel diary is recorded, in which occasional activities may be underrepresented. With the growing availability of human mobility data (e.g. the cellular network data), the information on people's travel behaviour can be obtained at a finer scale and over a longer period with a low cost (although this type of data may not contain the rich socioeconomic information as in the travel survey). The high spatial resolution can enable a more precise analysis on not only the travel behaviour but also the built environment. For instance, the modelling of mode choice can also take into consideration the built environment conditions along the route.

Besides, the quality of some of the data used this research is constantly improving and more data sources are emerging in recent years, so that the measurement of the built environment can be further improved. For instance, the POI data can now be combined with the data of customers' ratings from the Dazhongdianping website (similar to Foursquare). As a result, I am able to measure not only the number of various types of facilities, but also the quality of the goods and services, which could further enhance the behavioural realism of the model.

As mentioned before, the BEATIM model focuses on the short-term decisions in the complex system of land use and transport. The model can be extended both upwards and downwards to include other relevant longer-term or shorter-term interactions within this system, so that more comprehensive examination can be conducted on the interplay between the built environment and travel. For instance, the residence and work locations are taken as pre-determined and exogenous in my model. Future research could incorporate sub-models of home and work location choices, to take into account the long-term interactions between the built environment and the spatial distribution of households and businesses. Similarly, the expected travel time between locations, which is now taken as exogenous, can also be replaced with real-time travel time estimated from a module of route assignment and traffic flow, which corresponds to the 'very short-term' changes in the LUTI system.

Last, the cross-regional differences in the built environment-travel relationship are subject to further research. Both the review of existing findings and the findings from this research suggest that there are non-negligible differences between the results derived from different urban contexts. A better understanding on the factors that give rise to these differences could contribute to a more systematic understanding of this issue and enhance the generalisability of individual studies.

Appendix A Results from individual studies in the meta-analysis

Table A-1 Elasticity of VMT with respect to density

Study	N	y	x	e	In meta-analysis?
(Zegras, 2010)[a]	14,729	Household VKT	Dwelling unit density	-0.04	Y
(Guerra, 2014)	20,075	Latent VKT (1994)	Population (100s) per hectare	-0.21***	
(Guerra, 2014)	33,282	Latent VKT (2007)	Population (100s) per hectare	-0.31***	
(Guerra, 2014)	20,075	Latent VKT (1994)	Jobs (100s) per hectare	-0.11***	
(Guerra, 2014)	33,282	Latent VKT (2007)	Jobs (100s) per hectare	-0.11***	
(Guerra, 2014)	20,075	VKT (if any) (1994)	Population (100s) per hectare	-0.04**	Y
(Guerra, 2014)	33,282	VKT (if any) (2007)	Population (100s) per hectare	-0.04**	Y
(Guerra, 2014)	20,075	VKT (if any) (1994)	Jobs (100s) per hectare	-0.01 Ψ	Y
(Guerra, 2014)	33,282	VKT (if any) (2007)	Jobs (100s) per hectare	-0.01 Ψ	Y
(Ewing et al., 2015)†	58,011	Household VMT (if any)	Activity density within one mile (pop + emp per square mile in 1000s)	-0.05*	Y
(Salon, 2015)[b]	130,901	Weekday nonwork VMT	Population density	-0.03**	

Ψ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$

† Elasticities calculated by myself.

a Elasticities from this paper are combined elasticities for car ownership and car use via simulation and therefore do not contain a significance level.

b Only the elasticities of significant variables are provided in this paper.

Table A-2 Elasticity of VMT with respect to diversity

Study	N	y	x	e	In meta-analysis?
(Zegras, 2010)	14,729	Household VKT	Diversity index	-0.01	Y
(Guerra, 2014)	20,075	Latent VKT (1994)	Destination diversity	0.17***	
(Guerra, 2014)	33,282	Latent VKT (2007)	Destination diversity	0.17***	
(Guerra, 2014)	20,075	VKT (if any) (1994)	Destination diversity	-0.06***	Y
(Guerra, 2014)	33,282	VKT (if any) (2007)	Destination diversity	-0.06***	Y
(Ewing et al., 2015)†	58,011	Household VMT (if any)	Job-population balance within one-quarter mile	-0.03*	Y
(Ewing et al., 2015)†	58,011	Household VMT (if any)	Land use entropy within one mile	-0.10***	Y
(Salon, 2015)	130,901	Weekday nonwork VMT	Activity mix at home	-0.05**	
(Salon, 2015)	60,346	One-way commute VMT	Activity mix at home	-0.07**	
(Salon, 2015)	60,346	One-way commute VMT	Activity mix at work	0.36**	

Ψ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$

† Elasticities calculated by myself

Table A-3 Elasticity of VMT with respect to destination accessibility

Study	N	y	x	e	In meta-analysis?
(Zegras, 2010)	14,729	Household VKT	Distance to CBD	-0.23[a]	Y
(Guerra, 2014)	20,075	Latent VKT (1994)	Car accessibility (the number of jobs accessible by car to a household weighted by a negative exponential decay function for travel time)	0.38***	

(Guerra, 2014)	33,282	Latent VKT (2007)	Car accessibility (the number of jobs accessible by car to a household weighted by a negative exponential decay function for travel time)	0.28***	
(Guerra, 2014)	20,075	VKT (if any) (1994)	Car accessibility (the number of jobs accessible by car to a household weighted by a negative exponential decay function for travel time)	-0.27***	Y
(Guerra, 2014)	33,282	VKT (if any) (2007)	Car accessibility (the number of jobs accessible by car to a household weighted by a negative exponential decay function for travel time)	-0.21***	Y
(Guerra, 2014)	20,075	Latent VKT (1994)	Kilometres (10s) to Zocalo	-0.20***	Y
(Guerra, 2014)	33,282	Latent VKT (2007)	Kilometres (10s) to Zocalo	-0.22***	Y
(Ewing et al., 2015)†	58,011	Household VMT (if any)	Percentage of regional employment within 10 min by car	-0.05***	Y
(Ewing et al., 2015)†	58,011	Household VMT (if any)	Percentage of regional employment within 30 min by transit	-0.07***	Y
(Salon, 2015)	130,901	Weekday nonwork VMT	Regional job access at home	0.07**	
(Salon, 2015)	130,901	Weekday nonwork VMT	Local job access at home	-0.06**	
(Salon, 2015)	60,346	One-way commute VMT	Local job access at home	-0.16**	
(Salon, 2015)	60,346	One-way commute VMT	Regional job access at home	0.15**	

Ψ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$

† Elasticities calculated by myself.

a Sign reversed.

Table A-4 Elasticity of VMT with respect to road network

Study	N	y	x	e	In meta-analysis?
(Zegras, 2010)	14,729	Household VKT	4-way intersections per km	-0.02	Y
(Zegras, 2010)	14,729	Household VKT	3-way intersections per km	0.14	Y
(Guerra, 2014)	20,075	Latent VKT (1994)	Intersections per hectare	-0.07**	Y
(Guerra, 2014)	33,282	Latent VKT (2007)	Intersections per hectare	-0.06**	Y
(Ewing et al., 2015)†	58,011	Household VMT (if any)	Intersection density within one mile	-0.21***	Y
(Ewing et al., 2015)†	58,011	Household VMT (if any)	Percentage 4-way intersections within one mile	-0.06***	Y
(Salon, 2015)	130,901	Weekday nonwork VMT	Road density at home	-0.12**	
(Salon, 2015)	60,346	One-way commute VMT	Road density at home	-0.06**	

Ψ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$

† Elasticities calculated by myself.

Table A-5 Elasticity of VMT with respect to public transport service

Study	N	y	x	e	In meta-analysis?
(Zegras, 2010)	14,729	Household VKT	Distance to Metro	-0.20[a]	Y
(Guerra, 2014)	20,075	Latent VKT (1994)	Within a half kilometre of the metro	-0.02***	Y
(Guerra, 2014)	33,282	Latent VKT (2007)	Within a half kilometre of the metro	-0.02***	Y

Ψ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$

a Sign reversed.

Table A-6 Elasticity of walk trips with respect to density

Study	N	y	x	e	In meta-analysis?
(Boarnet, Joh, et al., 2011)†	1,370	Walking trips per person per day	Residential units per acre	-0.78	
(Boarnet, Joh, et al., 2011)†	1,370	Walking trips per person per day	Neighbourhood business per acre	0.59	Y
(Joh et al., 2012)†	825	Walking trips per person per day ('high-walk' attitude group)	Business per acre	0.23*	Y
(Joh et al., 2012)†	605	Walking trips per person per day ('low-walk' attitude group)	Business per acre	0.27	Y
(Witten et al., 2012)†	1,315	Self-report time of leisure physical activity (if any)	Dwelling density	0.19	
(Witten et al., 2012)†	1,575	Self-report time of walking (if any)	Dwelling density	0.17	
(Witten et al., 2012)†	1,619	Accelerometer-measured counts of physical activity (weekday)	Dwelling density	0.15*	
(Witten et al., 2012)†	1,512	Accelerometer-measured counts of physical activity (weekend)	Dwelling density	0.13*	
(Song et al., 2013)	2,676	Ratio of active travel time to total time (obligatory trips)	Population density	0.17**	Y
(Song et al., 2013)	2,676	Ratio of active travel distance to total distance (obligatory trips)	Population density	0.18**	Y
(Song et al., 2013)	3,309	Ratio of active travel time to total time (discretionary trips)	Population density	0.05	Y
(Song et al., 2013)	3,309	Ratio of active travel distance to total distance (discretionary trips)	Population density	0.03	Y

Study	N	y	x	e	In meta-analysis?
(Lee et al., 2014)	6,246	Walk/bike mode choice for home-based work trips	Population density in the one-quarter mile buffer area of D	0.31	Y
(Lee et al., 2014)	6,246	Walk/bike mode choice for home-based work trips	Employment density in the one-quarter mile buffer area of O	0.35	Y
(Lee et al., 2014)	6,246	Walk/bike mode choice for home-based work trips	Employment density in the one-quarter mile buffer area of D	0.13	Y
(Lee et al., 2014)	10,413	Walk/bike mode choice for home-based other trips	Population density in the one-quarter mile buffer area O	0.29	Y
(Lee et al., 2014)	10,413	Walk/bike mode choice for home-based other trips	Population density in the one-quarter mile buffer area D	0.48	Y
(Ewing et al., 2015)†	14,627	Household walk trips (if any)	Activity density within one-quarter mile (pop + emp per square mile in 1000s)	0.01 Ψ	Y

Ψ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$

† Elasticities calculated by myself

Table A-7 Elasticity of walk trips with respect to diversity

Study	N	y	x	e	In meta-analysis?
(Witten et al., 2012)†	1,315	Self-report time of leisure physical activity	Mixed land use	0.27	
(Witten et al., 2012)†	1,575	Self-report time of walking	Mixed land use	0.22	
(Witten et al., 2012)†	1,619	Accelerometer-measured counts of physical activity (weekday)	Mixed land use	0.08	
(Witten et al., 2012)†	1,512	Accelerometer-measured counts of physical	Mixed land use	0.11	

Study	N	y	x	e	In meta-analysis?
		activity (weekend)			
(Song et al., 2013)	2,676	Ratio of active travel time to total time (obligatory trips)	Land-use balance (inverse of the distance between each area's functionality mix and the national norm in a geometric space)	0.13	Y
(Song et al., 2013)	2,676	Ratio of active travel distance to total distance (obligatory trips)	Land-use balance (same as above)	0.11**	Y
(Song et al., 2013)	3,309	Ratio of active travel time to total time (discretionary trips)	Land-use balance (same as above)	0.01	Y
(Song et al., 2013)	3,309	Ratio of active travel distance to total distance (discretionary trips)	Land-use balance (same as above)	0.07	Y
(Song et al., 2013)	2,676	Ratio of active travel time to total time (obligatory trips)	Neighbours' land-use balance (same as above)	0.35*	Y
(Song et al., 2013)	2,676	Ratio of active travel distance to total distance (obligatory trips)	Neighbours' land-use balance (same as above)	0.40**	Y
(Song et al., 2013)	3,309	Ratio of active travel time to total time (discretionary trips)	Neighbours' land-use balance (same as above)	0.02	Y
(Song et al., 2013)	3,309	Ratio of active travel distance to total distance (discretionary trips)	Neighbours' land-use balance (same as above)	0.09	Y
(Lee et al., 2014)	6,246	Walk/bike mode choice for home-based work trips	Entropy index at O in the one-quarter mile buffer area	1.15	Y
(Lee et al., 2014)	10,413	Walk/bike mode choice for home-based other trips	Dissimilarity index at O in the one-quarter mile buffer area	0.16	Y
(Lee et al., 2014)	10,413	Walk/bike mode choice for home-based other trips	Dissimilarity index at D in the one-quarter mile buffer area	0.91	Y

Study	N	y	x	e	In meta-analysis?
(Ewing et al., 2015)†	14,672	Household walk trips (if any)	Land use entropy within one-half mile	0.10***	Y

Ψ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$

† Elasticities calculated by myself using formula in Table

Table A-8 Elasticity of walk trips with respect to accessibility

Study	N	y	x	e	In meta-analysis?
(Witten et al., 2012)†	1,315	Self-report time of leisure physical activity (if any)	Neighbourhood Destinations Accessibility Index (density of facilities)	0.31*	
(Witten et al., 2012)†	1,575	Self-report time of walking (if any)	Neighbourhood Destinations Accessibility Index (density of facilities)	0.31	
(Witten et al., 2012)†	1,619	Accelerometer-measured counts of physical activity (weekday)	Neighbourhood Destinations Accessibility Index (density of facilities)	0.17*	
(Witten et al., 2012)†	1,512	Accelerometer-measured counts of physical activity (weekend)	Neighbourhood Destinations Accessibility Index (density of facilities)	0.12	
(Song et al., 2013)	3,309	Ratio of active travel time to total time (discretionary trips)	Retail centre–home distance	-0.31**	
(Song et al., 2013)	3,309	Ratio of active travel distance to total distance (discretionary trips)	Retail centre–home distance	-0.46**	
(Cao, 2015a)[a]	1,194	Probability of conducting 'a lot more walking'	Change in accessibility (from principal axis factoring)	0.04	
(Ewing et al., 2015)†	14,672	Household walk trips (if any)	Percentage of regional employment within 30 min by transit	0.07*	

Ψ *p* < .10, * *p* < .05, ** *p* < .01, *** *p* < .001

† Elasticities calculated by myself

a This study employed a 5-point ordinal scale of changes in walking and biking after relocation, from ‘a lot less now’ to ‘a lot more now’. Since these responses are interrelated, only the elasticities on the probability of choosing ‘a lot more’ is included.

Table A-9 Elasticity of walk trips with respect to road network design

Study	**N**	**y**	**x**	**e**	**In meta-analysis ?**
(Boarnet, Joh, et al., 2011)†	1,370	Walking trips per person per day	Percentage intersections 4-way	0.13	Y
(Joh et al., 2012)†	825	Walking trips per day (‘high-walk’ attitude group)	Intersection density	0.47	Y
(Joh et al., 2012)†	605	Walking trips per day (‘low-walk’ attitude group)	Intersection density	-2.54*	Y
(Joh et al., 2012)†	825	Walking trips per day (‘high-walk’ attitude group)	Four-way intersections	-0.04	Y
(Joh et al., 2012)†	605	Walking trips per day (‘low-walk’ attitude group)	Four-way intersections	-0.28	Y
(Witten et al., 2012)†	1,315	Self-report time of leisure physical activity (if any)	Street connectivity (number of intersections with ⩾ 3 intersecting streets per square kilometre within a meshblock)	0.30*	
(Witten et al., 2012)†	1,575	Self-report time of walking (if any)	Street connectivity (same as above)	0.14	
(Witten et al., 2012)†	1,619	Accelerometer-measured counts of physical activity (weekday)	Street connectivity (same as above)	0.16*	

Study	N	y	x	e	In meta-analysis?
(Witten et al., 2012)†	1,512	Accelerometer-measured counts of physical activity (weekend)	Street connectivity (same as above)	0.16*	
(Lee et al., 2014)[a]	10,413	Walk/bike mode choice for home-based other trips	Total roadway length divided by total area in the one-quarter mile buffer area at D	0.69	Y
(Lee et al., 2014)	10,413	Walk/bike mode choice for home-based other trips	Number of intersections divided by total number of intersections and dead ends in the one-quarter mile buffer area at O	0.99	Y
(Ewing et al., 2015)†	14,672	Household walk trips (if any)	Percentage 4-way intersections within one quarter mile	0.03**	Y

Ψ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$

† Elasticities calculated by myself

a Only significant elasticities (at least at 0.1 level) are provided in this paper.

Table A-10 Elasticity of walk trips with respect to public transport service

Study	N	y	x	e	In meta-analysis?
(Cao, 2015a) †	1,194	Probability of conducting ‘a lot more walking’	Change in transit (from principal axis factoring)	0.06	
(Ewing et al., 2015)†	14,672	Household walk trips (if any)	Transit stop density within one-half mile	0.04***	Y

Ψ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$

† Elasticities calculated by myself.

Table A-11 Elasticity of transit trips with respect to density

Study	N	y	x	e	In meta-analysis?
(Lee et al., 2014)	6,246	Transit mode choice for home-based work trips	Population density at D in the one-quarter mile buffer	0.28	Y
(Lee et al., 2014)	6,246	Transit mode choice for home-based work trips	Employment density at O in the one-quarter mile buffer	0.35	Y
(Lee et al., 2014)	6,246	Transit mode choice for home-based work trips	Employment density at D in the one-quarter mile buffer	0.22	Y
(Lee et al., 2014)	10,413	Transit mode choice for home-based other trips	Population density at O in the one-quarter mile buffer	0.24	Y
(Lee et al., 2014)	10,413	Transit mode choice for home-based other trips	Population density at D in the one-quarter mile buffer	0.44	Y

Ψ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$

Table A-12 Elasticity of transit trips with respect to diversity

Study	N	y	x	e	In meta-analysis?
(Lee et al., 2014)	6,246	Transit mode choice for home-based work trips	Entropy index at O in the one-quarter mile buffer	1.19	Y
(Lee et al., 2014)	10,413	Transit mode choice for home-based other trips	Dissimilarity at O in the one-quarter mile buffer	0.14	Y
(Lee et al., 2014)	10,413	Transit mode choice for home-based other trips	Dissimilarity at D in the one-quarter mile buffer	0.83	
(Ewing et al., 2015)†	6,719	Household transit trips (if any)	Land use entropy within one-half mile	0.07	Y
(Ewing et al., 2015)†	6,719	Household transit trips (if any)	Job-population balance within one mile	0.12*	

Ψ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$

† Elasticities calculated by myself.

Table A-13 Elasticity of transit trips with respect to road network design

Study	N	y	x	e	In meta-analysis?
(Lee et al., 2014)	10,413	Transit mode choice for home-based other trips	Total roadway length divided by total area in the one-quarter mile buffer area at D	0.60	Y
(Lee et al., 2014)	10,413	Transit mode choice for home-based other trips	Number of intersections divided by total number of intersections and dead ends in the one-quarter mile buffer area at O	0.85	Y

Ψ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$

Table A-14 Elasticity of transit trips with respect to destination accessibility

Study	N	y	x	e	In meta-analysis?
(Ewing et al., 2015)†	6,719	Household transit trips (if any)	Percentage of regional employment within 30 min by transit	0.04*	

Ψ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$

† Elasticities calculated by myself.

Appendix B Summary of built environment features

Table B-1 Descriptive statistics of built environment features

	All		**Within 2nd ring road**		**2nd to 3rd ring road**		**3rd to 4th ring road**		**4th to 5th ring road**	
	Mean **Min**	**SD** **Max**	**Mean** **Min**	**SD** **Max**	**Mean** **Min**	**SD** **Max**	**Mean** **Min**	**SD** **Max**	**Mean** **Min**	**SD** **Max**
Population density	15435	9478	21643	7656	22114	8478	17709	8634	9361	6630
	1396	40500	11450	40500	3169	40363	3169	40363	1396	32305
Employment density	10105	11456	14856	9220	16975	13613	12567	13714	4260	5034
	343	72114	2817	33475	1702	72114	1249	72114	343	38597
Distance to the city center	8882	3622	3042	1131	5976	1121	8503	1482	12109	1853
	409	16890	409	5120	3676	9123	5832	12044	8502	16890
Distance to the nearest commercial cluster	1540	1295	888	507	1054	582	1121	614	2167	1633
	30	7319	178	2760	161	2674	51	2977	30	7319
Retail density	41	40	71	60	56	40	46	33	23	26
	0	389	0	389	0	213	0	226	0	152
Entertainment density	35	34	54	29	54	35	42	33	18	24
	0	247	0	130	0	179	1	247	0	142
Land use mix	1.01	0.37	1.20	0.28	1.15	0.32	1.08	0.34	0.86	0.38
	0	1.74	0.33	1.64	0.00	1.74	0.00	1.66	0	1.72
Primary road density	1416	1251	1863	1346	1825	1339	1587	1155	1009	1098
	0	7274	0	7273	0	6176	0	7274	0	5632
Secondary road density	4068	2489	4205	1943	4820	2113	4988	2559	3206	2476

	All		Within 2nd ring road		2nd to 3rd ring road		3rd to 4th ring road		4th to 5th ring road	
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
	Min	**Max**	**Min**	**Max**	**Min**	**Max**	**Min**	**Max**	**Min**	**Max**
	0	15623	0	9511	0	14705	0	15623	0	13794
Tertiary road density	2981	3167	8150	3767	3594	2445	2037	2072	1697	1962
	0	17235	2	17235	0	16110	0	14764	0	13056
Parking density	1145	1968	1612	2085	1804	2475	1650	2474	446	835
	0	14668	0	9351	0	14515	0	14668	0	5709
Distance to the nearest subway station	1381	1063	660	372	952	624	1135	791	1915	1214
	25	5863	236	1794	86	3276	25	3445	158	5863
Bus coverage	0.46	0.23	0.65	0.16	0.57	0.20	0.53	0.17	0.32	0.21
	0.00	1.00	0.17	1.00	0.08	0.95	0.11	0.95	0.00	0.94
Facade quality	1.47	0.16	1.42	0.14	1.51	0.12	1.53	0.13	1.45	0.18
	0.87	1.92	1.11	1.75	1.11	1.87	1.04	1.90	0.87	1.93
Facade continuity	0.18	0.08	0.27	0.10	0.19	0.06	0.19	0.07	0.15	0.07
	0.01	0.52	0.04	0.52	0.04	0.40	0.04	0.43	0.01	0.36

Table B-2 Correlation matrix of built environment features

	Den_ pop	Den_ emp	Dist_ city	Dist_ comclu	Den_ retail	Den_ ent	Mix	Den_ road _pri	Den_ road _sec	Den_ road _ter	Den_ parking	Dist_ sub	Cov_ bus	Facade_ qua	Facade_ conti
Den_pop	**1**	**0.43**	**-0.58**	**-0.48**	**0.37**	**0.48**	**0.41**	**0.13**	**0.35**	**0.35**	**0.25**	**-0.46**	**0.56**	**0.24**	**0.24**
Den_emp	**-**	**1**	**-0.47**	**-0.37**	**0.43**	**0.62**	**0.31**	**0.25**	**0.42**	**0.24**	**0.53**	**-0.41**	**0.44**	**0.26**	**0.14**
Dist_centre	**-**	**-**	**1**	**0.55**	**-0.47**	**-0.5**	**-0.31**	**-0.28**	**-0.29**	**-0.54**	**-0.32**	**0.52**	**-0.55**	-0.07	**-0.38**

Dist_comclu	**-**	**-**	**-**	**1**	**-0.43**	**-0.5**	**-0.23**	**-0.19**	**-0.39**	**-0.22**	**-0.3**	**0.39**	**-0.49**	**-0.26**	**-0.2**
Den_retail	**-**	**-**	**-**	**-**	**1**	**0.74**	**0.29**	**0.12**	**0.33**	**0.36**	**0.48**	**-0.31**	**0.52**	**0.14**	**0.23**
Den_ent	**-**	**-**	**-**	**-**	**-**	**1**	**0.36**	**0.18**	**0.49**	**0.35**	**0.56**	**-0.42**	**0.63**	**0.26**	**0.17**
Mix	**-**	**-**	**-**	**-**	**-**	**-**	**1**	**0.21**	**0.33**	**0.24**	**0.14**	**-0.21**	**0.49**	**0.2**	**0.12**
Den_road_pri	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**1**	**0.18**	**0.13**	**0.19**	**-0.19**	**0.24**	**0.13**	**0.22**
Den_road_sec	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**1**	0.07	**0.44**	**-0.41**	**0.56**	**0.33**	0.05
Den_road_ter	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**1**	**0.19**	**-0.3**	**0.35**	-0.02	**0.32**
Den_parking	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**1**	**-0.3**	**0.3**	**0.25**	**0.08**
Dist_sub	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**1**	**-0.47**	**-0.25**	**-0.21**
Cov_bus	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**1**	**0.24**	**0.22**
Facade_qua	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**1**	**-0.12**
Facade_conti	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**1**

Appendix C Supplementary results for Chapter 5

Table C-1 Distribution of activity purposes

Type	**Purpose**	**Percentage**
Commute	Work	21.7%
	School	3.7%
Non-commute	Sleep	0.1%
	Dine out	2.4%
	Personal business	2.7%
	Housekeeping	0.1%
	Entertain	6.8%
	Shopping	10.8%
	Visit friends	1.7%
	Escorting	4.6%
	Fetching goods	0.4%
	Others	0.6%
Back home	Back home	44%
Total		100.0%

Table C-2 Confusion matrix of the predictions of activity plans with and without built environment variables (two activities in the day)

Independent variables: socioeconomic, built environment (accuracy=61.2%)			
	h-d-h-d-h (obs)	h-d-s-h (obs)	h-s-d-h (obs)
h-d-h-d-h (sim)	3888	1522	806
h-d-s-h (sim)	170	222	92
h-s-d-h (sim)	21	18	44
Independent variables: socioeconomic only (accuracy=61.1%)			
h-d-h-d-h (sim)	3908	1539	829
h-d-s-h (sim)	156	213	88
h-s-d-h (sim)	15	10	25

Note: Multinomial logit model is used.

Table C-3 Confusion matrix of the predictions of activity plans with and without built environment variables (three activities in the day)

Independent variables: socioeconomic, built environment (accuracy=39.2%)							
	1 (obs)	2 (obs)	3 (obs)	4 (obs)	5 (obs)	6 (obs)	7 (obs)
1 (sim)	588	117	278	75	213	104	81
2 (sim)	3	4	2	6	1	2	1
3 (sim)	31	4	40	5	10	5	7
4 (sim)	19	7	14	37	7	19	13
5 (sim)	4	1	1	0	4	0	0
6 (sim)	84	54	45	58	50	220	63
7 (sim)	0	0	1	1	1	5	5
Independent variables: socioeconomic only (accuracy=39.2%)							
1 (sim)	625	123	326	81	233	108	83
2 (sim)	3	3	0	3	0	2	1
3 (sim)	4	0	3	1	1	2	1
4 (sim)	17	11	6	37	4	15	17
5 (sim)	1	0	0	0	0	0	0
6 (sim)	79	50	46	59	48	226	64
7 (sim)	0	0	0	1	0	2	4

Note: 1=h-d-h-d-h-d-h, 2=h-d-h-d-s-h, 3=h-d-s-h-d-h, 4=h-d-s-s-h, 5=h-s-d-h-d-h, 6= h-s-d-s-h, 7= h-s-s-d-h. Multinomial logit model is used.

Table C-4 AUCs for the prediction of whether a non-commute activity purpose is included in the activity plan with and without built environment variables (given that the number of non-commute activities > 0)

Activity types	**AUC (built environment variables included)**	**AUC (built environment variables not included)**
Shopping	0.56	0.55
Entertainment	0.52	0.60
Dining out	0.48	0.55
Personal business	0.51	0.51
Escorting/picking up/dropping off	0.47	0.55
Others	0.48	0.52

Note: Binary logistic regression is used.

Table C-5 Frequency distribution of non-commute activities

Purposes	**Percentages**
Shopping	36%
Entertainment	22%
Dining out	8%
Personal business	9%
Escorting	15%
Others	9%
Sum	*100%*

Table C-6 Frequency distribution of activity plans

# activities = 1 (62.2%)		# activities = 2 (24.9%)		# activities = 3 (7.2%)	
h-d-h	100%	h-d-h-d-h	68%	h-d-h-d-h-d-h	38%
Sum	*100%*	h-s-d-h	11%	h-d-s-s-h	9%
		h-d-s-h	21%	h-s-d-s-h	15%
		Sum	*100%*	h-s-s-d-h	7%
				h-s-d-h-d-h	15%
				h-d-s-h-d-h	16%
				Sum	*100%*

Table C-7 Link functions for ordinal regression model

Name	**Distribution**	**Link function**	**Inverse link**	**Density**
Logit	Logistic	$\text{Log}[\frac{\gamma}{1-\gamma}]$	$1/[1+\exp(\theta)]$	$\exp(-\theta)/[1+\exp(-\theta)]^2$
Probit	Normal	$\Phi^{-1}(\gamma)$	$\Phi(\theta)$	$\emptyset(\theta)$
Log-log	Gumbel (max)	$-\log[-\log(\gamma)]$	$\exp(-\exp(-\theta))$	$\exp(-\exp(-\theta)-\theta)$
Clog-log	Gumbel (min)	$-\log[-\log(1-\gamma)]$	$1-\exp(-\exp(\theta))$	$\exp[-\exp(\theta)+\theta]$
Cauchit	Cauchy	$\text{Tan}[\pi(\gamma-0.2)]$	$\frac{\arctan(\theta)}{\pi}+0.5$	$1/[\pi(1+\theta^2)]$

Source: (Christensen, 2010)

Table C-8 Confusion matrix on the number of commute activities

	1 (sim)	2 (sim)	3 (sim)
1 (obs)	9927 (10936)	1022 (13)	0 (0)
2 (obs)	936 (1330)	413 (19)	0 (0)
3 (obs)	60 (84)	25 (1)	0 (0)
Percentage of correct prediction=83.5%			
Sim total/obs total= 100.4%			

Note: Uncalibrated results are shown in parentheses.

Table C-9 Confusion matrix of the number of non-commute activities

	0 (sim)	1 (sim)	2 (sim)	3 (sim)
0 (obs)	11547 (12265)	800 (114)	33 (1)	0 (0)
1 (obs)	2595 (3170)	925 (5362)	5060 (48)	0 (0)
2 (obs)	610 (915)	476 (2866)	2736 (41)	0 (0)
3 (obs)	107 (202)	147 (1000)	968 (20)	0 (0)
Percentage of correct prediction=58.5%				
Sim total/obs total= 99.7%				

Note: Uncalibrated results are shown in parentheses.

Table C-10 Quota of alternative sampling for distance bands

Distance band (km)	**Shop**	**Entertain**	**Dine out**	**Personal business**	**Escort**	**Others**
0-1	4.5	3.5	4	2.5	3	1
1-2	2.5	2.5	2	1.5	2.5	1
2-3	1	1	1	1	1	1
3-4	0.5	0.5	0.5	1	1	0.5
4-5	0.5	0.5	0.5	0.5	0.5	0.5
5-10	0.5	1	1	2	1	2
>10	0.5	1	1	1.5	1	3

Table C-11 Distribution of travel time given the activity type and position in the activity plan

Type	Position	Before am peak	Am peak	Before noon	Afternoon	Pm peak	After pm peak
Work	1st tour	12%	74%	7%	6%	0%	0%
	2nd tour	0%	5%	3%	90%	2%	0%
	3rd tour	0%	0%	1%	64%	23%	12%
School	1st tour	34%	60%	1%	4%	0%	0%
	2nd tour	0%	0%	1%	97%	2%	0%
	3rd tour	0%	0%	0%	29%	36%	36%
Shopping	1st tour	4%	47%	38%	9%	1%	0%
	2nd tour	0%	6%	20%	57%	10%	7%
	3rd tour	0%	0%	1%	46%	29%	24%
Entertain	1st tour	26%	49%	17%	5%	1%	1%
	2nd tour	0%	2%	4%	46%	18%	29%
	3rd tour	0%	0%	0%	19%	26%	55%
Dining out	1st tour	5%	19%	44%	18%	12%	2%
	2nd tour	0%	2%	19%	20%	42%	18%
	3rd tour	0%	0%	0%	0%	76%	24%
Personal business	1st tour	10%	49%	26%	14%	0%	0%
	2nd tour	0%	4%	15%	74%	4%	2%
	3rd tour	0%	0%	6%	70%	8%	17%
Escorting	1st tour	15%	61%	9%	12%	1%	1%
	2nd tour	0%	1%	5%	77%	11%	5%
	3rd tour	0%	0%	1%	71%	23%	6%
Others	1st tour	8%	40%	38%	13%	1%	1%
	2nd tour	0%	3%	19%	61%	9%	8%
	3rd tour	0%	0%	3%	40%	27%	31%
Total		12%	50%	14%	18%	3%	3%

Table C-12 Quota of alternative sampling for detour distance bands

<table>
<tr><th>Distance range (km)</th><th>Shop</th><th>Entertain</th><th>Dine out</th><th>Personal business</th><th>Escort</th><th>Others</th></tr>
<tr><td>0-1</td><td>7</td><td>6</td><td>8</td><td>5</td><td>4.5</td><td>4</td></tr>
<tr><td>1-2</td><td>1.5</td><td>1.5</td><td rowspan="2">1</td><td>1.5</td><td>1.5</td><td>1</td></tr>
<tr><td>2-3</td><td>0.5</td><td>1</td><td>0.5</td><td>1</td><td>1</td></tr>
<tr><td>3-4</td><td rowspan="2">0.5</td><td rowspan="2">0.5</td><td rowspan="2">0.5</td><td>0.5</td><td>0.5</td><td>0.5</td></tr>
<tr><td>4-5</td><td>0.5</td><td>0.5</td><td>0.5</td></tr>
<tr><td>5-10</td><td rowspan="2">0.5</td><td>0.5</td><td rowspan="2">0.5</td><td>1</td><td>1</td><td>1.5</td></tr>
<tr><td>>10</td><td>0.5</td><td>1</td><td>1</td><td>1.5</td></tr>
</table>

References

Acheampong, R. A., & Silva, E. (2015). Land use–transport interaction modeling: A review of the literature and future research directions. *Journal of Transport and Land Use, 8*(3).

Aditjandra, P. T., Cao, X., & 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*(1), 22-32. doi:10.1016/j.tra.2011.09.001

Aditjandra, P. T., Mulley, C., & Nelson, J. D. (2013). The influence of neighbourhood design on travel behaviour: Empirical evidence from North East England. *Transport Policy, 26*, 54-65. doi:10.1016/j.tranpol.2012.05.011

Agyemang-Duah, K., & Hall, F. L. (1997). Spatial transferability of an ordered response model of trip generation. *Transportation Research Part A: Policy and Practice, 31*(5), 389-402.

American Planning Association. (2002). Policy guide on smart growth. Retrieved from https://www.planning.org/policy/guides/adopted/smartgrowth.htm

American Planning Association. (2006). *Planning and urban design standards*: John Wiley & Sons.

Appleyard, D. (1980). Livable streets: protected neighborhoods? *The ANNALS of the American Academy of Political and Social Science, 451*(1), 106-117.

Arentze, T. A., & Timmermans, H. J. (2004). A learning-based transportation oriented simulation system. *Transportation Research Part B: Methodological, 38*(7), 613-633.

Atkinson, G. (2003). *Construction quality and quality standards: The European perspective*: Routledge.

Auld, J., & Mohammadian, A. (2009). Framework for the development of the agent-based dynamic activity planning and travel scheduling (ADAPTS) model. *Transportation Letters, 1*(3), 245-255.

Auld, J., & Mohammadian, A. (2011). Planning-constrained destination choice in activity-based model: agent-based dynamic activity planning and travel scheduling. *Transportation Research Record: Journal of the Transportation Research Board*(2254), 170-179.

Auld, J., & Mohammadian, A. K. (2012). Activity planning processes in the Agent-based Dynamic Activity Planning and Travel Scheduling (ADAPTS) model. *Transportation Research Part A: Policy and Practice, 46*(8), 1386-1403.

Axhausen, K. W., & Garling, T. (1992). Activity-based approaches to travel analysis: Conceptual frameworks, models and research problems. *Transport Reviews, 12*(4), 323-341.

Bain, L., Gray, B., & Rodgers, D. (2012). *Living streets: Strategies for crafting public space*: John Wiley & Sons.

Balmer, M., Rieser, M., Meister, K., Charypar, D., Lefebvre, N., & Nagel, K. (2009). MATSim-T: Architecture and simulation times *Multi-agent systems for traffic and transportation engineering* (pp. 57-78): IGI Global.

Batty, M. (2009). Urban modeling. *International Encyclopedia of Human Geography. Oxford, UK: Elsevier*.

Batty, M. (2013). Visually-driven urban simulation: Exploring fast and slow change in residential location. *Environment and Planning A, 45*(3), 532-552.

Beebee, H., Hitchcock, C., & Menzies, P. (2009). *The Oxford handbook of causation*: Oxford University Press.

Beijing Transportation Research Center. (2011). *Transportation Report of Beijing 2010*. Retrieved from http://www.bjtrc.org.cn/InfoCenter/NewsAttach//aeb7c878-d31e-4f08-982f-3c17c717c87b.pdf

Beijing Transportation Research Center. (2013). *Transportation Report of Beijing 2012*. Retrieved from Beijing: http://www.bjtrc.org.cn/InfoCenter/NewsAttach/2013%E5%B9%B4%E5%8C%97%E4%BA%AC%E4%BA%A4%E9%80%9A%E5%8F%91%E5%B1%95%E5%B9%B4%E6%8A%A5.pdf

Bellemans, T., Kochan, B., Janssens, D., Wets, G., Arentze, T., & Timmermans, H. (2010). Implementation framework and development trajectory of FEATHERS activity-based simulation platform. *Transportation Research Record: Journal of the Transportation Research Board*(2175), 111-119.

Ben-Akiva, M. E., & Lerman, S. R. (1985). *Discrete choice analysis: Theory and application to travel demand* (Vol. 9): MIT press.

Bento, A. M., Cropper, M., Mobarak, A. M., & Vinha, K. (2005). The impact of urban spatial structure on travel demand in the United States. *Social Science Electronic Publishing, 87*(3), 466--478.

Bhat, C. R., & Eluru, N. (2009). A copula-based approach to accommodate residential self-selection effects in travel behavior modeling. *Transportation Research Part B-Methodological, 43*(7), 749-765. doi:10.1016/j.trb.2009.02.001

Bhat, C. R., Guo, J., Srinivasan, S., & Sivakumar, A. (2004). Comprehensive econometric microsimulator for daily activity-travel patterns. *Transportation Research Record: Journal of the Transportation Research Board*(1894), 57-66.

Bhat, C. R., & Pulugurta, V. (1998). A comparison of two alternative behavioral choice mechanisms for household auto ownership decisions. *Transportation Research Part B: Methodological, 32*(1), 61-75.

Boarnet, M. G., & Crane, R. (2001). *Travel by design: The influence of urban form on travel*: Oxford University Press.

Boarnet, M. G., Forsyth, A., Day, K., & Oakes, J. M. (2011). The street level built environment and physical activity and walking: Results of a predictive validity study for the Irvine Minnesota Inventory. *Environment and Behavior, 43*(6), 735-775. doi:10.1177/0013916510379760

Boarnet, M. G., Joh, K., Siembab, W., Fulton, W., & Nguyen, M. T. (2011). Retrofitting the suburbs to increase walking: Evidence from a land-use-travel study. *Urban Studies, 48*(1), 129-159.

Bowman, J., & Bradley, M. (2005). *Activity-based travel forecasting model for SACOG: Technical memos numbers 1-11*. Retrieved from http://jbowman.net

Bradley, M., & Bowman, J. (2006). *A summary of design features of activity-based microsimulation models for US MPOs.* Paper presented at the White Paper for the Conference on Innovations in Travel Demand Modeling, Austin, TX.

Brandt, E., & Rasmussen, M. (2002). Assessment of building conditions. *Energy and buildings, 34*(2), 121-125.

Brownson, R. C., Hoehner, C. M., Day, K., Forsyth, A., & Sallis, J. F. (2009). Measuring the built environment for physical activity: State of the science. *American journal of preventive medicine, 36*(4), S99-S123. e112.

Byles, J., Curryer, C., Vo, K., Forder, P., Loxton, D., & McLaughlin, D. (2016). Changes in housing among older women: Latent class analysis of housing patterns in older Australian women. *Urban Studies*. doi:10.1177/0042098016661309

Cambridge Systematics. (2002). San Francisco Travel Demand Forecasting Model development: Executive summary. *Prepared for San Francisco County Transportation Authority*.

Cao, X. (2015a). Examining the impacts of neighborhood design and residential self-selection on active travel: A methodological assessment. *Urban Geography, 36*(2), 236-255. doi:10.1080/02723638.2014.956420

Cao, X. (2015b). Heterogeneous effects of neighborhood type on commute mode choice: An exploration of residential dissonance in the Twin Cities. *Journal of Transport Geography, 48*, 188-196. doi:10.1016/j.jtrangeo.2015.09.010

Cao, X., & Fan, Y. (2012). Exploring the influences of density on travel behavior using propensity score matching. *Environment and Planning B-Planning & Design, 39*(3), 459-470. doi:10.1068/b36168

Cao, X., Mokhtarian, P. L., & Handy, S. (2009). Examining the impacts of residential self‐selection on travel behaviour: A focus on empirical findings. *Transport Reviews, 29*(3), 359-395. doi:10.1080/01441640802539195

Castiglione, J., Bradley, M., & Gliebe, J. (2015). *Activity-based travel demand models: A primer*: Transport Research Board.

Castle, C. J. E., & Crooks, A. T. (2006). Principles and concepts of agent-based modelling for developing geospatial simulations. *Ucl Centre for Advanced Spatial Analysis*(2006).

Cervero, R. (1989). Jobs-housing balancing and regional mobility. *Journal of the American Planning Association, 55*(2), 136-150.

Cervero, R. (1996). Jobs-housing balance revisited: Trends and impacts in the San Francisco Bay Area. *Journal of the American Planning Association, 62*(4), 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*(4), 475-490. doi:10.1080/01944360608976767

Cervero, R., & Kockelman, K. (1997). Travel demand and the 3Ds: Density, diversity, and design. *Transportation Research Part D-Transport and Environment, 2*(3), 199-219. doi:10.1016/s1361-9209(97)00009-6

Cervero, R., & Radisch, C. (1996). Travel choices in pedestrian versus automobile oriented neighborhoods. *Transport Policy, 3*(3), 127-141.

Chapin Jr, F. S. (1968). Activity systems and urban structure: A working schema. *Journal of the American Institute of Planners, 34*(1), 11-18.

Chatman, D. G. (2008). Deconstructing development density: Quality, quantity and price effects on household non-work travel. *Transportation Research Part A-Policy and Practice, 42*(7), 1008-1030. doi:10.1016/j.tra.2008.02.003

Chatman, D. G. (2009). Residential choice, the built environment, and nonwork travel:

Evidence using new data and methods. *Environment and Planning A, 41*(5), 1072-1089. doi:10.1068/a4114
Chen, C., & McKnight, C. E. (2007). Does the built environment make a difference? Additional evidence from the daily activity and travel behavior of homemakers living in New York City and suburbs. *Journal of Transport Geography, 15*(5), 380-395. doi:10.1016/j.jtrangeo.2006.11.010
Chen, H., Jia, B., & Lau, S. S. Y. (2008). Sustainable urban form for Chinese compact cities: Challenges of a rapid urbanized economy. *Habitat International, 32*(1), 28-40.
Cho, G.-H., & Rodriguez, D. A. (2015a). Location or design? Associations between neighbourhood location, built environment and walking. *Urban Studies, 52*(8), 1434-1453. doi:10.1177/0042098014537691
Cho, G.-H., & Rodriguez, D. A. (2015b). Neighborhood design, neighborhood location, and three types of walking: Results from the Washington DC area. *Environment and Planning B-Planning & Design, 42*(3), 526-540. doi:10.1068/b130222p
Christensen, R. H. B. (2010). Ordinal regression models for ordinal data. *R package version, 22*.
Chudyk, A. M., Winters, M., Moniruzzaman, M., Ashe, M. C., Gould, J. S., & McKay, H. (2015). Destinations matter: The association between where older adults live and their travel behavior. *Journal of Transport & Health, 2*(1), 50-57.
Colabianchi, N., Dowda, M., Pfeiffer, K. A., Porter, D. E., Almeida, M. J. C., & Pate, R. R. (2007). Towards an understanding of salient neighborhood boundaries: Adolescent reports of an easy walking distance and convenient driving distance. *International Journal of Behavioral Nutrition and Physical Activity, 4*(1), 66.
Crane, R. (1995). On form versus function: Will the "New Urbanism" reduce traffic or increase it? *Journal of Planning Education & Research, 15*(2), 117-126. Retrieved from http://www.escholarship.org/uc/item/7bj9g6bg
Croissant, Y. (2012). Estimation of multinomial logit models in R: The mlogit packages. *R package version 0.2-2. URL: http://cran.r-project.org/web/packages/mlogit/vignettes/mlogit.pdf*.
Cullen, G. (1961). *The concise townscape*: Routledge.
De Vos, J. (2015). The influence of land use and mobility policy on travel behavior: A comparative case study of Flanders and the Netherlands. *Journal of Transport and Land Use, 8*(1), 171-190. doi:10.5198/jtlu.2015.709
De Vos, J., Derudder, B., Van Acker, V., & Witlox, F. (2012). Reducing car use: changing attitudes or relocating? The influence of residential dissonance on travel behavior. *Journal of Transport Geography, 22*, 1-9. doi:10.1016/j.jtrangeo.2011.11.005
Dittmar, H., & Ohland, G. (2004). *The new transit town : Best practices in transit-oriented development*: Island Press.
Domencich, T. A., & McFadden, D. (1975). Urban travel demand-a behavioral analysis. *Canadian Journal of Economics, 10*(4).
Doyle, S., KellySchwartz, A., Schlossberg, M., & Stockard, J. (2006). Active community environments and health: The relationship of walkable and safe communities to individual health. *Journal of the American Planning Association, 72*(1), 19-31.
Ellder, E. (2014). Residential location and daily travel distances: The influence of trip purpose. *Journal of Transport Geography, 34*, 121-130. doi:10.1016/j.jtrangeo.2013.11.008

Eom, H.-J., & Cho, G.-H. (2015). Exploring thresholds of built environment characteristics for walkable communities: Empirical evidence from the Seoul Metropolitan area. *Transportation Research Part D-Transport and Environment, 40*, 76-86. doi:10.1016/j.trd.2015.07.005

Ettema, D., de Jong, K., Timmermans, H., & Bakema, A. (2007). PUMA: Multi-agent modelling of urban systems *Modelling land-use change* (pp. 237-258): Springer.

Ewing, R., & Cervero, R. (2001). Travel and the built environment: A synthesis *Land Development and Public Involvement in Transportation: Planning and Administration* (pp. 87-114).

Ewing, R., & Cervero, R. (2010). Travel and the Built Environment. *Journal of the American Planning Association, 76*(3), 265-294. doi:10.1080/01944361003766766

Ewing, R., Clemente, O., Neckerman, K. M., Purciel-Hill, M., Quinn, J. W., & Rundle, A. (2013). *Measuring Urban Design*: Island Press/Center for Resource Economics.

Ewing, R., Greenwald, M., Zhang, M., Walters, J., Feldman, M., Cervero, R., & Thomas, J. (2009). Measuring the impact of urban form and transit access on mixed use site trip generation rates: Portland pilot study. *US Environmental Protection Agency, Washington, DC*.

Ewing, R., & Handy, S. (2009). Measuring the unmeasurable: Urban design qualities related to walkability. *Journal of Urban design, 14*(1), 65-84.

Ewing, R., Tian, G., Goates, J. P., Zhang, M., Greenwald, M. J., Joyce, A., . . . Greene, W. (2015). Varying influences of the built environment on household travel in 15 diverse regions of the United States. *Urban Studies, 52*(13), 2330-2348. doi:10.1177/0042098014560991

Feng, J., Dijst, M., Prillwitz, J., & Wissink, B. (2013). Travel time and distance in international perspective: A comparison between Nanjing (China) and the Randstad (The Netherlands). *Urban Studies, 50*(14), 2993-3010. doi:10.1177/0042098013482504

Freeman, L., Neckerman, K., Schwartz-Soicher, O., Quinn, J., Richards, C., Bader, M. D. M., . . . Rundle, A. G. (2013). Neighborhood walkability and active travel (walking and cycling) in New York City. *Journal of Urban Health-Bulletin of the New York Academy of Medicine, 90*(4), 575-585. doi:10.1007/s11524-012-9758-7

Gärling, T., Kwan, M. P., & Golledge, R. G. (1994). Computational-process modelling of household activity scheduling. *Transportation Research Part B: Methodological, 28*(5), 355-364.

Giles-Corti, B., Bull, F., Knuiman, M., McCormack, G., Van Niel, K., Timperio, A., . . . Boruff, B. (2013). The influence of urban design on neighbourhood walking following residential relocation: Longitudinal results from the RESIDE study. *Social Science & Medicine, 77*, 20-30. doi:10.1016/j.socscimed.2012.10.016

Gim, T.-H. T. (2013). The relationships between land use measures and travel behavior: A meta-analytic approach. *Transportation Planning and Technology, 36*(5), 413-434. doi:10.1080/03081060.2013.818272

Giuliano, G., & Dargay, J. (2006). Car ownership, travel and land use: A comparison of the US and Great Britain. *Transportation Research Part A Policy & Practice, 40*(2), 106-124.

Goulias, K. G. (2002). *Transportation systems planning: Methods and applications*: CRC Press.

Goulias, K. G. (2009). Travel behavior and demand analysis and prediction. In R. A. Meyers (Ed.), *Encyclopedia of Complexity and Systems Science* (pp. 9536-9565). New York, NY: Springer New York.

Greene, W. H., & Hensher, D. A. (2010). *Modeling ordered choices: A primer*: Cambridge University Press.

Greenwald, M. J. (2006). *The relationship between land use and trip internalization behaviors: Evidence and Implications.* Paper presented at the TRB 2006 Annual Meeting, CD-ROM.

Guerra, E. (2014). The built environment and car use in Mexico City: Is the relationship changing over time? *Journal of Planning Education and Research, 34*(4), 394-408. doi:10.1177/0739456x14545170

Hägerstraand, T. (1970). What about people in regional science? *Papers in Regional Science, 24*(1), 7-24.

Hagenaars, J. A., & McCutcheon, A. L. (2002). *Applied latent class analysis*: Cambridge University Press.

Haklay, M. (2010). How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. *Environment and Planning B: Planning and Design, 37*(4), 682-703.

Handy, S. (1993). *Regional versus local accessibility-implications for nonwork travel*. Retrieved from

Handy, S., Cao, X., & 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*(6), 427-444. doi:10.1016/j.trd.2005.05.002

Hanson, S., & Giuliano, G. (2004). *The geography of urban transportation*: Guilford Press.

Harvey, C. (2014). *Measuring streetscape design for livability using spatial data and methods.* (the Degree of Master of Science), The University of Vermont and State Agricultural College.

Henson, K., Goulias, K. G., & Golledge, R. (2009). An assessment of activity-based modeling and simulation for applications in operational studies, disaster preparedness, and homeland security. *Transportation Letters, 1*(1), 19-39.

Heres-Del-Valle, D., & Niemeier, D. (2011). CO2 emissions: Are land-use changes enough for California to reduce VMT? Specification of a two-part model with instrumental variables. *Transportation Research Part B: Methodological, 45*(1), 150-161. doi:10.1016/j.trb.2010.04.001

Ho, C. Q., & Mulley, C. (2013). Multiple purposes at single destination: A key to a better understanding of the relationship between tour complexity and mode choice. *Transportation Research Part A-Policy and Practice, 49*, 206-219. doi:10.1016/j.tra.2013.01.040

Hong, J., Shen, Q., & Zhang, L. (2014). How do built-environment factors affect travel behavior? A spatial analysis at different geographic scales. *Transportation, 41*(3), 419-440. doi:10.1007/s11116-013-9462-9

Hoskin, R. (2012). The dangers of self-report. Retrieved from http://www.sciencebrainwaves.com/the-dangers-of-self-report/

Huang, X., Cao, X., Cao, X., & Yin, J. (2016). How does the propensity of living near rail transit moderate the influence of rail transit on transit trip frequency in Xi'an? *Journal of*

Transport Geography, 54, 194-204.

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. doi:10.1016/j.tra.2015.08.007

Joh, K., Nguyen, M. T., & Boarnet, M. G. (2012). Can built and social environmental factors encourage walking among individuals with negative walking attitudes? *Journal of Planning Education and Research, 32*(2), 219-236. doi:10.1177/0739456x11427914

Kamruzzaman, M., Baker, D., Washington, S., & Turrell, G. (2013). Residential dissonance and mode choice. *Journal of Transport Geography, 33*, 12-28. doi:10.1016/j.jtrangeo.2013.09.004

Kelly, S. G. (2013). *Decarbonising the English residential sector: Modelling policies, technologies and behaviour within a heterogeneous building stock.* (PhD), University of Cambridge.

Kim, C., & Wang, S. (2015). Empirical examination of neighborhood context of individual travel behaviors. *Applied Geography, 60*, 230-239. doi:10.1016/j.apgeog.2014.10.017

Kitamura, R. (1996). *Applications of models of activity behavior for activity based demand forecasting.* Paper presented at the Proceedings of Activity-Based Travel Forecasting Conference.

Kitchin, R. (2014). The real-time city? Big data and smart urbanism. *Geojournal, 79*(1), 1-14.

Klinger, T., & Lanzendorf, M. (2016). Moving between mobility cultures: What affects the travel behavior of new residents? *Transportation, 43*(2), 243-271.

Knaap, G., & Talen, E. (2005). New Urbanism and Smart Growth: A few words from the academy. *International Regional Science Review, 28*(2), 107-118.

Knuiman, M. W., Christian, H. E., Divitini, M. L., Foster, S. A., Bull, F. C., Badland, H. M., & Giles-Corti, B. (2014). A longitudinal analysis of the influence of the neighborhood built environment on walking for transportation. *American Journal of Epidemiology, 180*(5), 453-461. doi:10.1093/aje/kwu171

Kockelman, K. (1997). Travel behavior as function of accessibility, land use mixing, and land use balance: Evidence from San Francisco Bay Area. *Transportation Research Record, 1607*(1), 116-125. doi:10.3141/1607-16

Kostof, S. (1999). *The city assembled : The elements of urban form through history*: Thames and Hudson.

Krizek, K. J. (2003). Residential relocation and changes in urban travel: Does neighborhood-scale urban form matter? *Journal of the American Planning Association, 69*(3), 265-281. doi:10.1080/01944360308978019

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). *Imagenet classification with deep convolutional neural networks.* Paper presented at the Advances in neural information processing systems.

Kulkarni, A., & McNally, M. G. (2000). *A microsimulation of daily activity patterns.* Paper presented at the The 80th Annual Meeting of the Transportation Research Board.

Lamiquiz, P. J., & Lopez-Dominguez, J. (2015). Effects of built environment on walking at the neighbourhood scale: A new role for street networks by modelling their configurational accessibility? *Transportation Research Part A-Policy and Practice, 74*,

148-163. doi:10.1016/j.tra.2015.02.003
Lang, J. (1994). *Urban design: the American experience*: John Wiley & Sons.
Lanza, S. T., Flaherty, B. P., & Collins, L. M. (2010). *Latent class and latent transition analysis*: Wiley.
Lee, J. S., Nam, J., & Lee, S. S. (2014). Built environment impacts on individual mode choice: An empirical study of the Houston-Galveston Metropolitan Area. *International Journal of Sustainable Transportation, 8*(6), 447-470. doi:10.1080/15568318.2012.716142
Lehnerer, A. (2009). *Grand urban rules*: 010 Publishers.
Li, F., Luo, J., Lu, X., Deng, B., & Gan, L. (2016). Influencing factors for the asset distribution and changing trend of Chinese households. *Management World*(2), 45-56.
Lin, J., & Long, L. (2008). What neighborhood are you in? Empirical findings of relationships between household travel and neighborhood characteristics. *Transportation, 35*(6), 739-758. doi:10.1007/s11116-008-9167-7
Lin, J., & Yang, A. (2009). Structural analysis of how urban form impacts travel demand: Evidence from Taipei. *Urban Studies, 46*(9), 1951-1967. doi:10.1177/0042098009106017
Linzer, D. A., & Lewis, J. B. (2011). poLCA: An R package for polytomous variable latent class analysis. *Journal of Statistical Software, 42*(10), 1-29.
Liu, F., Janssens, D., Cui, J., Wang, Y., Wets, G., & Cools, M. (2014). Building a validation measure for activity-based transportation models based on mobile phone data. *Expert Systems with Applications, 41*(14), 6174-6189.
Logan, J. R., Bian, Y., & Bian, F. (1999). Housing inequality in urban China in the 1990s. *International Journal of Urban & Regional Research, 23*(23), 7-25.
Lowe, D. G. (1999). *Object recognition from local scale-invariant features.* Paper presented at the International Conference on Computer Vision.
Lund, H. (2003). Testing the claims of New Urbanism: Local access, pedestrian travel, and neighboring behaviors. *Journal of the American Planning Association, 69*(4), 414-429. doi:10.1080/01944360308976328
Lyon, N. B. (1978). *Yonghe Street Revitalization Project*. Toronto.
Ma, H., Arentze, T. A., & Timmermans, H. J. (2012). *An agent based model of dynamic activity-travel scheduling*. Paper presented at the Transportation Research Board 91st Annual Meeting.
Maat, K., & Timmermans, H. J. P. (2009). A causal model relating urban form with daily travel distance through activity/travel decisions. *Transportation Planning and Technology, 32*(2), 115-134. doi:10.1080/03081060902861285
Maat, K., Wee, B. V., & Stead, D. (2005). Land use and travel behaviour: Expected effects from the perspective of utility theory and activity-based theories. *Environment and Planning B: Planning and Design, 32*(1), 33-46.
Madlener, R., & Sunak, Y. (2011). Impacts of urbanization on urban structures and energy demand: What can we learn for urban energy planning and urbanization management? *Sustainable Cities & Society, 1*(1), 45–53.
Malleson, N. (2014). Calibration of Simulation Models. *Encyclopedia of Criminology & Criminal Justice, 40*, 115-118.
Manaugh, K., & El-Geneidy, A. M. (2015). The importance of neighborhood type dissonance in

understanding the effect of the built environment on travel behavior. *Journal of Transport and Land Use, 8*(2), 45-57. doi:10.5198/jtlu.2015.718

Marcus, C. C., & Francis, C. (1997). *People places: Design guidlines for urban open space*: John Wiley & Sons.

Martinussen, M., & Kroger, J. (2013). Meta-analytic studies of identity status and personality: introduction and overview. *Identity, 13*(3), 189-200.

Mashhadi, A., Quattrone, G., Capra, L., & Mooney, P. (2012). *On the accuracy of urban crowd-sourcing for maintaining large-scale geospatial databases.* Paper presented at the Proceedings of the Eighth Annual International Symposium on Wikis and Open Collaboration.

McCormack, G. R., Shiell, A., Giles-Corti, B., Begg, S., Veerman, J. L., Geelhoed, E., . . . Emery, J. C. H. (2012). The association between sidewalk length and walking for different purposes in established neighborhoods. *International Journal of Behavioral Nutrition and Physical Activity, 9*, 12. doi:10.1186/1479-5868-9-92

McFadden, D. (1972). Conditional logit analysis of qualitative choice behavior. *Frontiers in Econometrics*, 105-142.

McFadden, D. (1974). The measurement of urban travel demand. *Journal of public economics, 3*(4), 303-328.

McFadden, D. (1978). Modeling the choice of residential location. *Cowles Foundation Discussion Papers, 673*(477), 72-77.

McMillan, T. E. (2007). The relative influence of urban form on a child's travel mode to school. *Transportation Research Part A: Policy and Practice, 41*(1), 69-79. doi:10.1016/j.tra.2006.05.011

McNally, M. G. (2007). The four-step model *Handbook of Transport Modelling: 2nd Edition* (pp. 35-53): Emerald Group Publishing Limited.

Merlin, L. A. (2015). Can the built environment influence nonwork activity participation? An analysis with national data. *Transportation, 42*(2), 369-387. doi:10.1007/s11116-014-9554-1

Miller, E., & Roorda, M. (2003). Prototype model of household activity-travel scheduling. *Transportation Research Record*(1831), 114-121.

Millward, H., Spinney, J., & Scott, D. (2013). Active-transport walking behavior: Destinations, durations, distances. *Journal of Transport Geography, 28*, 101-110. doi:10.1016/j.jtrangeo.2012.11.012

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*(3), 204-228. doi:10.1016/j.trb.2007.07.006

Montgomery, J. (1998). Making a city: Urbanity, vitality and urban design. *Journal of Urban design, 3*(1), 93-116.

Moughtin, C. (2003). *Urban design: street and square*: Routledge.

Naess, P. (2010). Residential location, travel, and energy use in the Hangzhou Metropolitan Area. *Journal of Transport & Land Use, 3*(3), 27-59.

Naess, P. (2013). Residential location, transport rationales and daily-life travel behaviour: The case of Hangzhou Metropolitan Area, China. *Progress in Planning, 79*(79), 5-54.

Naess, P. (2015). Built environment, causality and travel. *Transport Reviews, 35*(3), 275-291.

doi:10.1080/01441647.2015.1017751
Nasar, J. L. (1983). Adult viewers' preferences in residential scenes: A study of the relationship of environmental attributes to preference. *Environment and Behavior, 15*(5), 589-614.
Nasri, A., & Zhang, L. (2012). Impact of metropolitan-level built environment on travel behavior. *Transportation Research Record*(2323), 75-79. doi:10.3141/2323-09
Nasri, A., & Zhang, L. (2015). Assessing the impact of metropolitan-level, county-level, and local-level built environment on travel behavior: Evidence from 19 US urban areas. *Journal of Urban Planning and Development, 141*(3). doi:10.1061/(asce)up.1943-5444.0000226
NavInfo. (2009). Digital map products. Retrieved from http://www.navinfo.com/en/product/vehiclenavigation.aspx
Neuman, M. (2005). The compact city fallacy. *Journal of Planning Education and Research, 25*(1), 11-26.
Parolek, D. G., Parolek, K., & Crawford, P. C. (2008). *Form based codes: A guide for planners, urban designers, municipalities, and developers*: John Wiley & Sons.
Pendyala, R. M., Kitamura, R., Kikuchi, A., Yamamoto, T., & Fujji, S. (2005). *FAMOS: The Florida activity mobility simulator.* Paper presented at the Proceedings of the 84th annual meeting of the transportation research board.
Pinjari, A. R., & Bhat, C. R. (2011). Activity-based travel demand analysis. *A Handbook of Transport Economics, 10*, 213-248.
Porta, S., Romice, O., Maxwell, J. A., Russell, P., & Baird, D. (2014). Alterations in scale: Patterns of change in main street networks across time and space. *Urban Studies, 51*(16), 3383-3400.
Quercia, D., O'Hare, N. K., & Cramer, H. (2014). *Aesthetic capital: What makes london look beautiful, quiet, and happy?* Paper presented at the ACM Conference on Computer Supported Cooperative Work & Social Computing.
Rajamani, J., Bhat, C. R., Handy, S., Knaap, G., & Song, Y. (2003). Assessing impact of urban form measures on nonwork trip mode choice after controlling for demographic and level-of-service effects. *Travel Demand and Land Use 2003: Planning and Administration*(1831), 158-165.
Rasmussen, C. E., & Ghahramani, Z. (2001). *Occam's Razor.* Paper presented at the neural information processing systems.
Rasouli, S., & Timmermans, H. (2014a). Activity-based models of travel demand: Promises, progress and prospects. *International Journal of Urban Sciences, 18*(1), 31-60. doi:10.1080/12265934.2013.835118
Rasouli, S., & Timmermans, H. (2014b). Applications of theories and models of choice and decision-making under conditions of uncertainty in travel behavior research. *Travel Behaviour and Society, 1*(3), 79-90. doi:10.1016/j.tbs.2013.12.001
Richardson, L. (1990). Narrative and sociology. *Journal of Contemporary Ethnography, 19*(1), 116-135. doi:doi:10.1177/089124190019001006
Rogerson, R., Sadler, S., Wong, C., & Green, A. (2010). Planning sustainable communities–skills and learning to envision future communities: an introduction. *Town Planning Review, 81*(5), 505-522.
Roorda, M. J., Miller, E. J., & Habib, K. M. (2008). Validation of TASHA: A 24-h activity scheduling

microsimulation model. *Transportation Research Part A: Policy and Practice, 42*(2), 360-375.

Said, S. Y., Zubir, S. S. S., & Rahmat, M. N. (2014). Measuring physical changes in an urban regeneration scheme. *WIT Transactions on Ecology and the Environment, 191*, 1165-1174.

Salon, D. (2015). Heterogeneity in the relationship between the built environment and driving: Focus on neighborhood type and travel purpose. *Research in Transportation Economics, 52*, 34-45. doi:10.1016/j.retrec.2015.10.008

Salon, D., Boarnet, M. G., Handy, S., Spears, S., & Tal, G. (2012). How do local actions affect VMT? A critical review of the empirical evidence. *Transportation Research Part D-Transport and Environment, 17*(7), 495-508. doi:10.1016/j.trd.2012.05.006

Salvini, P., & Miller, E. J. (2005). ILUTE: An operational prototype of a comprehensive microsimulation model of urban systems. *Networks and Spatial Economics, 5*(2), 217-234.

Santayana, G. (1955). *The sense of beauty: Being the outline of aesthetic theory* (Vol. 238): Courier Corporation.

Sarkar, C., Webster, C., Pryor, M., Tang, D., Melbourne, S., Zhang, X., & Liu, J. (2015). Exploring associations between urban green, street design and walking: Results from the Greater London boroughs. *Landscape and Urban Planning, 143*, 112-125. doi:10.1016/j.landurbplan.2015.06.013

Scheiner, J., & Holz-Rau, C. (2013). Changes in travel mode use after residential relocation: a contribution to mobility biographies. *Transportation, 40*(2), 431-458. doi:10.1007/s11116-012-9417-6

Schutt, R. K. (2011). *Investigating the social world: The process and practice of research*: Pine Forge Press.

Shabanpour, R., Javanmardi, M., Fasihozaman, M., Miralinaghi, M., & Mohammadian, A. (2017). Investigating the applicability of ADAPTS activity-based model in air quality analysis. *Travel Behaviour and Society*.

Simmonds, D., Waddell, P., & Wegener, M. (2013). Equilibrium v. dynamics in urban modelling. *Environment & Planning B Planning & Design, 40*(6), 1051-1070.

Singleton Jr, R. A., Straits, B. C., & Straits, M. M. (1993). *Approaches to social research*: Oxford University Press.

Sixth National Population Census Office of State Council. (2010). *The Sixth National Population Census*.

Song, Y., Preston, J. M., & Brand, C. (2013). What explains active travel behaviour? Evidence from case studies in the UK. *Environment and Planning A, 45*(12), 2980-2998. doi:10.1068/a4669

Southworth, M. (2005). Designing the walkable city. *Journal of Urban Planning and Development, 131*(4), 246-257.

Sperry, B. R., Burris, M. W., & Dumbaugh, E. (2012). A case study of induced trips at mixed-use developments. *Environment and Planning B: Planning and Design, 39*(4), 698-712.

Susilo, Y. O., Williams, K., Lindsay, M., & Dair, C. (2012). The influence of individuals' environmental attitudes and urban design features on their travel patterns in sustainable neighborhoods in the UK. *Transportation Research Part D-Transport and*

Environment, 17(3), 190-200. doi:10.1016/j.trd.2011.11.007

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., . . . Rabinovich, A. (2015). *Going deeper with convolutions.* Paper presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

Talen, E. (2013). *Charter of the new urbanism (2nd Edition)*: McGraw-Hill Professi.

Tan, J., & Zhao, Y. (2012). An empirical study of the housing-price-to-income ratio: Evidences from Beijing and China. *China Land Sciences, 26*(9), 66-70.

Tracy, A. J., Su, P., Sadek, A. W., & Wang, Q. (2011). Assessing the impact of the built environment on travel behavior: A case study of Buffalo, New York. *Transportation, 38*(4), 663-678. doi:10.1007/s11116-011-9337-x

Veldhuisen, J., Timmermans, H., & Kapoen, L. (2000). RAMBLAS: A regional planning model based on the microsimulation of daily activity travel patterns. *Environment and Planning A, 32*(3), 427-443.

Voordt, T., & Wegen, H. B. (2005). *Architecture in use: An introduction to the programming, design and evaluation of buildings*: Routledge.

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*(10), 1594-1607. doi:10.1016/j.tra.2012.08.001

Wang, Z., Deng, X., & Wong, C. (2016). Integrated land governance for Eco-Urbanization. *Sustainability, 8*(9), 903.

Witten, K., Blakely, T., Bagheri, N., Badland, H., Ivory, V., Pearce, J., . . . Schofield, G. (2012). Neighborhood built environment and transport and leisure physical activity: Findings using objective exposure and outcome measures in New Zealand. *Environmental Health Perspectives,* pp. 971-977.

Wohlwill, J. F. (1976). Environmental aesthetics: The environment as a source of affect. *Human behavior and environment: Advances in theory and research, 1*, 37-86.

Wong, C., Baker, M., Webb, B., Hincks, S., & Schulze-Baing, A. (2015). Mapping policies and programmes: The use of GIS to communicate spatial relationships in England. *Environment and Planning B: Planning and Design, 42*(6), 1020-1039.

Wu, F., & Li, Z. (2005). Sociospatial differentiation: Processes and spaces in subdistricts of Shanghai. *Urban Geography, 26*(2), 137-166.

Yang, L., Hipp, J. A., Adlakha, D., Marx, C. M., Tabak, R. G., & Brownson, R. C. (2015). Choice of commuting mode among employees: Do home neighborhood environment, worksite neighborhood environment, and worksite policy and supports matter? *Journal of Transport & Health, 2*(2), 212-218. doi:10.1016/j.jth.2015.02.003

Yang, Z., Cai, J., Ottens, H. F. L., & Sliuzas, R. (2013). City profile: Beijing. *Cities, 31*, 491-506.

Yarlagadda, A. K., & Srinivasan, S. (2007). Modeling children's school travel mode and parental escort decisions. *Transportation, 35*(2), 201-218. doi:10.1007/s11116-007-9144-6

Yasmin, F., Morency, C., & Roorda, M. J. (2015). Assessment of spatial transferability of an activity-based model, TASHA. *Transportation Research Part A: Policy and Practice, 78*, 200-213.

Yasmin, F., Morency, C., & Roorda, M. J. (2017). Macro-, meso-, and micro-level validation of an activity-based travel demand model. *Transportmetrica A: Transport Science, 13*(3), 222-249.

Zegras, C. (2010). The built environment and motor vehicle ownership and use: Evidence from Santiago de Chile. *Urban Studies, 47*(8), 1793-1817. doi:10.1177/0042098009356125

Zhao, P. (2011). Car use, commuting and urban form in a rapidly growing city: Evidence from Beijing. *Transportation Planning and Technology, 34*(6), 509-527. doi:10.1080/03081060.2011.600049

Zhao, P. (2014). Private motorised urban mobility in China's large cities: The social causes of change and an agenda for future research. *Journal of Transport Geography, 40*, 53-63. doi:10.1016/j.jtrangeo.2014.07.011

Zhao, P. (2015). The determinants of the commuting burden of low-income workers: Evidence from Beijing. *Environment and Planning A*. doi:10.1177/0308518x15597112

Zhao, P., & Lu, B. (2011). Managing urban growth to reduce motorised travel in Beijing: One method of creating a low-carbon city. *Journal of Environmental Planning and Management, 54*(7), 959-977. doi:10.1080/09640568.2010.547684

Zhao, P., Lu, B., & De Roo, G. (2011). The impact of urban growth on commuting patterns in a restructuring city: Evidence from Beijing. *Papers in Regional Science, 90*(4), 735-754. doi:10.1111/j.1435-5957.2010.00343.x

Zhao, Z., & Zhao, J. (2017). Car Pride: Psychological Structure and Behavioral Implications. Retrieved from https://dusp.mit.edu/publication/car-pride-psychological-structure-and-behavioral-implications

Zhou, B., & Kockelman, K. M. (2008). Self-selection in home choice use of treatment effects in evaluating relationship between built environment and travel behavior. *Transportation Research Record*(2077), 54-61. doi:10.3141/2077-08