ABSTRACT

Title of Dissertation:	UNDERSTANDING AND MODELLING TIME USE, WELL BEING AND DYNAMICS IN ACTIVITY-TRAVEL BEHAVIOR: A CHOICE BASED APPROACH
	Han Dong, Doctor of Philosophy, 2018
Dissertation directed by:	Cinzia Cirillo, Associate Professor, Civil and Environmental Engineering

Understanding the determinants of activity and travel related choices is critical for policy-makers, planners and engineers who are in charge of the management and design of large scale transportation systems. These systems, and their externalities, are interwoven with human actions and communities' evolution. Traditionally, individual decision-making and travel behaviour studies are based on random utility models (RUM) and discrete choice analysis. To extend the ability of modellers to represent and forecast complex travel behaviour, this dissertation expands existing models to accommodate the influence of variables other than the traditional socio-demographics or level of service variables. In this thesis, technology innovations, psychological factors, and perceptions of future uncertainty are integrated into the classical RUMs and their effects on activity-travel decision making are investigated.

Technology innovations, such as telecommunication, online communities and

entertainment, release individual's time and space constraints. They also modify people's activity and travel choices. An integrated discrete-continuous RUM is proposed to study individuals' participation in leisure activities, which is an important component of activity scheduling analysis and tour/trip formation. Leisure alternatives considered include: computer/internet related activity, in-home activity, and out-of-home activity. Compared to previous discrete-continuous models, interdependence among activities and the related time usage is explored using a modelling structure that accommodate full correlation among decision variables of different types.

Standard random utility models are extended by including attitudes and perceptions as latent variables; these constructs are expected to enhance the behavioural representation of the choice process. A simultaneous structural model is proposed to represent the mutual effects existing between psychological factors and activity choices. Biases due to endogeneity in psychological factors and activity choices are taken into consideration in the model. To further extend the behavioural realism of our model, this thesis proposes a new simultaneous equation model formulation that links psychological indicators to activity participation and time use decisions. Unlike previous studies, the proposed method allows the psychological factors to be correlated with time use decisions and serve as an attribute in time use choice model. A new iterative simulated maximization estimation method is also proposed to accommodate possible endogeneity bias in the model system. A simulation experiment shows that the estimation method produces consistent and unbiased estimation results. Moreover, a real case study is also implemented in the context of participation in leisure activities, linking emotions, activity involvement and time use.

After exploring individual's decisions on activity and time use choices, a dynamic discrete choice model framework is proposed to accommodate stochasticity in individual behaviour over time. Following previous studies, activity patterns are decomposed into tour and stop sequences. Accordingly, a tour choice model and a stop choice model are jointly formulated under a unified framework with a hierarchical structure where stop choices are assumed to be conditional on tour choices. The results indicate that individuals are sensitive to current and future changes in travel and activity characteristics and that a dynamic formulation better represents multi-day travel behaviour.

UNDERSTANDING AND MODELLING TIME USE, WELL BEING AND DYNAMICS IN ACTIVITY-TRAVEL BEHAVIOR:
A CHOICE BASED APPROACH

by

Han Dong

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
[Doctor of Philosophy]
[2018]

Advisory Committee:
Professor Cinzia Cirillo Chair
Professor Gang-Len Chang
Professor Paul M. Schonfeld
Professor Giovanni Baiocchi
Dr. Frederick W. Ducca

Table of Contents

List of Tables

List of Abbreviations

ATUS: American Time Use Survey
ABM: Activity Based Model
BA: Bachelor's degree
CDF: Cumulative Distribution Function
CNL: Crossed Nested Logit Model
DDCM: Dynamic Discrete Choice Model
GEV: General Extreme Value
GHDM: Generalized Heterogeneous Data Model
GNL: General Nested Logit
ICT: Information Communication Technology
IID: Independent and Identical Distributed
IIA: Independent and Identical Alternative
LPC: Pure In-Home Computer Use Activities
LH: Pure In-Home Other Leisure Activities
LOH: Pure Out-Of-Home Leisure Activities
LH&LPC: Multiple In-Home Leisure and Computer Use Activities
LH&LOH: Multiple In-Home and Out-Home Leisure Activities
MACML: Maximum Approximate Composite Marginal Likelihood
MAE: Mean Absolute Error
MDP: Markov Decision Process
ML: Maximum Likelihood
MNL: Multinomial Logit Model
MNP： Multinomial Probit Model
MXL: Mixed Logit Model
NL: Nested Logit Model
NO: No Leisure Activities
OLS: Ordinary Least Squares
PCL: Paired Combinatorial Logit
RUM: Random Utility Models
SLL: Simulated Log Likelihood
SML: Simulated Maximum Likelihood
SWB: Subjective Well-Being
EV: Extreme Value
WB: Well-Being

Chapter 1: Introduction

1.1 Background and Motivation

Concerns about congestion, emissions and land use patterns have driven governments to consider policies aimed at better managing the current transportation system. Such policies include shared driving, public transit incentive, road pricing, and telecommuting, which influence individual's behaviours and their travel choices by modifying travel cost and time usage. Understanding the determinants of activities and travel thus become a critical factor for transportation policymakers, planners and engineers to design and manage effectively transportation systems.

Activity-based travel demand approach has gained popularity in the recent two decades, proving to have a better performance in capturing important demand components than trip and tour-based modelling. More importantly, Activity-Based Modelling (ABM) supplies a valid framework for investigating the influence of activity duration and travel time on activity choices (Ben-Akiva et al 1996, Ben-Akiva and Bowman 1998; Bowman and Ben-Akiva 2001). ABM allows multiple tours to be connected over the course of the same day enabling the analyst to investigate the effects of inter-tour temporal spatial constraints. In ABMs, demand for activity and travel is viewed as a choice among all possible combinations of activity and travel in the course of daily life. The model uses a day time frame because people are assumed to organize activity and travel behaviour with the unit of a day, which allows within day scheduling decisions to be interacted with constraints of time and space. The daily activity pattern consists of multiple tours, which are tied up by an overarching activity pattern, thereby explicitly representing the ability of individuals to make inter-tour and at-home/out-of-home trade-offs.

1.2 Negative Correlation in Activity Choice Models

The discrete choice model is the predominant modelling tool in travel demand modelling, which is a mathematical model of a decision process that maps aspects of the choice situation. The model describes the choices of individuals through evaluating payoffs of choosing different alternatives. These payoffs depend on the attributes of the alternative and characteristics of the individuals. Random utility models of discrete choice (RUMs) are the most prevalent choice model (Ben-Akiva and Lerman, 1985). The wide RUM family consists of two main categories: the GEV models based on the assumption that the errors are type I EV distributed and the Probit model for which the errors are assumed to be multivariate normal.

The multinomial logit (MNL) model (McFadden 1973), the most widely applied discrete choice model, has the advantage of closed and simple form of mathematical structure and feasibility of estimation. However, the independence of irrelevant alternative (IIA) property implies the changes in the probability of any alternative will have the same proportional impact on all other alternatives. Another relaxation form in the GEV family is nested logit (NL) model (Williams, 1977), which allows dependence between the utilities of alternatives in the same nest. However, the remaining restriction on the equality and non-negativity dependence between all pairs of alternatives in a group is unrealistic, especially in the context of activity choices. Other models in the GEV family allow more flexible correlations among alternatives, which include: Paired Combinatorial Logit (Chu 1989), Cross-Nested Logit (Vovsha 1997). However, the key assumption in the GEV estimation process that any correlation between the error terms is necessarily non-negative.

There is no fundamental reason indicating that negative correlations should not occur from a behavioural perspective in the real world. Substitution effects should impose negative correlations among activity choices with similar purpose due to time and space constraints and individual's needs. Whereas, in the context of travel demand modelling, most models founded on the principle of utility maximization are relied on GEV models (e.g. Bowman 1998; Fosgerau, 1998; Wen and Koppelman, 2000; Bowman and Ben-Akiva, 2001). There are no previous studies clearly indicating the consequence of model output with the existence of negative correlations.

In Chapter 4, we investigate the bias in the estimation of coefficients and correlation terms deriving from two popular models in transportation planning models, Nested logit (NL) and Cross-nested Logit (CNL), under the condition that error terms are negatively correlated. To this aim we design several experiments, we estimate NL and CNL models and compare the results to those obtained by estimating a Multinomial Probit model (MNP). The performance of the three models is assessed with respect to the coefficients' estimates, the ability to recover the correlations among alternatives and the cross-validation results. The analysis is conducted both on simulated and real data.

1.3 Participation in Activities and Time Usage

The impetus for moving towards activity-based travel demand modelling is motivated by the causal process from which travel demand is derived. These processes and their dynamics are integrated within various spatial-temporal and socio-demographic systems that involve trade-offs between earning and consumption of resources and commodities (Habib, 2007). In the case of travel demand modelling, time and energy are spent to earn resources. Several studies in the past have investigated the theoretical aspects of the role of time in the entire demand system modelling (Becker 1965; Evan 1972). These studies indicate that time use is related to all other consumption processes. Here, we are concerned with the relation between time and space allocation of activities.

The traditional travel demand model focuses on the out-of-home activities and trips that involve travel or in-home activities that are indicated to have a substitution effect on some types of out-of-home activities (i.e. leisure or social). For the reason that time-space constraints play an important role in shaping people's activity patterns (Pendyala, 2002; Yamamoto et al., 2004; Kitamura et al., 2006), the choice of computer use is regarded as a replacement of physical activities. Computer related activities release spatial constraints, which also enable time to be allocated from travel to other activities. As a consequence, it would influence the overall activity-travel choices and the time use pattern. However, this substitution effects of computer use and correlations among time use and activity involvement are neglected in most travel demand models, which may result in the inability of the modelling framework to accurately capture and reflect individual demand for activities and formation of activity pattern.

In Chapter 5, we propose an integrated discrete continuous econometric framework which jointly estimate the activity choices and time usage decision. The model is designed to explore the effects that computer use for leisure and relaxing has on activity-travel patterns, including social and commute trips. The integrated model proposed also captures the potential correlation across activity involvement choices and associated time usage decisions.

1.4 Emotions, and Well-being in Activity Participation

Activity demand is not an arbitrary or random phenomenon. It has deep roots in basic human physiology, sociology and many other branches of basic social science. Maslow (1970) identifies that basic human needs is influencing us to engage in different type of activities. Maslow categorizes the basic needs of human beings into three broad categories: cognitive, conative, and aesthetic. As a component of cognitive, subjective well-being and emotions influence our decision processes on activity and time use (Leung and Lee, 2005; Lloyd and Auld, 2002; Robinson and Martin, 2008). However, classical discrete choice models activity-travel decisions only based on individuals' socio-demographics, activity location, and time of day, ignoring the joint effects that exist among emotions, activity choices, and time use. Although it might be difficult to integrate emotions into the current planning model due to data limitation and complexity in modelling, it is still worth to explore to what extent emotions affect activity-time use patterns in activity-based research and related disciplines (i.e.

marketing, social science, political science, health studies).

Incorporating well-being into studies of activity participation has the potential to more fully explain the activity and travel pattern. As subjective well-being is gaining interests in transportation researches, well-being or emotion related questionnaires are being added to activity diaries to explain what the feeling or satisfaction of their experiences are (Pavot and Diener, 1993; Diener et al., 1985; Västfjäll and Gärling, 2007; Olsson et al., 2012). This has tended to be in the form of using a simple add-on question per activity: “How are you feeling during this activity?” The answer choice tends to be seven-point scales ranging from 0 to 6. However, scale indicators cannot be used as explanatory variables, mainly for two reasons:

- Measurement errors: Scale is arbitrary and discrete. Justification bias of respondents may produce exaggerated responses (Ben-Akiva et al. 1999).
- No forecasting possibility: indicators cannot be predicted in the future.

In order to quantify the attitudes, perceptions and other psychological constructs, explain how they are formed, and influence choices, a model framework relied on RUM is proposed by Ben-Akiva et al. (1999). In the model, unobservable psychological constructs are quantified use an integration of latent variable model with discrete choice models. A simultaneous estimator is used, which incorporates indicators provided by responses to survey questions related to attitudes, perceptions, and motivations to aid in estimating the model. The model is then further developed to investigate the impacts of psychological factors on activity travel decisions (e.g. Vovsha and Bradley, 2004), mode choices (e.g. Kamrgianni and Polydoropoulou, 2013; Kamrgianni et al., 2015), and vehicle type choices (e.g. Daziano and Bolduc el al., 2013; Mabit et al., 2014). Emotions is referred as endogenous when the experience is relevant to the decision at hand and is regarded as an integral part of the goal setting and goal striving process (Zeelenberg et al., 2008). When incorporating such psychological variables into a choice model, resulting bias and inconsistent estimates cannot be ignored. However, limited studies identify the endogeneity of the latent variable and resulting bias.

In this dissertation, I explore the relation between individual's activity choice and psychological well-being with a recursive modelling system. Mutual effects between activities and emotions during chosen activities are investigated. Average treatment effects (ATE) and marginal effects are computed to further explore individual's choices under different emotion levels. In addition to that, we also explore the potential influence of subjective well-being on daily leisure activity participation (choice of leisure activity) and time use (duration of each activity) by proposing an extended hybrid choice modelling framework. Based on the discrete-continuous model formulation and recursive model developed in previous chapters, we introduce a comprehensive model system that is able to investigate the role of emotions into individual's activity-travel choice and time use decision. The derived model jointly estimates decisions related to time use, activity participation, and emotion levels. An iterative simulated maximization likelihood method is also proposed in the model estimation to avoid the bias caused by endogeneity.

1.5 Dynamics in Activity Choice Models

Important questions concerning how activity patterns are derived as well as how and why they change have gained great interests in the recent years. As a result, most researchers use activity type (work, school, shop, leisure, etc.) as a preliminary mean to explain when and how activity patterns are formed. For example, “work”, “school” and other “mandatory” or recurring activities are always assumed to be fixed in space and time and thus have higher priority in modelling, whereas “discretionary” activities are assumed to be more flexible and have lower priority. Based on these assumption, fixed or higher priority activities are assumed to be planned first in trip (e.g. Kitamura et al. 2000), tour (e.g. Shiftan 1998; Bowman and Ben-Akiva 2001) and activity scheduling models (e.g. Arentze and Timmermans 2000; Miller and Roorda, 2003).

Daily patterns are not a random event but rather a sequence of coherent set of actions (Miller, 2005). Accordingly, our daily activity schedule is formed with a sequence of interconnected activities, which are conducted by human-beings to fulfil their basic needs. The determination or planning of the activity schedule is a dynamic procedure that changes in time and space would influence the overall decision process. The time budget constraints on activity scheduling is presented through the relationship between daily time frame and activity durations. Any extension or contraction of the time spent on activities and travelling would influence the time frame of the overall activity schedule. The spatial constraint is mainly expressed in the relationship between long-term choices (i.e. home location, work location, and school location) and short-term activity locations. As stated, an individual’s daily activity schedule is formed with tours that are normally defined to start and end with activities with high priorities, such as work or school. Individuals seek to find the optimal routes and short-term locations between the origin and the destination of each tour to reduce their cost on travelling, and therefore maximize their overall utilities. Thus, individuals' future perceptions on activities is critical in activity scheduling and future choices are expected to influence current decisions on activity type and location choices. Accordingly, individuals need to modify their activity patterns due to changes in the time needed to reach all daily activity locations. However, most demand models assume that individuals have to choose tour or activity patterns from predefined choice sets, which restrain the ability of the model to capture uncertainty and individuals' stochastic behaviours on activity scheduling.

It is common in activity-travel demand modelling to consider a typical day as the time-span for activity-based model. A typical day assumption refers to a hypothesis that activities happen in that day have all features of a week. Models with the typical day assumption have smaller data requirements and computation burden. Whereas the major difference in favour of activity-based approach against trip-based approach is to capture within-day variations in travel behaviour. However, our activity pattern presents significant variations across the week in addition to the within-day variability (Schlich and Axhausen, 2003; Doherty et al., 2004). Moreover, the within-day variations of activity-travel behaviour are intricately related to the day-to-day variations (Habib and miller, 2008). Approaches towards the within-day variations in a typical day model thus becomes incomplete if day-to-day variations are not addressed properly.

The dissertation addresses these critical issues in travel demand modelling. Contrary to classical approach, Chapter 8 presents an integrated dynamic modelling system of activity scheduling. The overall activity schedule model system is constructed with a tour choice model and a stop choice model with a hierarchical structure. In the system, stop choice model are conditional on the tour choice. The whole activity pattern is split into several tours by locations such as home, work, and school. Tour choice model is formulated as a Markov decision process. Choices of intermediate stops in each tour are then approached with a dynamic discrete choice model. For the explicit consideration of within-day dynamics in time-use behaviour, time of day variables (AM peak period, PM peak period, and night period) and remaining time of the day are considered as variables both in the tour level and stop level choice model. Weekday indicators are considered as a variable in the tour level choice model to account for the day-to-day variations. In the work, a seven-day smart phone collected household travel survey is used.

1. 6 Contributions

As described above, classical econometric models, the primary tool used in travel demand modelling, are unable to accommodate substitution effects due to involvement in activity related to new technology and humans' subjective well-being and stochastic activity-travel behaviour. Even though progresses has been made on conceptualizing the model across different disciplines, more efforts are needed to cope with the exploding technology improvement and the understanding of human's recognition and perceptions. This dissertation focuses on proposing several advanced modelling frameworks that quantitatively and empirically investigate the impacts of new technology, human psychological factors, and stochastic behaviours on activity choice and pattern formation to enhance the flexibility and applicability of travel demand modelling.

The major contributions of this dissertation to the state of the art include:

1. The activity and time use study by closing the existing knowledge gap with innovative and econometric models that are flexible and transferable in nature, which can be easily applied across disciplines. The literature surrounds existing modelling framework that evaluates traveller's activity choice, time use decisions, and scheduling.
2. Synthetic experiments and a real case study are conducted to test the performance of GEV type models and Probit model when negative correlations present among random components of each alternative.
3. The development of a discrete-continuous choice model to capture the interdependence among multiple activity choices and related time usage and the substitution effects of computer use on in-home/ out-of-home activities in Chapter 5.
4. The use of recursive choice model system to capture the mutual effects between activity choices and psychological factors in Chapter 6.
5. The use of an extended hybrid choice model to study the impacts of psychological factors on activity choices and time usage decisions in Chapter 7.
6. The development of an iterative simulated log-likelihood estimation method that has been prove to produce consistent estimates and corrected inference for correcting endogeneity bias in the hybrid choice modelling framework in Chapter 7.
7. An integrated dynamic activity scheduling model constructed with a dynamic discrete choice model and a Markov decision model are proposed to approach

individuals' stochastic activity-travel behaviour in Chapter 8. Within day and day-to-day variations are captured in the model through dynamic variables.

Furthermore, it is expected that proposed models could be incorporated into a macroscopic planning model which is able to produce a robust evaluation of the travel demand.

Chapter 2: Literature Review

2.1 Activity Involvement and Scheduling

Activity involvement and time use analysis are important components of activity choice modelling. A large number of studies investigated the factors that affect an individual's activity involvement and travel patterns. In this context, socio-demographic characteristics, individual and family schedules, spatial and temporal constraints are found to significantly influence activity participation and travel behaviour (McNally et al., 2007; Kemperman and Timmermans, 2008). Several other studies revealed that individual's activity location choices, which are constrained by space and time, are always associated with daily travel patterns and activity schedules (e.g. Bhat and Gossen, 2004; Bhat and Lockwood, 2004; Lin and Wang, 2015). The location choice problem between at-home and out-of-home is particularly important for discretionary activities. Related studies showed that activity attributes have a greater impact on the activity location than socio-demographics based on their marginal effects. What's more, the characteristics of activities conducted prior and directly following the individual activity have a significant impact on its location choice. Also, longer work duration and commuting time could lead to lower participation in short, temporally and personally flexible out-of-home discretionary activities (Akar et al., 2011; Akar et al., 2012).

2.1.1 Impacts of Computer and Internet Usage on Activity Choice and Time Usage

Internet usage, other than physical activities, is expected to increase the spatial and temporal flexibility of everyday activities (Schwanen and Kwan, 2008). Since the 90s, researchers in transportation have attempted to disentangle the effects of information communication technology (ICT) on travel patterns (e.g. Hamer et al., 1991; Pendyala et al., 1991; Balepur et al., 1998). Their work indicated that the effects of teleworking and of other online activities on personal travel are balanced or outweighed by new trip generation (Handy and Mokhtarian 1996, Mokhtarian, 1991, 1997, 1998; Mokhtarian et al., 1995, Mokhtarian and Salomon, 1997). More recently, researchers began to pay attention to the impacts of ICTs on the involvement in other physical activities, such as shopping, leisure, and social activities. Complementarity and substitution are the most common effects found to be associated with internet use. Mokhtarian et al. (2006) explored the potential impacts of internet use on leisure trips. This study indicated that internet use enables relocation of time to other activities by replacing traditional leisure activity with ICT-based counterparts.

With relation to the effects of e-commerce on shopping trips, it is difficult to reach a definitive consensus on the changes in travel behaviour due to e-shopping. If, on one hand some studies found that the expansion of the e-commerce has contributed to the reduction of shopping trips, but only in a limited way (Mokhtarian, 2004; Weltervreden, 2007; Visser and Lanzendorf, 2004); on the other hand, some other studies found that e-shopping could induce even more physical shopping trips (Douma et al., 2004; Cao et al. 2010; Wilson et al., 2007; Farag, 2007).

The impact of internet use on business and personal travel has been explored by Wang and Law (2007). A positive effect of internet usage was found on the participation in out-of-home recreation and its associated travel activities. Robinson and Martin (2010) indicated that internet users seem to spend less time on other types of activity but have a higher frequency of social trips compared to non-users.

Existing papers more specifically dealing with the effect of social media usage on decisions related to travel and activity participation are more reviews or conceptual papers (Aguiléra et al., 2012; Dal Fiore et al., 2014) rather than empirical works, and the few latter are only partially covering the issue. Ben Elia et al. (2014) use data gathered in 2007, when social network usages were much lower, to investigate the relationships between use of ICT mobile devices, activity, and travel. Le Vine et al. (2016) focus on the relationship between internet use and time spent traveling or in out-of-home activities. To the state of the art, the most recent studies take a different perspective, since they rather assess the possibility of "harvesting transport-related information from social media" (Gal-Tzur et al., 2014) for example to better monitor traffic flows (D'Andrea et al., 2016), incidents (Zhang et al., 2015; Gu et al., 2016), service disruptions (Pender et al., 2014), transit performances (Dey et al., 2016), mode choice (Mondschein, 2015), O/D matrices (Lee et al., 2016) or activity locations (Hasan et al., 2014; Maghrebi et al., 2015). Works more specifically dealing with the impacts of social media on travel demand are limited to the study of route choice (Chen et al., 2015) or derived empirical rules from data streams rather than formal models

(Gkiotsalitis et al., 2014). To sum up, previous knowledge in this area is still rather fragmented and more empirical work is needed in particular.

2.1.2 Impacts of Emotions on Activity Choices

Daily activities, like working, studying, escorting, shopping, and various other forms of pastimes, constitute a key part of people's lives. Among them, leisure activity plays an important role for providing opportunities to improve the quality of life and to satisfy social needs. It also influences individual's daily emotions and satisfaction. Emotions during and after activity are also critical in determining whether individuals want to maintain their involvement (Biddle et al., 2003). As a component of subjective well-being (SWB), together with life satisfaction, emotions represent individuals' cognitive and affective evaluation of their own life (Diener, 1984; Kahneman 1999; Kahneman and Krueger 2006). Happiness, as a part of SWB, has attracted a plethora of cross-disciplinary research in recent year. From a broad context, earlier studies pointed out that SWB could be affected by culture, personality, household and individual socio-demographic. (Diener et al, 2003; Myers and Diener 1995; Ryan and Deci 2001; Dolan et al, 2008).

Several studies have examined the relationship between happiness and activity-travel behaviour. Enam et al. (2015) analyzed the relationship between SWB and daily activity engagement choice of elderly Americans. Focusing on the same population group, Ravulaparthy et al. (2016) investigated the link between time use and subjective well-being among different activities. Two latent class models are jointly estimated to explore the link between time use decisions and people's subjective well-beings.

Ettema et al. (2010) reveled that cognitive and affective components exist between happiness and daily work commute. Bergstad et al. (2012) found that people's mood, and cognitive subjective well-being can be influenced by out-of-home activities. In addition to that, Archer et al. (2012) suggested that out-of-home activities are associated with higher levels of happiness than in-home activities. Diener et al. (1999) pointed out that fulfilment of psychological needs as well as leisure activities may become an important source of individual's SWB. Participation and opportunities for leisure activities that predict SWB could vary across individuals and cultures (Diener et al. 2003; Iwasaki, 2007). Most of the previous studies found that a positive relationship exists between SWB and participation in physical leisure activities (Leung and Lee, 2005), social activities (Lloyd and Auld, 2002; Robinson and Martin, 2008). Kahneman and Krueger (2006) investigated the relationship between emotions and time allocations to each activity. Their research indicated that leisure activity has the highest net effects, which means that this type of activity could bring a higher level of positive emotions (happy, warm, enjoying myself) than any other type of activity. However, Enam et al. (2015) debated that the effects of happiness on leisure activity involvement is negative. A stream of studies tried to explain how psychological well-being is influencing individual's behaviour on travel mode choice (Ettema et al., 2016), and trip duration (Ettema et al., 2012; Stutzer and Frey, 2008). These studies investigated the role of emotions in a single decision-making process. However, in the real world, decision making on activity participation is always correlated with time use choices. Incorporating emotions into an activity-time use joint modelling framework constitutes the major contribution of this thesis.

As noted by Zeelenberg et al. (2008), emotions are referred as endogenous when the experience is relevant to the decision at hand, and are regarded as an integral part of the goal setting and goal striving process. When incorporating such psychological variables into choice model, resulting bias and inconsistent estimates could not be ignored. Fortunately, one way to correct this bias is through two-stage least square (2SLS) methods which have been discussed extensively in the literature (Achen, 1986; Amemiya 1978; Maddala 1983). Appropriate estimation procedures exist to account for the endogenous variable in a simultaneous equation system are introduced by Heckman (1978), Amemiya (1978), and Maddala (1983). Recently, Keshk (2003) introduced a program in STATA that implements a two-stage Probit least square method and provides all the necessary procedures for obtaining consistent estimates for the coefficients. However, these methods are only able to cope with endogenous variables in simple modelling framework. The work needs to be expanded to deal with the situation when endogeneity exists in a more complex model system. Bhat (2015) proposed a generalized heterogeneous data model (GHDM) which approach a mixed modelling system that jointly estimate mixed type of dependent variables. The applied Maximum Approximate Composite Marginal Likelihood (MACML) estimation approach is able to produce a consistent and unbiased estimation results and identifies endogeneity exists among endogenous dependent variables. Later, Bhat et al. (2016) used the modelling system to analyze a bundle of household choices on residential location, vehicle ownership, parents' commute mode choice, and children's school mode choice, along with measurement variables for two latent constructs. However, the modelling system assumes the error terms are independent across dependent variables.

2.2.2 Models on Activity scheduling

Various theoretical and analytical methods have been proposed to model activity scheduling behaviour, a consensus has not to be reached due to the complex nature of the problem. The most general tool that researchers always use to approach the problem is the Random Utility Maximization (RUM) framework. Models of trip chaining and activity scheduling (e.g. Adler and Ben-Akiva 1979; Kitamura, 1984; Ben-Akiva and Bowman, 1995) approach the optimal tours which result in individual's internal utility maximization. A strong assumption of RUM -model is that individuals make rational decisions to optimize their internal utility. As noted by Ben-Akiva et al (1998) that the combinations of tour elements (e.g. the activities of the tour, the timing and locations) results in a large pre-defined choice set.

Research in cognitive psychology has shown that the scheduling and execution of activities often involve a dynamic adjustment. Individuals would adjust their decisions due to unexpected conditions and constraints. To analyse the dynamic natural of activity pattern, Arentze and Timmermans (2009) developed a dynamic activity generation model based on the assumption that individuals' activities are driven by needs. The utility of an activity increases with the satisfied needs and decreases with needs it induces. Using the concept of a utility-of-time threshold they formulate a simple decision and learning rule that generates multi-day patterns based on the optimized utility. Later Arentze et al. (2011) developed an approach to estimate the dynamic models of activity generation on one-day travel-diary data. The dynamic model predicts multi-day activity patterns based on dynamic needs as well as day-varying preferences and time-budgets. Cirillo and Axhausen (2010) proposed a dynamic model of activity choice and scheduling that integrated a mixed logit model to approach state-dependency of choices using Mobidrive multi-week dataset.

Most existing theories and models of activity scheduling behaviour reveal behaviour patterns rather than decision processes. As an alternative, process models focus on the travellers' choices by assuming behaviour process and incorporating constraints, habits, learning, etc. The research attracts increasing attention. For instance, Pendyala et al. (1998) derived a rule-based activity-scheduling algorithm to predict activity scheduling and mode choice. Markov chain is another way to approach the decision-making process. Goulias (1999) used a Markov process model and PSTP dataset to explore the time use and activity-pattern. Ben-Akiva (2010) proposed a planning-action model that transition among travel modes is modelled as a Markov process.

Other than that, numerous classifications of activity-based model with specific modules have been observed in the literature which link choice modelling to trip/activity generation, including utility-based micro-simulations and hybrid simulation/ computation process systems (Bowman and Ben-Akiva, 1997; Meyer and Miller, 2001). Such simulation systems mimic the travel decision making processes, including ALBATROSS (Arentze and Timmermans, 2000), PCATS (Kitamura et al. 2005), TRANSIMS (Guin 2010).

2.2 Previous studies on advanced discrete choice models

2.2.1 Review on Multiple Discrete-continuous Model

On a methodological viewpoint, most of the papers referenced so far are limited to the analysis of individuals' activity participation and ignore the time associated with each activity, which is important for activity scheduling and for the definition of travel patterns. Indeed, jointly considering both categorical and metric endogenous variables is traditionally seen as rather challenging. However, models that accommodate discrete and continuous decisions have recently emerged in the activity-based analysis (Bhat, 2005; Habib et al., 2008; Habib et al., 2009; Srinivasan et al., 2006; Copperman et al., 2007). Discrete-continuous models enable researchers to capture the correlation that potentially exists between individual's discrete and continuous choices.

Bhat (2005) developed multiple discrete-continuous extreme value (MDCEV) models and applied them to model participation in discretionary activity and the duration of time investment. The model framework was then adopted to analyze children's after school out-of-home activity-location engagement patterns and time allocations (Paleti et al. 2011). Pinjari and Bhat (2010) developed a multiple discrete-continuous nested extreme value (MDCEV) model to estimate non-worker activity time-use and scheduling behaviour. However, the MDCEV type models are restricted by the assumption of fixed total time budget allocated to the considered activities. This limits the ability of the analyst to analyze change in time use due to changes in the independent variables included in the model formulation. Habib et al. (2008) developed a discrete-continuous model to estimate the relationship between social contexts, activity starting time and activity duration. A multinomial logit model is employed to capture “with whom” choices of social activities and a hazard model is adopted to capture related activity durations and starting time. This framework poses assumptions on the correlation structure that can be estimated between the discrete and the continuous dependent variables.

More recently, Liu et al. (2014) introduced a discrete-continuous modelling framework, which relaxed the constraints outlined in previous researches. A multinomial Probit model is used to estimate discrete choices, and a regression is used to estimate the continuous decisions. Correlations across the discrete and the continuous parts are captured with a full variance-covariance matrix of the unobserved factors. The modelling framework was further extended by Liu and Cirillo (2015), which allows the specification of multiple regressions for each continuous component in the framework. This latter development will be adopted in this work to estimate joint models that describe leisure involvement (including social media), location (in home vs. out of home) and time spent on each of the considered activity types. Duration of different leisure activities could in fact have different determinants according to the kind of activity under consideration.

2.2.2 Review on Hybrid Choice Model

Hybrid choice models (HCM) gained attention in recent years for their capability of identifying unobservable factors and including such factors into a discrete choice analysis. The model can be viewed as an expanded discrete choice modelling framework, which incorporates different types of models into a single structure (Ben-

Akiva et al. 2002). A basic HCM incorporates a latent variable model specified by a parametric relationship into a discrete choice model based on utility functions. The model typically consists of a set of structural and measurement relationships. Three types of model are generally defined in the modelling framework, which are: the structural equation, the measurement function and the choice model. The structural equation (Eqn. 2.1) indicates the relationships between observable exogenous variables and the latent variables. The measurement equation builds the link between indicator variables and latent variables. In the measurement equation, indicators normally present individual's perceptions or responses to survey questions regarding different attitude, which are normally expressed by a linear (Eqn. 2.2a) or an ordinal (Eqn. 2.2b) form:

$$Z^* = X_{Lat}^T \beta_{Lat} + \eta, \eta \sim N\ (0, \sigma_\eta^2), \tag{2.1}$$

$$I = \varsigma + \Lambda Z^* + \varepsilon, \varepsilon \sim N\ (0, \sigma_\varepsilon^2) \tag{2.2a}$$

$$I = \begin{cases} 0, & if\ Z^* < 0 \\ 1, & if\ 0 \le Z^* < \gamma_1 \\ & \vdots \\ k-1, & if\ \gamma_{k-2} \le Z^* < \gamma_{k-1} \\ k, & if\ \gamma_{k-1} \le Z^* \end{cases} \tag{2.2b}$$

where Z^* indicates a vector of latent variables, X_{Lat} is a vector of observable explanatory variables, ς is the intercept. $\beta_{Lat}$and Λ are the coefficients related to X_{Lat} and Z^* respectively. η and ε are the error terms and are assumed to follow a normal distribution.

Choice model is usually formulated as a typical random utility-maximization model, which consists of the observable exogenous variables and latent variables. The utility function of the choice alternatives j can be expressed as follows:

$$U_j = X_j^T \beta_j + \vartheta_j(Z^*) + \xi_j, \tag{2.3}$$

where β_j are the coefficients that correspond to socio-demographic attributes, ϑ_j is the load of the latent variable on utility, and ξ follows a normal distribution when the choice model is formulated as a Probit model or a Gumbel distribution if the choice model is of GEV type.

The joint likelihood function of these three models is defined by assuming that the random disturbance terms are independent. Then it is possible to write the individual likelihood function as:

$$\mathrm{L}\big(\theta, X_{LAT}, X_j\big) = \int f(Y = j|\beta_j, \vartheta_j, \xi_j) f(I|\varsigma, \Lambda, \sigma_\varepsilon) f(Z^*|\beta_{Lat}, \sigma_\eta) dZ^* \tag{2.4}$$

Starting from this HCM modelling framework, several progresses have been made to incorporate latent psychological constructs into traditional choice models. Walker and Ben-Akiva (2001) first estimated HCM that integrated latent psychological constructs, such as attitudes and preferences, into traditional logit choice models. Bolduc et al. (2008) applied the hybrid choice model to study customer’s perceptions and attitudes towards technological innovations. Later, Bolduc and Alvarez-Daziano (2010) considered both a Probit and mixed multinomial logit discrete choice kernel in the HCM by using simulated maximum likelihood method. These models enable researchers to investigate the influence of psychological variable on the decision-making process. However, the earlier formulations proposed restricted the indicators in the measurement equation to be only of one type (continuous or ordinal). Bhat and Dubey (2014) proposed an integrated model for choices and latent variables based on a multivariate Probit (MNP) kernel, which enables a more flexible covariance structure of the random error terms. Measurement equations with continuous and ordinal indicators are allowed to be jointly estimated with choice model. However, as mentioned in previous section, the model needs to be improved to cope with the endogenous variables in the multiple decision-making processes.

Concerning the estimation method, three different approaches have been proposed and applied: the maximum likelihood (ML) method based on numerical integration (Kim et al. 2012; Glerum et al. 2013; Mabit et al. 2014), simulated maximum likelihood (SML) estimation (Kamrgianni et al. 2014; Jensen et al. 2013; Bhat and Dubey 2014), and Bayesian estimation (Alvarez-Daziano and Bolduc 2013). The ML calculates the likelihood value by numerically integrating the joint likelihood function. However, the approach becomes infeasible as the number of variables increases or when the choice model is formulated as a Multinomial Probit Model. In the latter case, a simulation-based estimation approach such as simulated likelihood and Bayesian estimation are more appropriate. SML is similar to the numerical integration, except that the simulated probabilities are used instead of exact probabilities. The simulated probabilities in SML are obtained by random draws of the latent variables from their probability distributions (Train, 2009). These estimation methods have been proved to produce consistent and unbiased results in previous studies. However, the methods described need to be enhanced when considering endogeneity due to psychological factors.

2.2.3 Review on Dynamic Discrete Choice Model

Dynamic discrete choice models (DDCMs) are widely used in economics and related fields. The models are used to approach decisions as sequences of stochastic discrete choices where at each time decision makers choose alternative that maximize their current and future utilities. They are useful tool for the evolution of price elasticity, intertemporal substitution, and new policies in marketing. In the structure of DDCMs, agents are forward looking and maximize expected intertemporal payoffs, with the knowledge of the evolution of product attributes such as price and technology. The earliest generation of research on DDCM includes Wolpin (1984) on fertility and child motality, Miller (1984) on job matching and occupation choice, Pakes (1984) on patent renewal, and Rust (1987) on machine replacement. In the pioneering work of dynamic model proposed by Rust (1987), the dynamic model is formulated as an optimal stopping problem and is used to estimate the optimal time to replace a bus engine. The model is conceived for singe agent, homogeneous product, and infinite time horizon. Random components of the model are assumed to additively separable, conditionally independent and extreme value distributed. Melnikov (2013) extended the model to consider a binary decision on printer purchases. Moreover, heterogeneous products and homogeneous consumers. Same as Rust's model, He forms the problem as an optimal stopping model, in which consumers will be out-of-market once they make a purchase. Lorincz (2005) expanded the Rust model by allowing consumer who already has a product to upgrade it instead of replacing it.

Despite the vast use of DDCM, there is limited application in the field of transportation.

In transportation, the majority of DDCMs account for consumer's previous actions. Ben-Aliva and Abou-Zeid (2007) proposed a DDCM with the integration of Hidden Markov Chain to model sequence of choice decisions of driving behaviours and the evolution of latent variables. Gao et al., (2010) proposed a policy routing choice model with a cumulative prospect theory utility function (a non-expected utility framework) to measure traveller's route choice when information of the stochastic en-route network is updated. Alternatively, Fosgerau et al. (2013) developed a dynamic route choice model where the path choice problem is formulated as a sequence of link choices. Cirillo et al. (2015) proposed a DDCM with regenerative optimal stopping formulation in order to capture vehicle purchase time and vehicle type choices in a dynamically evolving vehicle market. Later, Serulle and Cirillo (2017) proposed a mixed DDCM to accommodate taste variations in the coefficients in the model. The model is used to approach the optimal time to evacuate under an emergency situation. Other than that, DDCMs are also applied to approach lane changing behaviour of motorcycle drivers (Dong et al. 2017), and optimal departure time (Dong and Cirillo, 2018).

Chapter 3: Data Sources

3.1 American Time Use Survey (ATUS)

The primary data sources used in this Chapter4, Chapter 5, Chapter 6, and Chapter 7 are extracted from the 2013 American Time Use Survey (ATUS) (Bureau of Labor Statistics, 2014). The ATUS is designed and collected by the U.S. Bureau of Labor Statistics and contains detailed information on time use for each activity on which respondents have been involved the day before the interview. Activity related attributes include the start and end time of participation, activity type, and activity location; individual and household socioeconomic characteristics are also available in the basic module of the survey. Both in-home and out of home activities are reported, which makes ATUS particularly attractive for time use analysis and modelling.

In this dissertation we are interested in leisure activity involvement, in the location where those activities take place and the time spent for leisure. We distinguish between in-home and out of home leisure activities and between generic leisure activities and those involving the use of the computer. In particular, we refer to the ATUS category "*Computer use for leisure*"; this variable explicitly excludes games, listening to music, watching videos, e-mails, computer use for work and volunteer activities, which are included in different activity categories. Therefore, we argue that this activity category is mainly time spent online to use social media; a comparative study based on ATUS and a survey conducted by Nielsen supports our claim and concludes that "*the top* *leisure uses included in the ATUS variable are social networks, portals and search*". (Greenstein and Tucker, 2015).

A total of 5,612 observations are available for weekends, while 5,594 observations are available for weekdays. Household characteristics, land-use variables and time use information for each household, are the main variables extracted from the original dataset. Table 3.1 lists the basic statistics relative to the 2013 ATUS sample. We can observe that individuals with no leisure activity have the highest travel time to work, travel to social and entertainment activities and in average have more children. Individuals who use the computer for leisure activities have high income and spend also time on art and entertainment related activities. These trends are similar for weekdays and weekends. In average about 1.5 hours per day are spent on computer for leisure during weekdays and about 2.3 hours per day during weekends among observations who choose computer use as a method of leisure. While average time spent on all leisure activity is about 3.1 hours per day during weekdays and about 3.8 hours per day during weekends among all observations.

Table 3.1 Descriptive statistics of attributes in the basic module

	By activity types					
Variables	**NL**	**LPC**	**LH**	**LOH**	**LH& LPC**	**LH& LOH**
Weekday:						
Gender (female = 1; otherwise = 0)	0.558	0.579	0.572	0.557	0.579	0.464
Metropolitan status (metropolitan = 1; otherwise = 0)	0.843	0.895	0.828	0.784	0.839	0.850
Working status (full time = 1; otherwise = 0)	0.628	0.553	0.404	0.603	0.333	0.600
No. of people in household	2.396	2.289	2.185	2.134	2.191	2.136
Age (years)	42.0	41.1	50.5	41.6	49.3	50.0
Household income ($)	77025	85903	61418	63961	72588	59232

No. of children in Household		1.1	1.1	0.8	0.7	0.7	0.7
Household type	1.Married,	56.2%	60.5%	51.8%	43.8%	55.8%	47.0%
	2. Unmarried,	20.6%	13.2%	17.2%	22.7%	14.0%	16.5%
	3.Single,	23.2%	26.3%	31.0%	33.0%	30.0%	36.6%
	4. Group	<1.00%	<1.00%	<1.00%	<1.00%	<1.00%	<1.00%
Travel time related to working (hrs.)		0.496	0.431	0.297	0.378	0.198	0.42
Travel time related to socializing and communicating (hrs.)		0.032	0.016	<0.010	0.012	<0.010	<0.010
Travel time related to arts and entertainment (hrs.)		0.104	0.103	0.069	0.161	0.057	0.088
Time spent on in-home leisure activity (hrs.)		NA	NA	4.204	NA	3.811	3.079
Time spent on in-home computer use for leisure activity (hrs.)		NA	1.500	NA	NA	1.475	NA
Time spent on out-of-home leisure activity (hrs.)		NA	NA	NA	1.842	NA	1.066
Weekends:							
Gender (female = 1; otherwise = 0)		0.636	0.625	0.542	0.563	0.540	0.503
Metropolitan status (metropolitan = 1; otherwise = 0)		0.834	0.825	0.818	0.820	0.849	0.832
Working status (full time = 1; otherwise = 0)		0.597	0.463	0.433	0.494	0.394	0.463
No. of people in household		2.370	2.288	2.203	2.241	2.191	2.140
Age (years)		44.3	41.0	50.0	44.4	48.5	45.0
Household income ($)		74030	81025	61709	65120	68533	57727
No. of children in Household		1.1	0.8	0.8	0.9	0.7	0.7
Household type	1.Married	53.0%	53.8%	52.3%	44.8%	54.7%	41.5%
	2. Unmarried	21.8%	18.8%	16.9%	22.2%	13.1%	22.9%
	3.Single	25.2%	27.5%	30.7%	33.0%	32.0%	35.6%
	4. Group	<1.00%	<1.00%	<1.00%	<1.00%	<1.00%	<1.00%
Travel time related to working (hrs.)		0.105	0.061	0.099	0.027	0.139	0.080
Travel time related to socializing and communicating (hrs.)		0.212	0.138	0.114	0.148	0.082	0.148
Time spent on in-home leisure activity (hrs.)		NA	NA	4.948	NA	4.511	3.334
Time spent on in-home computer use for leisure activity (hrs.)		NA	2.333	NA	NA	1.546	NA
Time spent on out-of-home leisure activity (hrs.)		NA	NA	NA	3.453	NA	1.929

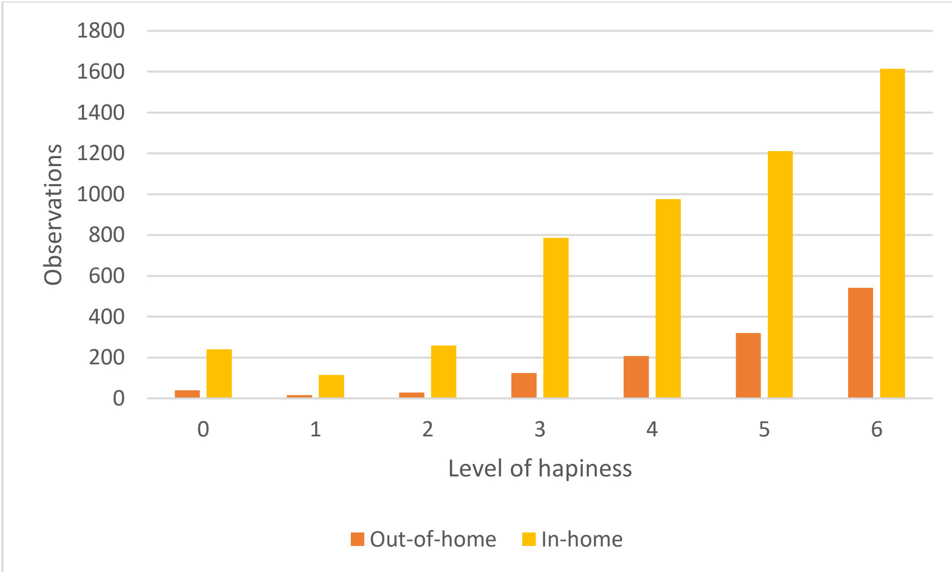

Figure 3.1 Distribution of Leisure Activity Participation and Well-Being Indicators

Besides basic module, the survey also includes Well-Being (WB) module in 2010, 2012, and 2013. The WB Module data files contain information related to how people felt during selected activities, as well as general health information. All respondents in the ATUS survey are selected in the WB module. Three activities from respondents' diaries are randomly selected and 7 questions related to the emotions during activities are asked. The module also includes general information on health status and life satisfaction of each respondent. In the Chapter 6 and Chapter 7, we selected *"Level of Happiness"* during selected activities as the emotion indicator (EI) to represent the emotions and feelings that respondent experienced. The emotion indicator is classified into 7 levels that ranges from 0 to 6 (0 is the lowest level of happiness, 6 is the highest level of happiness).

Besides that, we address the role of participations of other activities (i.e. working, shopping, and socializing) and respondents' social demographics in leisure activity choices. Respondent's age, employment status, education, and time spent on other activities are taken into consideration. Other explanatory variables, such as household income, household size, number of children, and location of household are also included in the model selection part. To enable the analysis, information from ATUS and WB model are linked to create sample. Table 3.2 indicates the characteristic's distribution in the combined dataset and the distribution of respondents' characteristics from entire population. Most variables in selected sample keep consistent with the entire population. Among selected population, time usage on commuting, time usage on working, number of household, and number of children have lower average value. While percentage of single respondents is higher among selected population.

Table 3.2 Definitions and sample statistics of variables in well-being module

	Min	**Max**	**Median**	**Mean**	**S.D.**
Travel time to work	0.00	4.52	0.00	0.13	0.38
Time spend on working	0.00	23.17	0.00	1.73	3.44
Travel time to shop	0.00	9.67	0.00	0.19	0.47
Time spend on shopping	0.00	14.00	0.00	0.32	0.73
Age	15.00	85.00	51.00	50.14	18.74
Male (dummy)	0.00	1.00	0.00	0.48	0.50
Income (*$10000)	1.18	15.00	5.50	6.04	4.03
Metropolitan status (dummy)	0.00	1.00	1.00	0.83	0.38
Higher than Bachelor degree (dummy)	0.00	1.00	0.00	0.11	0.31
Full time working status (dummy)	0.00	1.00	0.00	0.44	0.50
Half time working status (dummy)	0.00	1.00	0.00	0.10	0.29
Couple (dummy)	0.00	1.00	0.00	0.44	0.50
Single (dummy)	0.00	1.00	0.00	0.30	0.46
Number of household	1.00	11.00	2.00	2.48	1.45
Number of children in household	0.00	9.00	0.00	0.64	1.04
Well-being module only:					
Interaction during activity	0.00	1.00	1.00	0.55	0.50
Meaningful of the activity	0.00	6.00	4.00	3.98	2.00
Level of happiness	0.00	6.00	5.00	4.44	1.60
Level of stressfulness	0.00	6.00	0.00	1.03	1.63
level of tiredness	0.00	6.00	2.00	2.35	2.01

level of pain	0.00	6.00	0.00	1.00	1.69
level of sadness	0.00	6.00	0.00	0.64	1.39

3.2 “In the Moment” Household Travel Survey (ITM)

The primary data source used in Chapter 8 is from “In the Moment” Household Travel Survey. The travel Survey is conducted by RSG on behalf of the Madison County Council of Governments (MCCOG) in Anderson, Indianan and the Federal Highway Administration (FHWA) Office of Planning and Office of Transportation Policy Studies. The survey fully replaced the traditional telephone and web household travel diary survey with smartphone GPS data collection over a seven-day period. The smartphone’s sensor passively collected location data, while in-app survey questions obtain the remaining essential household travel survey data elements (Greene et al, 2016).

The survey is collected in the March of 2015 from 528 participants of 200 households in Madison county of Indiana, which results in 13689 trip records in the data set. Among all the participants, 242 participants from 138 households fully completed every survey for all seven assigned travel days. Survey contains trip purpose, trip party, detailed trip mode, and socio-demographics.

Table 3.3 Descriptive statistics

	Min	**Max**	**Mean**	**S.D.**
Individual variables:				
Male	0	1	0.43	0.50
Young (16-34)	0	1	0.35	0.48
Middle age (35-55)	0	1	0.43	0.50

Working status (full time)	0	1	0.63	0.48
Working status (part time)	0	1	0.10	0.30
Bachelor or higher degree	0	1	0.55	0.50
Student status	0	1	0.07	0.26
Low income ($15,000-$34,999)	0	1	0.28	0.45
Median income ($35,000-99,999)	0	1	0.41	0.49
Household variables:				
Household ownership	0	1	0.82	0.39
Household size	1	7	2.83	1.34
Number of vehicle	1	6	2.14	0.79
Number of adult	1	4	1.95	0.58
Number of children	0	5	0.88	1.15

In addition to the seven-day travel diary, personal and household characteristics have been collected, including gender, age, working status, education and household-level. Specifically, Table 3.3 shows the descriptive statistics of the socio-demographics of 242 participants.

One uniqueness of this data is that it provides one of the few, if any, longitudinal travel survey dataset within United States. It is rich in short-term travel behaviour adjustments.

Chapter 4: Bias in Discrete Choice Models of Activity Involvement Choices

4.1 Problem Description

Random utility models (RUM) have been developed considerably in the past three decades (Train 2009) and are extensively applied to many travel related behavioural choices. The wide RUM family consists of two main categories: the GEV models based on the assumption that the errors are type I EV distributed and the Probit model for which the errors are assumed to be multivariate normal. The GEV (McFadden 1978) models include the Multinomial Logit (MNL) and more flexible specifications that allow correlation across choice alternatives while maintaining a closed mathematical form for the choice probabilities. Model formulations that belong to the GEV family include: Nested Logit (Williams 1977); Paired Combinatorial Logit (Chu 1989), Cross-Nested Logit (Vovsha 1997) and General Nested Logit (Wen and Koppelman 2001). The GEV models have been widely applied to model travel mode choice (Hess et al. 2013), spatial location choice (Sener, Pendyala, and Bhat 2011), departure time and route choice (Bekhor, Toledo, and Prashker 2008), and transport networks (Shahhoseini, 2015).

Probit model allows very general error structures (Ben Akiva and Bolduc 1996; Karac-Mandic and Train 2003; Bhat 2011 ; Daziano and Achtnicht 2013; Daganzo 2014), but the associated choice probabilities requires the computation of multivariate normal distribution functions. In particular, the dimension of the integral depends on the number of correlation terms to be estimated and therefore increases rapidly with number of alternatives. Thus, despite the improvements of estimation techniques (Bhat 2003; Bhat 2001; Daziano and Bolduc 2013; Connors, Hess, and Daly 2014), MNL and other GEV models – mainly Nested and Cross Nested Logit are still those most frequently applied in practical applications involving planning, forecasting and feasibility assessments. However, GEV models are based on a set of specific mathematical properties, one of which is the non-negativity in unobserved correlations. Williams and Ortúzar (1982) presented this condition as rigorous and unambiguous. In reality, there is no fundamental reason why non-positive correlations should not occur also from a behavioural perspective. A negative correlation can appear when an explanatory variable or latent factor is omitted from the model specification for some reason (if it is not explicitly measured in the data). If this variable has opposite effects on the utilities of two alternatives, their error terms will have a negative correlation. For example, in mode choice, suppose that attitude towards the environment is part of the "true" model, with a positive sign in the utilities of transit, and a negative sign in the car alternatives. But, if this variable is omitted in the specified model, it will generate negative correlations between the transit and car alternatives. Many such examples can be imagined.

In this section, our motivation is to investigate the bias in the estimation of coefficients and correlation terms deriving from GEV models when errors are negatively correlated. To this scope we design several experiments, we estimate a MNP and we compare the results to those obtained by using NL and CN Logit models. The performance of the three models is assessed with respect to the coefficients' estimates, the ability to recover the correlation among alternatives and the market shares of out-of-sample datasets. Finally, real case study is applied to investigate the activity choice among in-home, out-of-home leisure activities, and leisure activities related with computer/ internet usage.

4.2 Review on GEV Model and Probit Model

4.2.1 The GEV Theory and Deficiencies in MNL

In GEV theory (McFadden 1978), the probability that a given choice maker (n) chooses alternative (i) within the choice set (C) is:

$$P(i \mid C) = \frac{y_i G_i(y_1, .., y_j)}{\mu G(y_1, .., y_j)} = \frac{e^{V_i + \ln G_i}}{\sum_{j \in C} e^{V_j + \ln G_j}} \tag{4.1}$$

whereby: $G_i = \partial G / \partial y_i$, J is the number of available alternatives, $y_i = e^{V_i}$, V_i is the systematic part of the utility function associated with alternative i, and G is a non-negative differentiable function which verifies some specific properties.

Any model that can be derived in this way is regarded as a GEV model. This formulation, therefore, defines the family of GEV models. GEV derived models must respect several distinct properties. These properties have no real behavioural intuition but are a mathematical requirement. The properties that the function G must exhibit are the following:

(1) $G \geq 0$ for all positive values of $y_j \forall$ j.

(2) G is homogeneous of degree one i.e. if each y_j is raised by some proportion ρ, G rises by proportion ρ. (Ben-Akiva and Francois 1983) showed that this condition can be relaxed to allow any degree of homogeneity.

(3) $G \rightarrow \infty$ as $y_j \rightarrow \infty$ for any j.

(4) The mixed partial derivatives of G exist and are continuous with non-positive even partial derivate. That is, $G_i \geq 0$ for all i , $G_{ij} = \partial G_i / \partial y_j \leq 0$ for all $j \neq i$ and non-negative odd mixed partial derivate $G_{ijk} = \partial G_{ij} / \partial y \geq 0$ for any distinct i, j, and k, and so on for higher order mixed partials.

These conditions are sufficient so that $f = e^{-G(e^{y_1},..,e^{y_j})}$ is an absolutely continuous multivariate extreme value distribution function. However, as noted by (Dagsvik 1995), these constraints also imply that the correlations reproduced by a GEV model are necessarily positive.

This general theory consists of a large family of specifications that includes the MNL itself. MNL's main advantage is in its analytical tractability. However, the hypothesis of errors identically distributed causes the property of independence from irrelevant alternatives (IIA), which results in failure to account for similarities between alternatives (Ben Akiva and Lerman 1985). Furthermore, the variance–covariance matrix of the MNL model is homoscedastic.

4.2.2 Nested Logit and Presence of Negative Correlations

Initially proposed by Williams (1977) and Daly and Zachary (1978), the Nested Logit

(NL) model is an extension of the MNL model designed to capture correlations among alternatives by partitioning choice sets into different nests. The NL model is designed for choice problems where the alternatives within each nest have correlated error terms; however, error terms between nests remain uncorrelated. Both MNL and NL models are instances of the GEV family. For the MNL model (Eqn.4.2):

$$G_i(y_1,..,y_j) = \sum_{j\in C} y_j^{\ \mu}, \quad P(i|C) = \frac{e^{\mu V_i}}{\sum_{j\in C} e^{\mu V_j}} \tag{4.2}$$

For the NL model (Eqn. 4.3):

$$G_i(y_1,..,y_j) = \sum_{m=1}^{M} \left(\sum_{j\in m} y_j^{\ \mu_m}\right)^{\frac{\mu}{\mu_m}}, \quad P(i|C) = \frac{e^{\mu_m V_i} \left(\sum_{i\in m} e^{\mu_m V_i}\right)^{\frac{\mu}{\mu_m}}}{\sum_{i\in m} e^{\mu_m V_i} \cdot \sum_{m=1}^{M} \left(\sum_{i\in m} e^{\mu_m Vi}\right)^{\frac{\mu}{\mu_m}}} \tag{4.3}$$

where μ is the scale factor, and μ_m is the parameter associated with nest m. The ratio of $\frac{\mu}{\mu_m}$ is the degree of independence or dissimilarity among the alternatives belonging to nest *m*. This ratio must be within a particular range for the model to be consistent with utility- maximizing behaviour. Following (McFadden 1978), it is possible to assert that the NL model is consistent with utility maximization when $0 < \frac{\mu}{\mu_m} \leq 1$. Furthermore, Daganzo and Kusnic (1993) presented that the correlation between any two alternatives belonging to nest *m* is equal to:

$$Corr(U_{mi}, U_{mj}) = 1 - \left(\frac{\mu}{\mu_m}\right)^2 \tag{4.4}$$

Börsch-Supan (1990), Kling and Herriges (1995), and Herriges and Kling (1996) provided tests of consistency of NL with utility maximization when the degree of dissimilarity is greater than 1 i.e. $\frac{\mu}{\mu_m} > 1$. Train, McFadden, and Ben- Akiva (1987) showed that in this case, consistency with utility- maximizing is appropriate for some specified range of the explanatory variables. Carrasco and Ortuzar (2002) discuss in great details the consistency conditions of Börsch-Supan and the successive corrections by Kling and Herriges (1995) and Herriges and Kling (1996). They highlight that from a behavioural standpoint a greater degree of substitution between nests than within them makes it impossible to test the hierarchical relationship between the different nesting levels. On the other hand, as noted by Train (2009), a negative value of the degree of dissimilarity i.e. $\frac{\mu}{\mu_m} < 0$, is inconsistent with utility maximization and implies that improving the attributes of an alternative (such as lowering its price) can reduce the probability of the alternative being chosen. Finally, when the degree of dissimilarity approaches zero i.e. $\frac{\mu}{\mu_m} \to 0$, NL approaches the "elimination by aspects" model suggested by Tversky (1972). It is interesting to note that the range of $0 < \frac{\mu}{\mu_m} \leq \sqrt{2}$ is the only proper range that is acceptable in terms of the demands for a correctly specified statistical correlation with corresponding values that lie on the range [-1,1].

Furthermore, when $1 < \frac{\mu}{\mu_m} \leq \sqrt{2}$, then there exists a negative correlation between any two alternatives within nest m. This fact certainly cannot be consistent with the GEV theory's assumptions which regard only the possibility of a positive correlation for mathematical reasons.

4.2.3 Cross Nested Logit and the presence of negative correlations

The Cross-Nested Logit (CNL) model was also originally proposed by Williams (1977) discussed in terms of its properties using simulated data by Williams and Ortúzar (1982) and further developed by Vovsha (1997). CNL is an extension of the NL model. However, in addition to the choice set being partitioned into nests each alternative may belong to more than one nest. General Nested Logit (GNL) developed by Wen and Koppelman (2001) is a broader specification than the CNL model. NL is a special case of the GNL model in which the coefficients are binary, either zero or one. Thus an alternative can only belong to one nest. Various formulations for the CNL model have been proposed in the literature (Bierlaire 2006). An adaptation of GNL to model route choice was proposed by Vovsha and Bekhor (1998)

The Paired Combinatorial Logit (PCL) specification is a particular example of the CNL model. PCL is another GEV- type model, proposed by Chu (1989) and later expanded by Koppelman and Wen (2000). It was applied extensively to model route choice by conveniently defining the similarity index (Prashker and Bekhor 1998; Gliebe, Koppelman, and Ziliaskopoulos 1999). In the Nested Logit model all alternatives in a common grouping are similar. In contrast, in the PCL model, each pair of alternatives can have a similarity relationship that is completely independent of the similarity relationship of other pairs of alternatives. This feature is highly desirable for route choice models, since each pair of routes may have different similarities.

Similar to the NL model, CNL also has a GEV generating function and derived probability (Eqn. 4.5):

$$G_i(y_1,..,y_j) = \sum_{m=1}^{M}\left(\sum_{j\in C}\left(\alpha_{jm}{}^{1/\mu} y_j\right)^{\mu_m}\right)^{\frac{\mu}{\mu_m}}$$

$$P(i\,|\,C) = \sum_{m=1}^{M}\frac{\left(\sum_{i\in m}\alpha_{jm}^{\frac{\mu_m}{\mu}} e^{\mu_m Vj}\right)^{\frac{\mu}{\mu_m}}}{\sum_{n=1}^{M}\left(\sum_{j\in C}\alpha_{jm}^{\frac{\mu_m}{\mu}} e^{\mu_m V_j}\right)^{\frac{\mu}{\mu_m}}}\cdot\frac{\alpha_{im}^{\frac{\mu_m}{\mu}} e^{\mu_m Vi}}{\sum_{j\in C}\alpha_{jm}^{\frac{\mu_m}{\mu}} e^{\mu_m V_j}} \tag{4.5}$$

where: $\alpha_{jm} \geq 0, \forall j$; $\sum_{m=1}^{M}\alpha_{jm} > 0, \forall j$; $\mu_m > 0, \forall m$; $\mu_m \geq \mu, \forall m$.

However, unlike NL, the correlation between alternatives in overlapping nests is not a simple formula. Papola (2004) proposed a conjecture regarding the approximate structure of this correlation (Eqn. 4.6):

$$Corr(U_{mi}, U_{mj}) = \sum_{m=1}^{M}\alpha_{im}^{1/2}\alpha_{jm}^{1/2}\left(1-\left(\frac{\mu}{\mu_m}\right)^2\right) \tag{4.6}$$

Abbe, Bierlaire, and Toledo (2007) presented a proof that this correlation has quite a messy structure which is derived from the joint cumulative distribution function (CDF) of the CNL utilities:

$$Corr(U_{mi},U_{mj}) = \frac{6\mu^2}{\pi^2}\iint_{\Re^2} x_i x_j \partial^2_{x_i x_j} F_{\varepsilon_i,\varepsilon_j}\left(x_i,x_j\right) dx_i dx_j - \frac{6\gamma^2}{\pi^2} \tag{4.7}$$

where: $$F_{\varepsilon_i,\varepsilon_j}\left(x_i,x_j\right) = \exp\left(-\sum_{m=1}^{M}\left(\left(\alpha_{im}^{1/\mu}e^{-x_i}\right)+\left(\alpha_{jm}^{1/\mu}e^{-x_j}\right)\right)^{\mu_m}\right)^{\frac{\mu}{\mu_m}}$$

This integral has no closed form and must be estimated using numeric procedures. In addition, this correlation (Eqn. 4.7) is always positive. Just like in the case of NL, there is no reason to suggest that in reality this assumption should always hold.

4.3 Synthetic Experiments

4.3.1 Rational

The inherited assumption of non-negative correlations is brought about by mathematical necessities. However, within elaborate nested structures there is no apparent reason why this assumption must hold. Therefore, we decided to put this to the test by creating artificial correlation structures using synthetic data generation and estimating MNP, and GEV models – NL (Experiment I) and CNL (Experiment II) to measure the obtained bias in the results. MNP unlike GEV can theoretically approximate any correlation structure without bias and should be used whenever the analyst believes that negative correlation could exists. In practice, some restrictions are set on the correlation structure for identification purposes and models' outcomes have some difficulties in their interpretation. For a further discussion of the properties of MNP see (Greene 2008).

4.3.2 Experiment I

A sample of 10 files (runs) each with 3,000 synthetic choice observations was created. The sample was created separately for two choice problems: a choice between three alternatives (Experiment Ia) and a choice between four alternatives (Experiment Ib). Each file contained the deterministic utility for each alternative (V_i) and the error components (ε_i).

The synthetic utilities – both the deterministic and stochastic parts were computed using a *standard normal distribution*. The alternative specific constant of 1st alternative was set to 0 for reasons of normalization. 21 artificial 'true' correlation values (ρ) were assumed to vary from -0.95 to 0.95 with 0.095 interval.

For each correlation (ρ_k), a correlation matrix was computed. For the three-alternative case the correlation matrix is showed as below:

$$Cor_k = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & \rho_k \\ 0 & \rho_k & 1 \end{bmatrix}, k = 1, ..., 21 \tag{4.8}$$

where: Cor_k represents the k^{th} correlation matrix and ρ_k is true value. In the case of four alternative choice set, the correlation matrix was defined separately for positive and negative correlations:

$$Cor_k = \left\{ \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & \rho_k & \rho_k \\ 0 & \rho_k & 1 & \rho_k \\ 0 & \rho_k & \rho_k & 1 \end{bmatrix} \text{if } \rho_k \geq 0 \text{ ,} \quad \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & -\rho_k & \rho_k \\ 0 & -\rho_k & 1 & \rho_k \\ 0 & \rho_k & \rho_k & 1 \end{bmatrix} \text{if } \rho_k < 0 \right\} \tag{4.9}$$

The vectors of errors of all the alternatives except ε_1 were multiplied by the Cholesky factorization of each correlation combination in order to transform the matrix into a product of a lower triangular matrix which is important to maintain the stability in the matrix estimation. The chosen alternative was the one with the maximum utility. Thus, for each of the 21 'true' correlations a corresponding vector of choices was matched.

A NL model was estimated with BIOGEME (Bierlaire 2003) for each of the 21 choice vectors in each of the 10 data sets (in total - 210 models). The NL model had a common nest which included all alternatives apart for 1st alternative. Figure 3.1 presents the structure for the three-alternative model and Figure 3.2 for the four-alternative model:

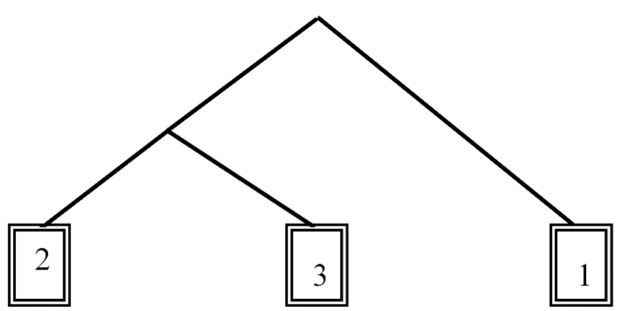

Figure 4. 1 NL model with three alternatives

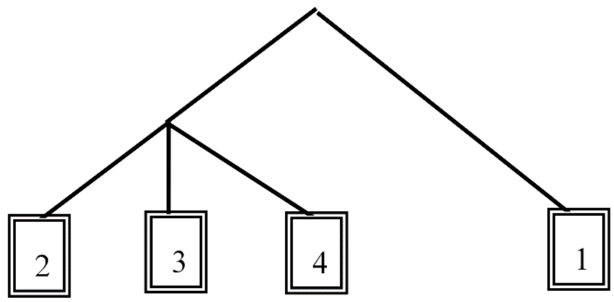

Figure 4. 2 NL model with four alternatives

The NL model was specified according to the following principles:

(1) The utility functions were specified as:

$$U_1 = \beta_i ' V_1$$
$$U_i = \beta_0^i + \beta_i ' V_i \tag{4.10}$$

where:

β_0^i is the alternative specific constant of alternative *i* and $\beta_i ' V_i$ is the observed utility components to alternative i.

(2) The coefficient of the Nest (μ_m) was left to be estimated.

(3) The logit scale (μ) was normalized to 1.

4.3.3 Experiment II

A sample of 10 files (runs) each with 3,000 synthetic choice observations was created using R. The choice was between four alternatives in a similar manner that the data was created in Experiment I.

The artificial correlations were derived from the combinations of the values $(0.75, 0.25, -0.25, -0.75)$ in groups of three. In total *k=20* combinations were created. For example, the combination (0.75, 0.75, 0.75) is the first, (0.75, 0.75, 0.25) the second, etc. The correlation matrix was defined as:

$$Cor_k = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & \rho_k^{12} & \rho_k^{13} \\ 0 & \rho_k^{12} & 1 & \rho_k^{23} \\ 0 & \rho_k^{13} & \rho_k^{23} & 1 \end{bmatrix}, \quad k = 1, \ldots, 20 \tag{4.11}$$

whereby: ρ_k^{ij} is the correlation between alternatives *i,j* and *k* is the combination's number.

Not all the combinations are viable. In five out of the 20 combinations the Cholesky factorization is invalid. This fact reduced the number of combinations from 20 to 15. Similar to Experiment I, the vectors of errors of all the alternatives except ε_1 were multiplied by the Cholesky factorization of each correlation combination. The chosen alternative was the one with the maximum utility. Thus, for each of the 15 'true' correlations a corresponding vector of choices was matched.

A CNL model was estimated with BIOGEME (Bierlaire 2003) for each of the 15 choice vectors in each of the 10 data sets. The CNL model had a PCL specification of three alternatives, except the first alternative. Each alternative has a shared nest with each of the other two alternatives. Figure 3.3 depicts the model structure:

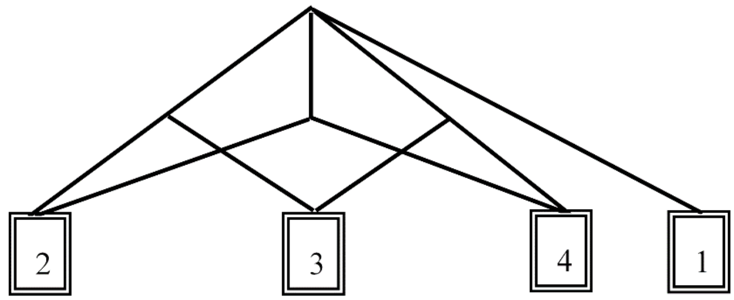

Figure 4. 3 Cross Nested Logit model with four alternatives

The CNL model was specified according to the following principles:

(1) The utility functions are specified as:

$$U_1 = \beta_i ' V_1$$
$$U_i = \beta_0^i + \beta_i ' V_i \qquad (4.12)$$

whereby: β_0^i is the alternative specific constant of alternative *i* and $\beta_i ' V_i$ is the overall utility component specified to alternative i.

(2) The coefficients of the three nests (μ_m) are left to be estimated.

(3) The logit scale (μ) is normalized to 1.

(4) The similarity coefficients (α_{im}), were estimated and the sums for each pair are constrained to 1.

The estimated correlations of the CNL model are computed using Papola's approximation (Eqn.4.6). As Papola's approximation is a conservative estimate of the real correlation, we believe this provides a reasonable estimate of the possible bias compared to the true values.

4.3.4 Normalization of the Covariance Matrix for MNP

In GEV models, the normalization for scale and level occurs automatically with the distributional assumptions that are placed on the error terms. As a result, normalization does not need to be considered for these models. However, with Probit models, normalization for scale and level does not occur automatically. The model should be normalized directly.

The Probit model has n alternatives, and utility function is expressed as $U_i = V_i + \varepsilon_i, i = 1, \ldots, n$. The vector of errors ε_i is normally distributed with zero mean. The procedure proposed by Train (2009) has been applied to normalize the Probit model and assure that all the parameters are identified. The differences with respect to first alternatives are taken, and the error differences is defined as $\tilde{\varepsilon}_{i1} = \varepsilon_i - \varepsilon_1$.

The covariance matrix for the vector of error differences take the form

$$\widetilde{\Omega} = \begin{pmatrix} \theta_{22} & \cdots & \theta_{2n} \\ \vdots & \ddots & \vdots \\ \theta_{n2} & \cdots & \theta_{nn} \end{pmatrix}$$

where: θ is related to the original σ, when the differences are taken against alternative 1. It is showed as follows:

$$\theta_{ab} = \sigma_{ab} + \sigma_{11} - \sigma_{1a} - \sigma_{1b} \tag{4.13}$$

The matrix is obtained using the $(n-1) \times n$ transformation matrix M_1 as

$$\Omega = M_{n-1} \Omega M_{n-1}' \tag{4.14}$$

where: $M_{n-1} = \begin{pmatrix} -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -1 & 0 & \cdots & 1 \end{pmatrix}$

4.4 Results of Synthetic Experiments

4.4.1 Experiment I: MNP and NL with three alternatives

Figure 4.4 presents the results of the estimated correlations $\tilde{\rho}_k$of the MNP and NL models with three alternatives and the true values. There appears no real difference between the results of the NL model and the true values for positive correlations. However, for negative correlations there is a growing gap between the true value and the estimation as the biased correlation estimates of NL still stay negative. We note that the estimated correlations for the MNP model were basically identical to the true values, as expected.

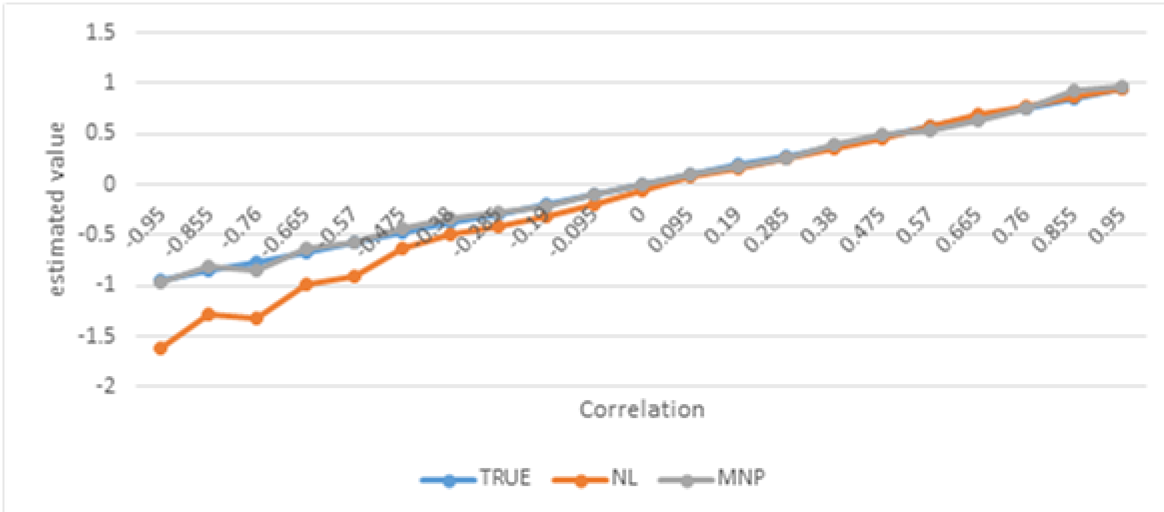

Figure 4. 4 Comparison of correlation ρ ̃_k between NL and MNP model

Table 4.1 presents the result of the estimated coefficients for the MNP and NL models (averaged over the 10 runs). As the correlation value increases, a smaller scale factor is revealed in the results of NL. When $\tilde{\rho}_k = 0$, the scale factor is approximately 1.37, which is consistent with the results in Train (2009). Table 4.1 shows that MNP does produce consistent results when synthetic data is created with error terms following normal distribution.

Table 4.1 Estimation results - Experiment Ia

		Alt1			Alt2			Alt3	
ρ	**Asc1**	**β11**	**β12**	**Asc2**	**β21**	**β22**	**Asc3**	**β31**	**β32**
TRUE	0.000	0.300	1.200	0.800	0.500	1.500	1.500	0.500	1.800
MNP:									
-0.950	0.000	0.295	1.199	0.795	0.481	1.515	1.518	0.506	1.830
-0.855	0.000	0.267	1.222	0.817	0.514	1.470	1.505	0.472	1.809
-0.760	0.000	0.310	1.227	0.796	0.516	1.541	1.526	0.500	1.862
-0.665	0.000	0.287	1.215	0.834	0.504	1.487	1.537	0.480	1.781
-0.570	0.000	0.282	1.222	0.824	0.492	1.532	1.506	0.525	1.828
-0.475	0.000	0.286	1.197	0.785	0.497	1.490	1.486	0.474	1.800
-0.380	0.000	0.321	1.220	0.866	0.486	1.529	1.558	0.485	1.821
-0.285	0.000	0.315	1.175	0.800	0.510	1.525	1.498	0.512	1.786
-0.190	0.000	0.300	1.242	0.818	0.521	1.547	1.542	0.516	1.856

-0.095	0.000	0.315	1.246	0.815	0.492	1.520	1.533	0.526	1.897
0.000	0.000	0.304	1.246	0.838	0.474	1.552	1.553	0.533	1.881
0.095	0.000	0.319	1.213	0.807	0.483	1.539	1.492	0.517	1.832
0.190	0.000	0.305	1.220	0.788	0.481	1.523	1.513	0.534	1.884
0.285	0.000	0.294	1.250	0.811	0.508	1.546	1.536	0.514	1.874
0.380	0.000	0.332	1.224	0.801	0.513	1.554	1.529	0.526	1.869
0.475	0.000	0.292	1.224	0.823	0.500	1.560	1.515	0.536	1.901
0.570	0.000	0.319	1.199	0.814	0.513	1.517	1.518	0.500	1.816
0.665	0.000	0.323	1.213	0.774	0.514	1.536	1.503	0.528	1.861
0.760	0.000	0.294	1.235	0.777	0.497	1.551	1.512	0.515	1.855
0.855	0.000	0.278	1.223	0.816	0.514	1.547	1.529	0.516	1.867
0.950	0.000	0.295	1.248	0.810	0.531	1.577	1.547	0.528	1.901
NL:									
-0.950	0.000	0.436	1.764	1.121	0.705	2.219	2.170	0.732	2.638
-0.855	0.000	0.385	1.774	1.151	0.740	2.109	2.131	0.673	2.572
-0.760	0.000	0.453	1.766	1.108	0.733	2.196	2.141	0.704	2.620
-0.665	0.000	0.416	1.757	1.178	0.721	2.122	2.170	0.681	2.526
-0.570	0.000	0.403	1.716	1.124	0.691	2.141	2.071	0.724	2.517
-0.475	0.000	0.398	1.666	1.062	0.693	2.061	2.027	0.651	2.458
-0.380	0.000	0.435	1.679	1.166	0.665	2.080	2.104	0.653	2.456
-0.285	0.000	0.432	1.607	1.082	0.697	2.071	2.024	0.690	2.409
-0.190	0.000	0.401	1.682	1.086	0.701	2.080	2.052	0.685	2.472
-0.095	0.000	0.422	1.664	1.067	0.647	2.023	2.019	0.687	2.487
0.000	0.000	0.405	1.648	1.085	0.620	2.042	2.027	0.694	2.444
0.095	0.000	0.421	1.606	1.060	0.636	2.024	1.953	0.672	2.400
0.190	0.000	0.396	1.590	1.008	0.620	1.966	1.951	0.680	2.408
0.285	0.000	0.374	1.620	1.043	0.653	1.989	1.974	0.652	2.391
0.380	0.000	0.427	1.564	1.002	0.651	1.977	1.932	0.656	2.346
0.475	0.000	0.373	1.564	1.028	0.638	1.978	1.912	0.674	2.390
0.570	0.000	0.406	1.541	1.038	0.653	1.932	1.933	0.635	2.316
0.665	0.000	0.413	1.558	0.989	0.653	1.953	1.908	0.675	2.366
0.760	0.000	0.374	1.566	0.982	0.623	1.951	1.904	0.648	2.330
0.855	0.000	0.347	1.520	1.003	0.632	1.900	1.887	0.634	2.279
0.950	0.000	0.364	1.547	0.996	0.655	1.938	1.912	0.645	2.326

4.4.2 Experiment I: MNP and NL with four alternatives

Figure 4.5 and Figure 4.6 present the results of the estimation of the correlations of the NL model with four alternatives and the comparison to the true values. Table 4.2 presents the comparison of the estimated coefficients of the MNP and NL models. The results are averaged over 10 runs.

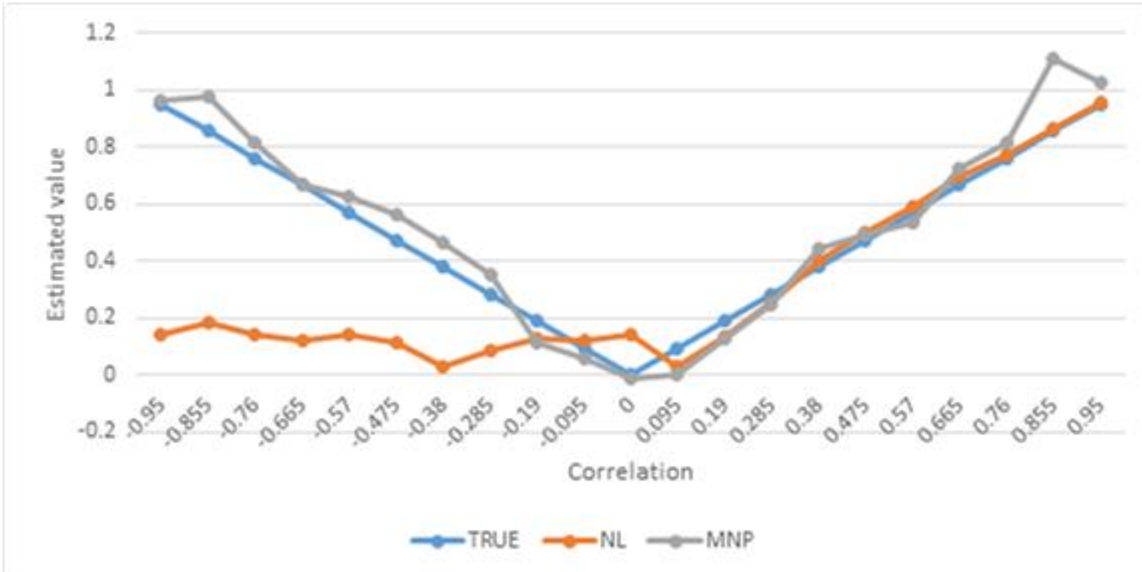

Figure 4. 5 Comparison of $\tilde{\boldsymbol{\rho}}_k^{\mathbf{12}}$ between NL and MNP model

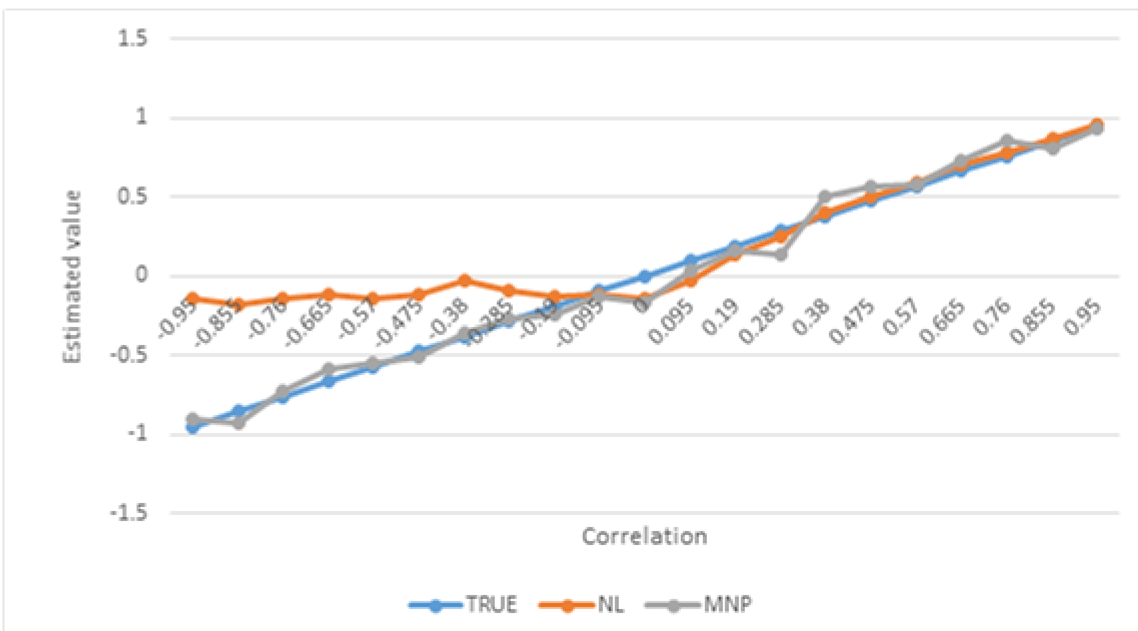

Figure 4. 6 Comparison of $\tilde{\boldsymbol{\rho}}_k^{\mathbf{13}}$ between NL and MNP model

The correlations shown in Figure 4.5 and Figure 4.6, are presented in the form of $\tilde{\rho}_k$.

Consistent with the results obtained in Experiment Ia, NL model presents bias estimates when negative correlation value exists in the correlation matrix. First, for positive correlation there appears no real difference between the results of the NL model and the true values. While with negative correlation, there is a growing gap between the true values and the estimation when correlation is decreasing as it can be seen in both figures. The MNP and true values were basically identical.

Table 4.2 presents the estimated coefficients obtained with the MNP and NL models. Apart for the differences which are attributed to the scale difference between the models, there is no significant difference in the coefficients obtained with NL and MNP. This result is quite remarkable as the correlations clearly show that there is a significant bias in the negative side. However, it seems that the coefficients in the NL model are not influenced by this fact.

Table 4.2 Estimation results - Experiment Ib

		Alt1			**Alt2**			**Alt3**			**Alt4**	
ρ	**Asc1**	**β11**	**β12**	**Asc2**	**β21**	**β22**	**Asc3**	**β31**	**β32**	**Asc4**	**β41**	**β42**
True	0.000	0.300	1.200	0.800	0.500	1.500	1.500	0.500	1.800	0.600	0.600	1.200
MNP:												
-0.950	0.000	0.332	1.197	0.844	0.524	1.527	1.563	0.527	1.844	0.610	0.627	1.199
-0.855	0.000	0.260	1.259	0.841	0.535	1.605	1.585	0.554	1.955	0.615	0.633	1.328
-0.760	0.000	0.308	1.269	0.865	0.531	1.549	1.563	0.508	1.887	0.611	0.626	1.248
-0.665	0.000	0.293	1.233	0.817	0.497	1.505	1.530	0.497	1.825	0.616	0.616	1.167
-0.570	0.000	0.292	1.233	0.831	0.499	1.536	1.527	0.529	1.889	0.543	0.636	1.273
-0.475	0.000	0.299	1.236	0.835	0.509	1.551	1.563	0.485	1.855	0.628	0.618	1.242
-0.380	0.000	0.288	1.202	0.833	0.473	1.443	1.507	0.492	1.769	0.640	0.601	1.152
-0.285	0.000	0.312	1.243	0.851	0.488	1.491	1.552	0.526	1.842	0.618	0.603	1.259
-0.190	0.000	0.320	1.167	0.729	0.520	1.531	1.470	0.525	1.849	0.568	0.594	1.194
-0.095	0.000	0.262	1.224	0.814	0.497	1.533	1.514	0.497	1.828	0.644	0.576	1.200
0.000	0.000	0.320	1.184	0.757	0.524	1.527	1.446	0.505	1.893	0.555	0.608	1.216
0.095	0.000	0.310	1.200	0.759	0.514	1.518	1.472	0.499	1.852	0.543	0.619	1.212
0.190	0.000	0.312	1.210	0.752	0.521	1.566	1.492	0.514	1.845	0.518	0.646	1.262
0.285	0.000	0.315	1.184	0.775	0.516	1.524	1.461	0.498	1.819	0.603	0.581	1.214
0.380	0.000	0.281	1.269	0.878	0.508	1.540	1.579	0.549	1.925	0.621	0.656	1.288
0.475	0.000	0.306	1.235	0.814	0.533	1.564	1.549	0.535	1.869	0.583	0.631	1.270

0.570	0.000	0.306	1.222	0.790	0.515	1.522	1.522	0.504	1.838	0.599	0.629	1.210
0.665	0.000	0.347	1.257	0.831	0.504	1.592	1.556	0.543	1.910	0.585	0.648	1.267
0.760	0.000	0.331	1.273	0.818	0.526	1.591	1.530	0.542	1.915	0.560	0.644	1.304
0.855	0.000	0.308	1.258	0.821	0.540	1.637	1.574	0.521	1.944	0.643	0.628	1.268
0.950	0.000	0.277	1.241	0.778	0.549	1.640	1.534	0.546	1.958	0.563	0.649	1.306
NL:												
-0.950	0.000	0.490	1.760	0.977	0.810	2.344	2.194	0.803	2.741	1.453	0.745	1.431
-0.855	0.000	0.366	1.774	0.931	0.787	2.375	2.130	0.786	2.770	1.383	0.724	1.522
-0.760	0.000	0.437	1.816	1.017	0.780	2.304	2.125	0.741	2.745	1.322	0.755	1.490
-0.665	0.000	0.431	1.789	0.989	0.739	2.262	2.124	0.725	2.641	1.284	0.761	1.445
-0.570	0.000	0.423	1.731	0.988	0.720	2.221	2.059	0.727	2.638	1.122	0.780	1.544
-0.475	0.000	0.417	1.733	1.034	0.713	2.195	2.100	0.671	2.580	1.178	0.764	1.532
-0.380	0.000	0.416	1.720	1.093	0.681	2.054	2.077	0.691	2.473	1.133	0.776	1.484
-0.285	0.000	0.441	1.743	1.086	0.690	2.101	2.098	0.737	2.547	1.052	0.770	1.616
-0.190	0.000	0.452	1.690	1.014	0.741	2.167	2.074	0.732	2.589	0.922	0.815	1.638
-0.095	0.000	0.378	1.762	1.121	0.704	2.165	2.125	0.690	2.534	0.976	0.792	1.654
0.000	0.000	0.460	1.693	1.068	0.726	2.130	2.030	0.691	2.601	0.783	0.853	1.714
0.095	0.000	0.437	1.700	1.044	0.717	2.115	2.038	0.691	2.553	0.771	0.856	1.687
0.190	0.000	0.433	1.661	1.029	0.702	2.111	2.018	0.691	2.477	0.748	0.860	1.694
0.285	0.000	0.439	1.657	1.089	0.698	2.067	2.016	0.670	2.441	0.815	0.804	1.685
0.380	0.000	0.362	1.630	1.085	0.654	1.978	2.004	0.689	2.410	0.823	0.824	1.612
0.475	0.000	0.393	1.589	1.031	0.680	1.984	1.962	0.674	2.351	0.774	0.786	1.594
0.570	0.000	0.404	1.602	1.032	0.662	1.963	1.972	0.647	2.366	0.784	0.815	1.566
0.665	0.000	0.434	1.582	1.037	0.622	1.971	1.945	0.661	2.327	0.763	0.794	1.553
0.760	0.000	0.420	1.585	1.001	0.648	1.952	1.894	0.657	2.323	0.711	0.785	1.590
0.855	0.000	0.387	1.560	0.983	0.649	1.960	1.926	0.613	2.283	0.764	0.758	1.524
0.950	0.000	0.352	1.562	0.986	0.667	1.989	1.919	0.657	2.359	0.726	0.786	1.584

4.4.3 Experiment II: MNP and CNL with three alternatives

Figure 4.7, Figure 4.8 and Figure 4.9 present the comparison between the true covariance and correlation parameters and the PCL model estimates. As noted five out of 20 correlation combinations were not positive semi-definite (i.e. the Cholesky factorization does not exist). These combinations were excluded. Table 4.3 lists the 15 resulting correlation combinations that were used in the experiment. The estimated correlations in the PCL model were computed according to (Papola, 2004) approximation.

Table 4.3 Correlation combinations

k	$\boldsymbol{\rho_k^{12}}$	$\boldsymbol{\rho_k^{13}}$	$\boldsymbol{\rho_k^{23}}$
1	0.75	0.75	0.75
2	0.75	0.75	0.25
3	0.75	0.25	0.25
4	0.75	0.25	-0.25
5	0.75	-0.25	-0.25
6	0.75	-0.25	-0.75
7	0.75	-0.75	-0.75
8	0.25	0.25	0.25
9	0.25	0.25	-0.25
10	0.25	0.25	-0.75
11	0.25	-0.25	-0.25
12	0.25	-0.25	-0.75
13	0.25	-0.75	-0.75
14	-0.25	-0.25	-0.25
15	-0.25	-0.25	-0.75

The results show that PCL specification with multiple nests is hard to estimate. Only about 10 out of 100 runs of the model obtained convergence; the log-likelihood function deriving from a PCL specification is highly nonlinear and non-convex, which causes the convergence failures reported. We note that MNP estimates were basically identical to the true values. Figure 4.7 and Figure 4.8 show the comparison between converged estimates of CNL model, MNP, and the true values. The results show that the estimated correlation have less bias when all three correlations are positive (k=1, 2, 3, 8). However, when negative correlation value appears, bias can result in the correlation matrix even for correlations with positive values.

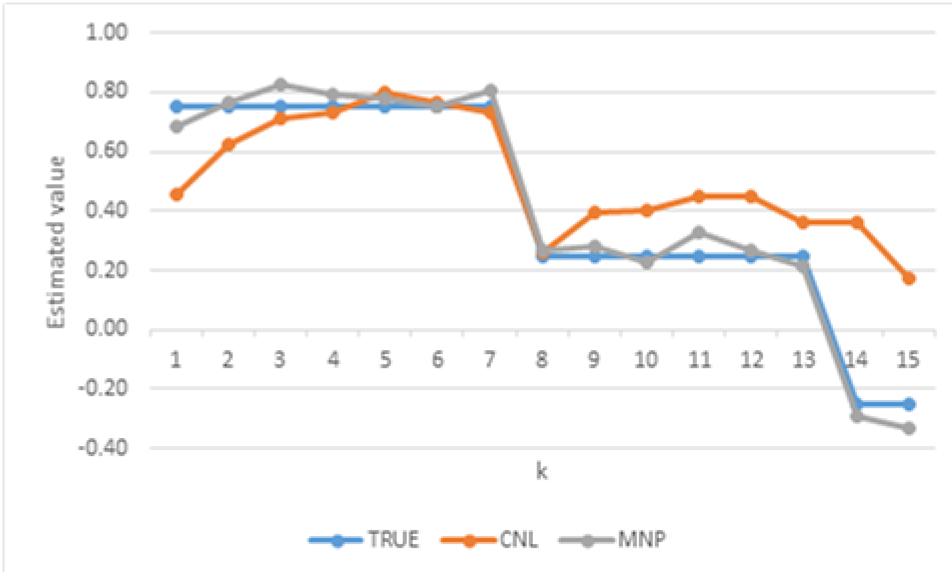

Figure 4. 7 Comparison of $\widetilde{\boldsymbol{\rho}}_{k}^{12}$ between CNL and MNP model

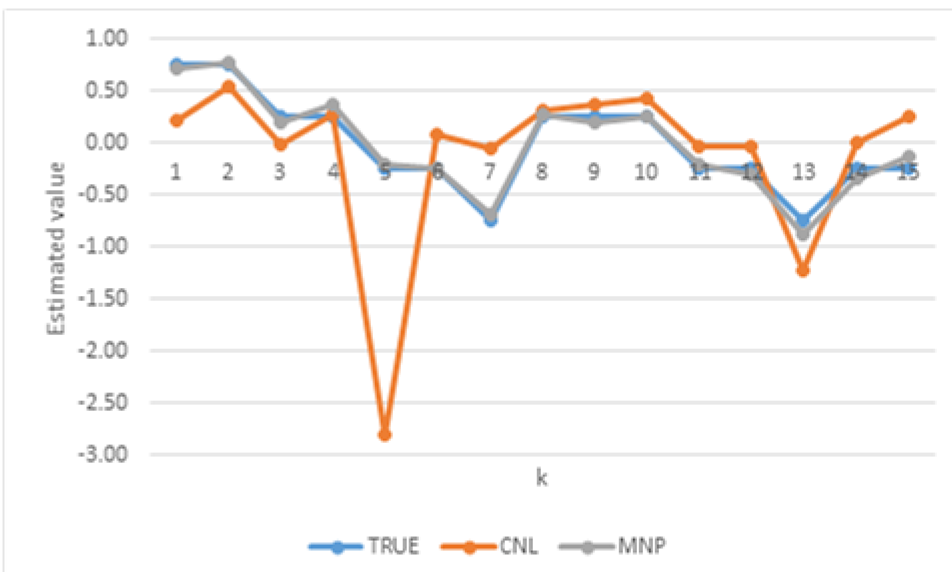

Figure 4. 8 Comparison of $\widetilde{\boldsymbol{\rho}}_{k}^{13}$ between CNL and MNP model

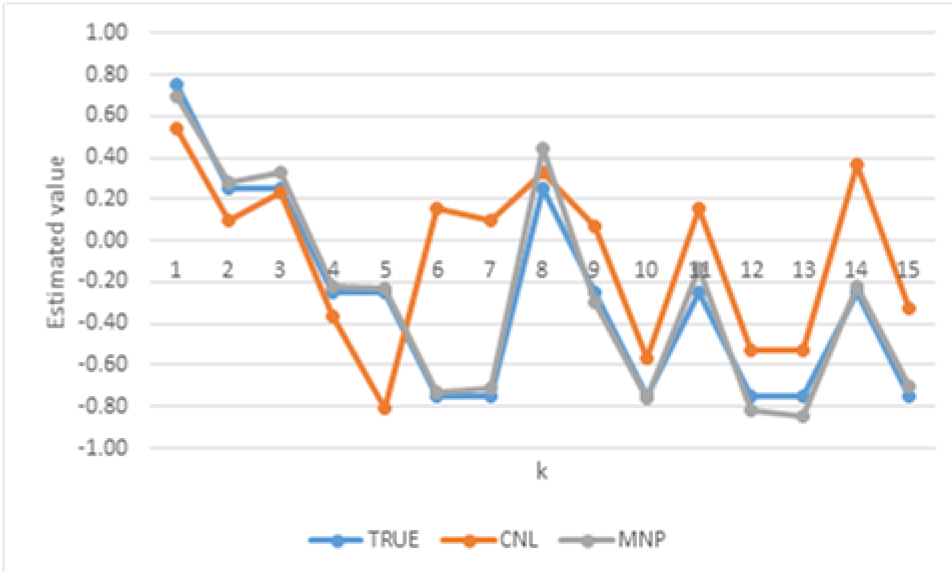

Figure 4. 9 Comparison of $\tilde{\boldsymbol{\rho}}_k^{23}$ between CNL and MNP model

Table 4.4 presents the results for the CNL that converged. Apart for the alternative specific constant, a constant scalar difference can be obtained for most of the results. The results obtained with MNP model are similar to the true value but are not presented in the chapter (can be obtained from the authors by request).

Table 4.4 Estimation results of Cross Nested Logit model

		Alt1			Alt2			Alt3			Alt4	
k	**Asc1**	**β11**	**β21**	**Asc2**	**β21**	**β22**	**Asc3**	**β31**	**β31**	**Asc4**	**β41**	**β41**
True	0.000	0.300	1.200	0.800	0.500	1.500	1.500	0.500	1.800	0.600	0.600	1.200
1	0.000	0.417	1.725	0.663	0.874	2.558	1.963	0.779	2.921	0.421	0.996	1.954
2	0.000	0.395	1.576	0.804	0.735	2.231	1.862	0.720	2.556	0.565	0.854	1.731
3	0.000	0.432	1.639	0.979	0.694	2.103	1.981	0.695	2.484	0.728	0.816	1.676
4	0.000	0.390	1.624	1.093	0.652	1.938	2.034	0.615	2.321	0.955	0.806	1.580
5	0.000	0.337	1.641	1.212	0.608	1.815	2.072	0.605	2.201	1.000	0.739	1.481
6	0.000	0.370	1.662	1.309	0.616	1.845	2.306	0.629	2.224	1.387	0.749	1.428
7	0.000	0.467	1.679	1.346	0.629	1.832	2.500	0.632	2.238	1.559	0.728	1.430
8	0.000	0.372	1.661	1.125	0.650	1.991	2.097	0.644	2.397	0.823	0.800	1.586
9	0.000	0.457	1.617	1.187	0.649	1.878	2.077	0.631	2.220	0.956	0.766	1.480
10	0.000	0.441	1.745	1.226	0.669	2.006	2.148	0.681	2.372	0.955	0.877	1.649
11	0.000	0.449	1.616	1.216	0.594	1.838	2.094	0.588	2.224	0.973	0.740	1.475
12	0.000	0.406	1.747	1.198	0.669	2.026	2.143	0.632	2.387	0.958	0.812	1.649

13	0.000	0.495	1.828	1.250	0.722	2.129	2.271	0.699	2.491	0.992	0.904	1.731
14	0.000	0.420	1.740	1.362	0.639	1.901	2.225	0.616	2.200	1.070	0.763	1.570
15	0.000	0.428	1.690	1.334	0.676	1.940	2.239	0.631	2.257	1.041	0.813	1.564

4.4.4 Model validation

In order to calculate the prediction power of the models under analysis, we calculate the market share on out-of-sample datasets. The estimated coefficients based on 2400 observations are applied to the reminder 600 observations. Table 3.5, Table 3.6, and Table 3.7 report the measure of the errors between observed and predicted market shares. In comparing the errors, we conclude that MNP and NL models provide a better fit when compared to CNL. The results indicate that there is not much difference between the predictions of the MNP and NL models for the three-alternative specification. In fact, the NL model has less apparent difference between the true and estimated shares. The results show that both MNP and NL provide reasonable market shares. The difference between true and estimated shares is smaller in NL compared to MNP. In contrast to the results with NL, the validation of CNL shows large differences between the true and estimated market shares. As mentioned in previous section, CNL cannot converge in most of the runs, which leads to the instability in the predictions.

Table 4.5 Difference in market share predictions - Experiment Ia

		MNP			**NL**	
ρ	**Alt1**	**Alt2**	**Alt3**	**Alt1**	**Alt2**	**Alt3**
-0.95	2.9%	-1.4%	-1.4%	-0.6%	0.4%	0.2%
-0.855	3.4%	-1.6%	-1.8%	0.2%	0.0%	-0.2%
-0.76	2.8%	-1.5%	-1.3%	-0.4%	0.1%	0.3%
-0.665	3.3%	-2.0%	-1.3%	0.4%	-0.7%	0.4%
-0.57	3.2%	-1.4%	-1.8%	0.6%	-0.2%	-0.3%
-0.475	2.6%	-1.0%	-1.7%	0.2%	-0.1%	-0.1%

-0.38	2.6%	-0.9%	-1.7%	0.4%	-0.2%	-0.2%
-0.285	1.9%	0.1%	-2.1%	-0.3%	0.8%	-0.5%
-0.19	2.1%	-0.6%	-1.5%	0.3%	-0.2%	-0.1%
-0.095	1.3%	-0.2%	-1.1%	-0.3%	0.0%	0.3%
0	1.8%	0.0%	-1.8%	0.4%	0.0%	-0.4%
0.095	1.6%	0.3%	-1.9%	0.3%	0.1%	-0.4%
0.19	1.2%	-0.4%	-0.8%	0.2%	-0.9%	0.7%
0.285	0.8%	0.8%	-1.7%	0.0%	0.3%	-0.2%
0.38	0.4%	0.9%	-1.4%	-0.1%	0.2%	-0.1%
0.475	0.3%	1.2%	-1.5%	-0.1%	0.4%	-0.3%
0.57	0.0%	1.6%	-1.5%	-0.3%	0.3%	0.0%
0.665	0.1%	1.0%	-1.1%	0.1%	-0.4%	0.4%
0.76	-0.4%	2.1%	-1.7%	-0.2%	0.7%	-0.6%
0.855	-1.0%	2.5%	-1.5%	-0.3%	0.3%	0.0%
0.95	-0.7%	1.6%	-0.9%	0.0%	-0.4%	0.4%

Table 4.6 Difference in market share predictions - Experiment Ib

	MNP				NL			
ρ	**Alt1**	**Alt2**	**Alt3**	**Alt4**	**Alt1**	**Alt2**	**Alt3**	**Alt4**
-0.950	2.2%	2.8%	-0.1%	-4.8%	-0.1%	0.2%	0.3%	-0.5%
-0.855	2.0%	1.7%	-0.4%	-3.4%	-0.1%	-0.4%	-0.1%	0.7%
-0.760	2.3%	1.7%	-1.0%	-2.9%	0.3%	-0.4%	-0.4%	0.4%
-0.665	1.8%	1.8%	-1.4%	-2.2%	-0.2%	0.1%	-0.6%	0.7%
-0.570	1.7%	1.6%	-1.5%	-1.8%	-0.1%	0.1%	-0.6%	0.6%
-0.475	1.2%	1.2%	-0.2%	-2.2%	-0.5%	0.1%	0.6%	-0.2%
-0.380	1.3%	1.5%	-2.0%	-0.9%	-0.3%	0.4%	-0.6%	0.5%
-0.285	1.6%	1.2%	-1.1%	-1.7%	0.1%	0.3%	0.0%	-0.5%
-0.190	1.4%	0.2%	-1.2%	-0.4%	-0.5%	0.1%	0.3%	0.0%
-0.095	2.1%	-0.4%	-1.7%	0.0%	0.5%	-0.5%	-0.1%	0.1%
0.000	2.2%	-0.4%	-2.6%	0.8%	0.3%	-0.1%	-0.8%	0.5%
0.095	1.2%	0.0%	-1.3%	0.2%	-0.3%	0.0%	0.6%	-0.2%
0.190	1.2%	0.4%	-1.8%	0.3%	0.1%	0.3%	-0.1%	-0.3%
0.285	0.8%	0.8%	-2.6%	1.0%	0.0%	0.7%	-0.3%	-0.4%
0.380	-0.1%	1.4%	-2.5%	1.2%	-0.1%	0.6%	-0.5%	0.0%
0.475	-0.4%	0.7%	-1.7%	1.3%	-0.2%	0.0%	0.4%	-0.2%
0.570	-0.3%	1.0%	-2.7%	1.9%	0.1%	0.1%	-0.2%	0.1%
0.665	-1.3%	0.4%	-2.3%	3.3%	-0.5%	-0.4%	0.1%	0.9%
0.760	-0.8%	0.6%	-2.4%	2.6%	0.4%	-0.3%	-0.3%	0.3%
0.855	-1.8%	0.9%	-2.4%	3.3%	-0.1%	-0.2%	0.1%	0.2%
0.950	-1.5%	1.2%	-2.9%	3.2%	0.2%	0.0%	-0.3%	0.1%

Table 4.7 Difference in market share predictions - Experiment II

	MNP				CNL			
k	**Alt1**	**Alt2**	**Alt3**	**Alt4**	**Alt1**	**Alt2**	**Alt3**	**Alt4**
1	0.5%	0.3%	-3.0%	2.2%	-8.8%	7.5%	-10.5%	11.7%
2	-1.5%	1.9%	-3.1%	2.7%	-7.7%	5.2%	-4.7%	7.3%
3	-0.1%	3.0%	-1.4%	-1.6%	-5.8%	6.5%	-9.7%	9.1%
4	0.8%	1.9%	-1.8%	-0.9%	-9.8%	17.5%	-26.0%	18.3%
5	0.6%	2.4%	-0.3%	-2.7%	-4.9%	6.7%	-1.8%	0.1%
6	2.7%	2.8%	-2.7%	-2.8%	-6.0%	12.9%	-8.6%	1.7%
7	2.0%	1.6%	-2.0%	-1.6%	-4.1%	9.1%	-0.1%	-4.9%
8	1.6%	-0.7%	-1.1%	0.2%	-8.1%	10.0%	-6.4%	4.5%
9	1.7%	2.6%	-3.4%	-0.9%	-6.5%	6.1%	-9.6%	9.9%
10	1.9%	1.2%	-2.3%	-0.8%	-4.7%	11.7%	-7.8%	0.9%
11	0.5%	1.8%	-1.7%	-0.6%	-4.9%	5.0%	-6.0%	5.9%
12	0.9%	2.4%	-1.1%	-2.2%	-3.6%	9.9%	-5.3%	-1.0%
13	2.3%	0.1%	-0.9%	-1.6%	-2.9%	10.6%	3.9%	-11.6%
14	2.7%	-0.9%	-0.4%	-1.3%	-6.1%	8.9%	-6.1%	3.3%
15	2.4%	2.3%	-2.8%	-1.9%	-6.2%	16.4%	-11.7%	1.5%

4.5 Evidence with Real Data

So far, our investigation has been based on synthetic data designed specifically for "known" correlation structures. We turn now our attention to a real case study where the primary data source is extracted from the 2013 American Time Use Survey (ATUS). The ATUS survey has been designed and collected by the Bureau of Labor Statistics on a yearly basis starting from 2003. ATUS questionnaire asks respondents to report their time use together with other information on daily activity episodes including the start and end time of participation, type and location of recorded activity. Socio-demographic information can also be obtained from the survey.

In this study, we consider observations for weekdays from ATUS 2013; 5595 observations are included in the final dataset used for model estimation. Household and individual characteristics, land-use variables are the main variables extracted from the original dataset. The dependent variable of our discrete choice model is the involvement in leisure activities. Six activity episodes have been selected and categorized according to their locations and types (including computer use for leisure):

- No leisure activities (NO);
- Pure in-home computer use activities (LPC): only choose computer use for leisure activity;
- Pure in-home other leisure activities (LH): only choose in-home leisure activities other than computer use;
- Pure out-of-home leisure activities (LOH): only choose out-of-home leisure activities;
- Multiple in-home leisure and computer use activities (LH&LPC): choose in-home computer use and other in-home leisure activities;
- Multiple in-home and out-home leisure activities (LH&LOH): choose in-home leisure activities without computer use and out-of-home leisure activities.

In addition to the models tested in the synthetic data experiments: MNP, NL, and CNL, there is added value to evaluate the performance of the Mixed Logit Model (MXL) (Cardell and Dunbar, 1980; Train, 2009). The MXL model is a highly flexible model that can approximate any random utility model (McFadden & Train, 2000) and has been widely applied in research and practice. In this case, the MXL model is applied to investigate the negativity of correlations among choices. In the MXL model the utility is specified as

$$U_{mj} = \beta' x_{mj} + \kappa_m' z_m + \varepsilon_j \tag{4.15}$$

where x_{mj} and z_{mj} are vectors of observed variables relating to alternative j, β is a vector of fixed coefficients, κ is a vector of random terms with zero mean, and ε_j is i.i.d extreme value. The terms in z_m define the stochastic portion of utility. The unobserved portion of utility is $\eta_{mj} = \kappa_m' z_m + \varepsilon_j$, which can be correlated over alternatives depending on the specification of z_m. The covariance between any two alternatives in nest k is specified as

$$Cov(U_{mi}, U_{mj}) = E(\kappa_m' z_m + \varepsilon_i)(\kappa_m' z_m + \varepsilon_j) = \sigma_m \tag{4.16}$$

A MNP with variance-covariance matrix is also estimated, using in house software coded in R language; as noted several times in this chapter MNP is able to correctly recover all types of correlation, including negative correlation if any.

In the NL model, LPC, LH, and LPC&LH are specified in nest B, which contain all home related leisure activities, while LOH and LH&LOH are in nest C, where all the alternatives have an out-of-home leisure episode. It is also conceivable that such correlations also exist between the LH and LH&LOH alternatives, given that they have the common aspect of involving in-home leisure activity. To test for the presence of such correlation, CNL and MXL model were fitted to the data, allowing LH to be shared by two nests. Figure 4.10 and Figure 4.11 present the model structures described.

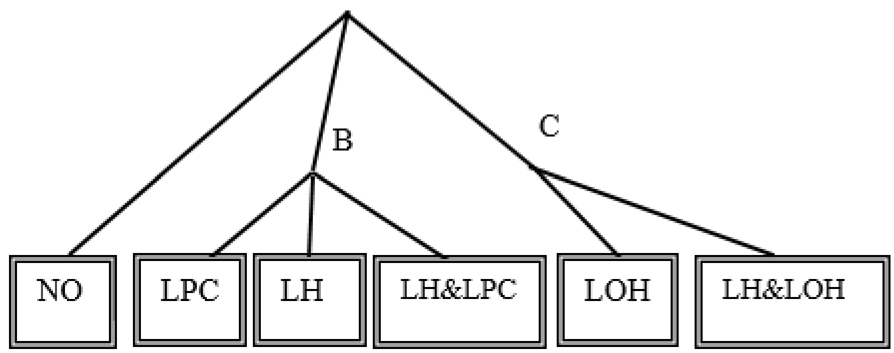

Figure 4. 10 NL model structure in real case study

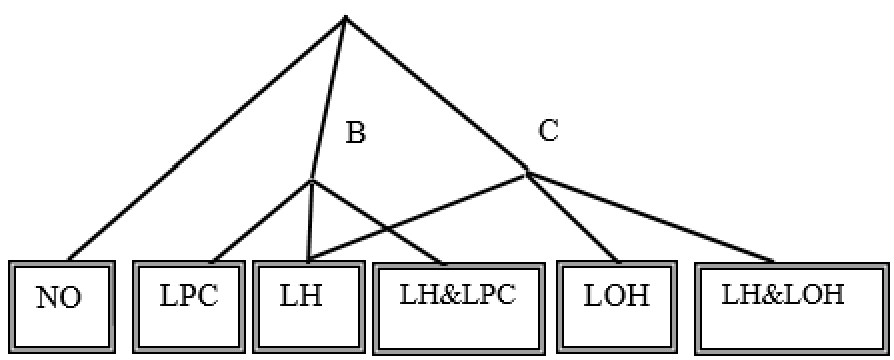

Figure 4. 11 CNL and MXL model structure in real case study

MXL, NL, and CNL models were estimated using BIOGEME. The MXL correlation matrix attests that there exist negative correlations between alternatives in Nest B and C. The same applies to correlation terms estimated with NL and CNL (shown in Table 4.8). Unfortunately, these correlation matrices cannot be compared directly with the one obtained by using Probit where notably the correlations are across differences in error terms with respect to the first alternative; for comparison purpose, the covariance matrices are normalized using (Eqn.4.14).

Table 4.8Covariance of difference matrix in real case

$$\Omega_{MNP} = \begin{pmatrix} 2.00 & 0.84 & 1.14 & 1.16 & 1.02 \\ 0.84 & 2.78 & 1.26 & 1.89 & 1.86 \\ 1.14 & 1.26 & 3.95 & 2.76 & 2.83 \\ 1.16 & 1.89 & 2.76 & 5.21 & 4.36 \\ 1.02 & 1.86 & 2.83 & 4.36 & 6.12 \end{pmatrix} \quad \Omega_{NL} = \begin{pmatrix} 3.29 & -1.72 & 1.64 & -1.72 & 1.64 \\ -1.72 & 3.29 & 1.64 & -1.72 & 1.64 \\ 1.64 & 1.64 & 3.29 & 1.64 & 0.76 \\ -1.72 & -1.72 & 1.64 & 3.29 & 1.64 \\ 1.64 & 1.64 & 0.76 & 1.64 & 3.29 \end{pmatrix}$$

$$\Omega_{CNL} = \begin{pmatrix} 3.29 & -2.15 & 1.64 & -8.40 & 1.64 \\ -2.15 & 3.29 & 1.39 & -2.15 & 1.39 \\ 1.64 & 1.39 & 3.29 & 1.64 & 1.37 \\ -8.40 & -2.15 & 1.64 & 3.29 & 1.64 \\ 1.64 & 1.39 & 1.37 & 1.64 & 3.29 \end{pmatrix} \quad \Omega_{MXL} = \begin{pmatrix} 3.01 & 1.36 & 1.64 & 1.36 & 1.64 \\ 1.36 & 3.23 & 1.87 & 1.36 & 1.87 \\ 1.64 & 1.87 & 3.51 & 1.64 & 1.87 \\ 1.36 & 1.36 & 1.64 & 3.01 & 1.64 \\ 1.64 & 1.87 & 1.87 & 1.64 & 3.51 \end{pmatrix}$$

Correlation matrix of NL, CNL, MXL

$$Cor_{NL} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -2.05 & 0 & -2.05 & 0 \\ 0 & -2.05 & 1 & 0 & -2.05 & 0 \\ 0 & 0 & 0 & 1 & 0 & -0.54 \\ 0 & -2.05 & -2.05 & 0 & 1 & 0 \\ 0 & 0 & 0 & -0.54 & 0 & 1 \end{pmatrix}$$

$$Cor_{CNL} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -2.31 & 0 & -6.11 & 0 \\ 0 & -2.31 & 1 & -0.15 & -2.31 & -0.15 \\ 0 & 0 & -0.15 & 1 & 0 & -0.16 \\ 0 & -6.11 & -2.31 & 0 & 1 & 0 \\ 0 & 0 & -0.15 & -0.16 & 0 & 1 \end{pmatrix}$$

$$Cor_{MXL} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -0.19 & 0 & -0.21 & 0 \\ 0 & -0.19 & 1 & 0.13 & -0.19 & 0.13 \\ 0 & 0 & 0.13 & 1 & 0 & 0.12 \\ 0 & -0.21 & -0.19 & 0 & 1 & 0 \\ 0 & 0 & 0.13 & 0.12 & 0 & 1 \end{pmatrix}$$

Table 4.9 shows the estimation results obtained by applying the models, along with degree of independence for Nest B and Nest C (μ_B, μ_C) of both models, factors α_{3B}, α_{3C} of LH for CNL, and covariance σ_B, σ_C for MXL. Estimation results are stable and variables maintain their sign and their significance across the three model specifications, with just few exceptions. Surprisingly, the Probit and MXL model present worse fit, while NL and CNL produce almost the same value of the final log-likelihood. The nested coefficients μ_B, μ_C are both significant, while the two additional parameters of CNL are not significant.

Table 4.9 Estimation results in real case (standard errors in parentheses)

		Coefficient			
	Name	**MNP**	**NL**	**CNL**	**MXL**
LPC	Constant	-1.091* (0.189)	-4.810* (2.22)	-7.030 (4.51)	-1.800* (0.24)
	Higher than BA Degree (Dummy)	0.714 (0.3)	1.010 (0.65)	1.300 (0.94)	0.538* (0.29)
	Full-time working status (Dummy)	-0.109 (0.245)	-0.254 (0.49)	0.252 (1.03)	-0.424* (0.25)
LH	Constant	2.520* (0.092)	2.980* (0.12)	2.960* (0.14)	3.040* (0.31)
	No. of children	-0.030* (0.037)	-0.142* (0.05)	-0.140* (0.05)	-0.147* (0.04)
	Full-time working status (Dummy)	-0.758* (0.073)	-0.674* (0.09)	-0.702* (0.11)	-0.744* (0.11)

	Age. Senior	-0.362*	-0.673*	-0.708*	-0.476*
	(Dummy)	(0.089)	(0.13)	(0.17)	(0.09)
	Age. Adult	-0.372*	-0.863*	-0.888*	-0.817*
	(Dummy)	(0.118)	(0.16)	(0.19)	(0.13)
	Age. Young	-0.646*	-1.090*	-1.170*	-0.847*
	(Dummy)	(0.153)	(0.22)	(0.28)	(0.16)
	Age. Teen	-0.579*	-1.040*	-1.130*	-0.904*
	(Dummy)	(0.162)	(0.24)	(0.30)	(0.17)
	Higher than BA Degree	-0.156*	-0.434*	-0.422*	-0.335*
	(Dummy)	(0.118)	(0.15)	(0.15)	(0.12)
LOH	Constant	-1.084*	-1.320*	-1.210*	-0.955*
		(0.153)	(0.61)	(0.56)	(0.19)
	No. of children	-0.209*	-0.320*	-0.313*	-0.278*
		(0.162)	(0.12)	(0.11)	(0.09)
	Single	0.197	0.273	0.295	0.317*
	(Dummy)	(0.186)	(0.25)	(0.23)	(0.19)
	Age. Teen	0.338*	1.210*	1.150*	0.957*
	(Dummy)	(0.309)	(0.43)	(0.39)	(0.31)
	Age. Young	0.632*	0.768*	0.737*	0.870*
	(Dummy)	(0.268)	(0.38)	(0.34)	(0.26)
LPC&LH	Constant	0.420*	-0.641	-0.857	0.837*
		(0.1)	(1.03)	(1.5)	(0.19)
	No. of children	-0.098*	-0.287*	-0.310*	-0.247*
		(0.056)	(0.10)	(0.11)	(0.06)
	Race. Black	-0.775*	-1.250*	-1.480*	-0.667*
	(Dummy)	(0.161)	(0.49)	(0.60)	(0.16)
	Full-time working status	-1.341*	-1.510*	-1.660*	-1.29*
	(Dummy)	(0.115)	(0.25)	(0.34)	(0.14)
	Higher than BA Degree	-0.004	0.605*	0.615*	0.307*
	(Dummy)	(0.167)	(0.35)	(0.37)	(0.17)
	BA Degree	0.557*	0.820*	0.894*	0.548*
	(Dummy)	(0.114)	(0.28)	(0.30)	(0.12)
	Age. Adult	-0.190*	-0.545*	-0.518*	-0.370*
	(Dummy)	(0.139)	(0.25)	(0.27)	(0.14)
LOH&LH	Constant	0.378*	0.603*	0.531*	0.704*
		(0.103)	(0.14)	(0.29)	(0.17)
	No. of children	-0.106*	-0.161*	-0.168*	-0.169*
		(0.054)	(0.07)	(0.07)	(0.05)
	Gender. Female	-0.475*	-0.371*	-0.390*	-0.278*
	(Dummy)	(0.086)	(0.10)	(0.13)	(0.09)
	Single	0.359*	0.357*	0.370*	0.313*
		(0.105)	(0.13)	(0.13)	(0.11)

Higher than BA Degree (Dummy)	-0.581* (0.176)	-0.857* (0.24)	-0.808* (0.22)	-0.819* (0.18)
BA Degree (Dummy)	0.025* (0.111)	-0.242* (0.14)	-0.217 (0.14)	-0.275* (0.11)
Age. Adult (Dummy)	-0.603* (0.125)	-0.427* (0.16)	-0.414* (0.15)	-0.495* (0.13)
μ_B		0.573*	0.375*	
μ_C		0.807*	0.927	
α_{3B}			0.143*	
α_{3C}			0.857*	
σ_B				-0.281
σ_C				0.233
Final Log-likelihood	-6420.89	-4975.36	-4975.02	-6213.43

*significant, p-value < 0.1

Consistently with what was conducted for the simulated datasets, we tested the ability of the models in Table 4.10 to reproduce market share in out-of-samples. We re-estimated the model on about 80% of the observations and we applied the model to the remaining observations. The results show that although MXL, MNP, and NL models have a good performance, NL produces better results when compared to MNP. CNL has the most biased results, mainly caused by the failure in reproducing the market share for the alternative LPC&LH.

Table 4.10 Validation results in real case study (predicted market shares)

Alternative	**Observed**	**Predicted value**				**Difference**			
		MNP	**NL**	**CNL**	**MXL**	**MNP**	**NL**	**CNL**	**MXL**
NO	9.20%	10.81%	10.86%	12.49%	9.48%	1.61%	1.66%	3.28%	0.27%
LPC	0.89%	3.90%	0.00%	0.00%	1.35%	3.00%	-0.89%	-0.89%	0.45%
LH	64.70%	63.47%	62.78%	70.51%	65.32%	-1.23%	-1.92%	5.81%	0.62%
LOH	3.84%	3.46%	2.78%	5.51%	3.42%	-0.38%	-1.06%	1.67%	-0.42%
LPC&LH	9.38%	8.07%	12.10%	0.41%	9.12%	-1.31%	2.72%	-8.98%	-0.26%
LOH&LH	11.97%	10.29%	11.48%	11.09%	11.31%	-1.68%	-0.50%	-0.89%	-0.67%

We finally analyse model elasticity and particularly calculate the effects on LH share caused when increasing of one unit the number of child in the household. Table 4.11 reports the changes in the aggregate share of LH activity (P_{LH}) over the initial value (P_{LH}^0):

$$\Delta P_{LH} = \frac{P_{LH} - P_{LH}^0}{P_{LH}^0} \tag{4.17}$$

where P_{LH}^0 and P_{LH} are, respectively, the aggregate probabilities of choosing activity LH before and after the variable number of children in the household has been modified. All probabilities are calculated by using sample enumeration (Munizaga et al., 2000).

Table 4.11 Policy analysis

Alternative	**MNP**	**NL**	**CNL**	**MXL**
NO	3.49%	18.41%	11.22%	7.67%
LPC	3.70%	37.86%	22.59%	8.80%
LH	0.72%	4.67%	-0.04%	0.01%
LOH	-10.84%	-12.47%	-17.50%	-6.74%
LPC&LH	-2.17%	-38.04%	-8.67%	-5.03%
LOH&LH	-2.37%	0.89%	-4.61%	-1.10%

It appears that MXL and MNP models produce similar results, while NL model has different results than the other three. The interpretation is quite straightforward. The NL and CNL models could produce biased modal shifts, when failing to account for correlation across observations and eventually different policy analysis results.

4.6 Summary

In this chapter we put forward the idea of a possible bias when trying to estimate GEV type choice models in the presence of negative correlations. GEV choice models like Nested Logit and Cross Nested Logit Model have been widely used in the past years. However, modelers hardly ever know in advance the correlation structure of their choice alternatives and tend to forget the fact that negative error correlation might bias their results. In these cases, MNP or MXL, that can overcome the non-negative correlation limitation, should be adopted; however, the simulation assisted estimation is often lengthy and difficult.

To understand the performance of GEV models when negative correlations appear between choices, three experiments are carried out for two of the most common GEV models- Nested Logit and Cross Nested Logit (Paired Combinatorial). The first two experiments use synthetic data that recreate artificial sets of different correlations in the choice vectors. Based on these datasets we estimate the MNP and GEV models, and we compare their estimates to the true values. An experiment based on the 2013 American Time Use Survey data was considered as a real case study where true values are unknown. However, estimated results obtained using MXL model indicate the negative correlations exist between activity choices. The three models were also validated by calculating market shares on out-of-sample observations.

The results with synthetic data (Experiment I and II) reveal that the GEV correlation estimates are biased in the presence of negative correlation, while the MNP estimates of the correlations are practically identical to the true values. In the case of NL, biased estimates of negative correlation have the same patterns for both simple three-alternative case and complex four-alternative case. The results are consistent with the key assumption of GEV model. In the case of CNL, the results from both correlation estimation and validation reveal that the PCL specification fails to estimate the true correlation even under the non-negative conditions. Evidently, more research is required to investigate the CNL model with PCL specification and its failure to achieve convergence.

The results obtained from the real case study attest that MNP, and GEV models produce similar estimates. Negative correlations have been estimated with NL and CNL models; direct comparison with MNP correlations is impossible given that the normalization of pProbit imposes to work with differences in error terms. While the model fit of NL and CNL is much better than the one obtained with MNP and MXL, NL and MXL models produce better aggregate choice probabilities when applied to an out-of-sample dataset for validation and when compared with MNP and CNL. Nevertheless, MNP and MXL do better than NL in sensitivity analysis when marginal changes are considered for policy analysis as they properly account for the (negative) correlation across alternatives.

Recently, researchers are working to make it easier to use flexible modelling specifications like MNP by providing more efficient estimation techniques that reduces the computational burden of simulations. This research shows that GEV models, which are notably homoscedastic, could only deal with limited correlation pattern and are not suited for negative correlations. Mixed Logit, which is not limited by the assumptions imposed by GEV is less restrictive. It is suggested that when lacking information on the data structure, more flexible model specifications should be used. However, these models still suffer from a high level of sophistication and expert knowledge is required to verify model identification and correct estimation. Probit and Mixed Logit models have no closed form, estimation is based on simulation and random drawing procedures, and computation time is significantly larger compared to straightforward GEV models. However, improvements made in both hardware and software are reducing this limitation and make flexible models more attractive to practitioners.

The counterintuitive evidence we provided in this chapter suggests that more research is needed in understanding the statistical and mathematical properties of discrete choice models. The important lesson for modellers and practitioners is to test many various model specifications with the same dataset including both estimation, and not less important, validation of the model coefficients as well as sensitivity analysis to key parameters.

Chapter 5: Discrete Continuous Model on Linking activity involvement to time use decisions

5.1 Problem Description

Social media platforms and online communities continue global expansion in recent years. In 2016, with a global population of 7.4 billion, 3.419 billion are internet users, of which 2.3 billion use social media (Wearesocial, 2016). Overall, it is estimated that two third of online adults are using social media platforms. The high penetration rate of social media is changing how individuals communicate and interact. By analyzing social media and its use, researchers are trying to understand people's thinking, communication patterns, health, beliefs, prejudices, group behaviours, which is relevant in social science and related disciplines. At the same time, the growing use of social media is also expected to modify travel patterns both indirectly, by changing activity needs and time spent at home or out-of-home, and directly, by modifying the perception and the utility of the time spent traveling, during which the use of social media is becoming ubiquitous. It is therefore also important to transportation researchers understanding the influences of social media on the time allocated to activities and ultimately on travel behaviour.

Most previous studies in transportation focused on the influence of ICT usage on working activity and commute trips (Wang and Law, 2007; Ben-Elia et al., 2014). Other studies turned their attention to the effects of internet usage on individual's attitudes towards time usage and involvement in other physical activities such as discretionary trips (Ferrell, 2005; Veenhof, 2006; Farag et al., 2006; Carrasco and Miller, 2009). However, limited studies have empirically investigated social media involvement, its effects on leisure activity participation, and the relative time use.

In this part, we propose an integrated econometric framework that accounts for the effects that internet usage for leisure and relaxing, which contains a major component of social media involvement, has on activity-travel patterns, including social and commute trips. The joint model proposed captures the potential correlation across activity involvement choices, the location where this activity takes place and time usage decisions associated to each chosen activity. A number of studies have shown that time-space constraints play an important role in shaping people's activity patterns (Pendyala, 2002; Yamamoto et al., 2004; Kitamura et al., 2006), and that time use affects individual's daily schedule (Bhat and Koppelman, 1999). Neglecting the correlation among spatial and temporal decisions may result in the inability of the modelling framework to accurately capture and reflect individual activity and time use patterns in our increasingly digitized world. It is the purpose of the study to identify the appropriate data for this problem, to formulate the model that account for both discrete (leisure participation and location) and continuous decisions (time spent on social media) and to quantify the impacts that the involvement on social media has on travel behaviour.

5.2 Data Descriptions

The primary data source used in this analysis is extracted from the 2013 American Time Use Survey (ATUS) (Bureau of Labor Statistics, 2014). The ATUS is designed and collected by the U.S. Bureau of Labor Statistics and contains detailed information on time use for each activity on which respondents have been involved the day before the interview. Activity related attributes include the start and end time of participation, activity type, and activity location; individual and household socioeconomic characteristics are also available in the dataset. Both in-home and out of home activities are reported, which makes ATUS particularly attractive for time use analysis and modelling.

In this study we are interested in leisure activity involvement, in the location where those activities take place and the time spent for leisure. We distinguish between in-home and out of home leisure activities and between generic leisure activities and those involving the use of the computer. In particular, we refer to the ATUS category "*Computer use for leisure*"; this variable explicitly excludes games, listening to music, watching videos, e-mails, computer use for work and volunteer activities, which are included in different activity categories. Therefore, we argue that this activity category is mainly time spent online to use social media; a comparative study based on ATUS and a survey conducted by Nielsen supports our claim and concludes that "*the top leisure uses included in the ATUS variable are social networks, portals and search*". (Greenstein and Tucker, 2015).

By combining "in home" and "out of home" with "use of computers" and "absence of use of computers" and also considering one versus multiple leisure activities, the resulting set of discrete choices over leisure activity types includes the following six alternatives:

- No leisure activities (NL) on the day of the survey;
- Pure in-home leisure activities that involve the use of the computer (LPC);
- Pure in-home other (than computer use) leisure activities (LH);
- Pure out-of-home leisure activities (LOH);
- Multiple in-home leisure activities, of which some require the use of the computer (LH&LPC);
- Multiple in-home and out-of-home leisure activities in which the computer is not in use (LH&LOH).

Table 5.1 Distribution of Leisure Activity

Category	**Obs. (Weekdays)**	**Obs. (Weekends)**
No leisure activity (NL)	543	519
Pure in-home computer use activities (LPC)	76	80
Pure in-home other leisure activities (LH)	3620	3773
Pure out-of-home leisure activities (LOH)	194	261
Multiple in-home leisure and computer use activities (LH&LPC)	523	502
Multiple in-home and out-of-home leisure activities (LH&LOH)	638	477

Each survey respondent, and the corresponding observation in the dataset can then be classified into one of the above six kinds of activity sequences. Table 4.1 provides the breakdown among activity sequences in the sample; a total of 5,612 observations are available for weekends, while 5,594 observations are available for weekdays. Household characteristics, land-use variables and time use information for each household, are the main variables extracted from the original dataset. Table 4.2 lists the basic statistics relative to the 2013 ATUS sample. We can observe that individuals with no leisure activity have the highest travel time to work, travel to social and entertainment activities and in average have more children. Individuals who use the computer for leisure activities have high income and spend also time on art and entertainment related activities. These trends are similar for weekdays and weekends. In average about 1.5 hours per day are spent on computer for leisure during weekdays and about 2.3 hours per day during weekends among observations who choose LPC. While average time spent on all leisure activity is about 3.1 hours per day during weekdays and about 3.8 hours per day during weekends among all observations.

Table 5.2 Descriptive statistics of attributes

Variables		By activity types					
		NL	**LPC**	**LH**	**LOH**	**LH& LPC**	**LH& LOH**
Weekday:							
Gender (female = 1; otherwise = 0)		0.558	0.579	0.572	0.557	0.579	0.464
Metropolitan status (metropolitan = 1; otherwise = 0)		0.843	0.895	0.828	0.784	0.839	0.850
Working status (full time = 1; otherwise = 0)		0.628	0.553	0.404	0.603	0.333	0.600
No. of people in household		2.396	2.289	2.185	2.134	2.191	2.136
Age (years)		42.0	41.1	50.5	41.6	49.3	50.0
Household income ($)		77025	85903	61418	63961	72588	59232
No. of children in Household		1.1	1.1	0.8	0.7	0.7	0.7
Household type	1.Married,	56.2%	60.5%	51.8%	43.8%	55.8%	47.0%
	2. Unmarried,	20.6%	13.2%	17.2%	22.7%	14.0%	16.5%
	3.Single,	23.2%	26.3%	31.0%	33.0%	30.0%	36.6%
	4. Group	<1.00%	<1.00%	<1.00%	<1.00%	<1.00%	<1.00%
Travel time related to working (hrs.)		0.496	0.431	0.297	0.378	0.198	0.42
Travel time related to socializing and communicating (hrs.)		0.032	0.016	<0.010	0.012	<0.010	<0.010

Travel time related to arts and entertainment (hrs.)	0.104	0.103	0.069	0.161	0.057	0.088
Time spent on in-home leisure activity (hrs.)	NA	NA	4.204	NA	3.811	3.079
Time spent on in-home computer use for leisure activity (hrs.)	NA	1.500	NA	NA	1.475	NA
Time spent on out-of-home leisure activity (hrs.)	NA	NA	NA	1.842	NA	1.066

Weekends:							
Gender (female = 1; otherwise = 0)		0.636	0.625	0.542	0.563	0.540	0.503
Metropolitan status (metropolitan = 1; otherwise = 0)		0.834	0.825	0.818	0.820	0.849	0.832
Working status (full time = 1; otherwise = 0)		0.597	0.463	0.433	0.494	0.394	0.463
No. of people in household		2.370	2.288	2.203	2.241	2.191	2.140
Age (years)		44.3	41.0	50.0	44.4	48.5	45.0
Household income ($)		74030	81025	61709	65120	68533	57727
No. of children in Household		1.1	0.8	0.8	0.9	0.7	0.7
Household type	1.Married	53.0%	53.8%	52.3%	44.8%	54.7%	41.5%
	2. Unmarried	21.8%	18.8%	16.9%	22.2%	13.1%	22.9%
	3.Single	25.2%	27.5%	30.7%	33.0%	32.0%	35.6%
	4. Group	<1.00%	<1.00%	<1.00%	<1.00%	<1.00%	<1.00%
Travel time related to working (hrs.)		0.105	0.061	0.099	0.027	0.139	0.080
Travel time related to socializing and communicating (hrs.)		0.212	0.138	0.114	0.148	0.082	0.148
Time spent on in-home leisure activity (hrs.)		NA	NA	4.948	NA	4.511	3.334
Time spent on in-home computer use for leisure activity (hrs.)		NA	2.333	NA	NA	1.546	NA
Time spent on out-of-home leisure activity (hrs.)		NA	NA	NA	3.453	NA	1.929

5.3 Modelling Framework

The econometric model system proposed captures the joint decisions of participation in leisure activity, where this activity takes place (in-home vs. out-of-home) if the leisure activity involves the use of the computer and the time spent on each of the leisure activity. The problem involves both discrete dependent variables (activity type) and continuous dependent variables (time use). A discrete-continuous model framework is adopted in the study to jointly estimate the leisure activity choice and the time spent on each leisure activity choice. An in-house software coded in R language is used to estimate the integrated model with variance-covariance matrix.

5.3.1 The Activity Choice Sub-model

Discrete choice analysis is adopted to model the choice of activity sequences. The discrete choice model forecasts the outcome of a categorical dependent variable Y_{DIS}. The six types of leisure activity sequences that were introduced in Chapter 3 thus constitute the discrete endogenous variable in the modelling framework. To each activity sequence i, i = 1 … 6, we associate a utility:

$$U_i = X_i^T \beta_i + \varepsilon_i, \tag{5.1}$$

where X_i are the socio-demographic attributes and activity related variables, β_i are the associated parameters to be estimated and ε_i are the error terms.

The decision maker is assumed to be rational and to choose the alternative with the highest utility. A multivariate Probit model is adopted for the discrete problem, and therefore the error terms follow a multivariate normal distribution with full, unrestricted covariance matrix. The Probit model is normalized to take into account the fact that the level and scale of the utility is irrelevant (Train, 2009).

The probability of choosing a given leisure activity sequence *i* can also be expressed in the way of difference:

$$\mathrm{P}(\mathrm{Y}_{DIS} = \mathrm{i}) = \int B(\widetilde{\mathrm{V}}_{i-j} + \tilde{\varepsilon}_{i-j} > 0, \forall j \neq i)\phi(\tilde{\varepsilon})d\tilde{\varepsilon} \tag{5.2}$$

where $B()$ is a Boolean indicator of whether the statement in parentheses holds is, $\phi(\tilde{\varepsilon})$ is the density of the error term in difference formation (i.e. $\tilde{\varepsilon}_{i-j} = \varepsilon_i - \varepsilon_j$), $\widetilde{\mathrm{V}}_{i-j} = X_i^T\beta_i - X_j^T\beta_j$.

5.3.2 The Time Usage Sub-model

Regressions are used to estimate the time spent on leisure activities, which are classified into three groups according to the location where they take place and, for in-home activities, the use of computers. The model formulation therefore includes the following three continuous variables as dependent variables: (a) time spent on out-of-home leisure activity (LOH), (b) time spent on in-home leisure activity without computer use (LH), and (c) time spent on PC for leisure purpose (LPC). For example, if individual chooses multiple leisure activities (e.g. LH&LPC), two regressions are used to estimate the time usage on (b) LH and (c) LPC following Eqn. (5.5). The time spent on single leisure activity $\mathrm{s} \in \{\mathrm{LPC}, \mathrm{LH}, \mathrm{LOH}\}$, $Y_{REG,s}$, can be expressed as a linear combination of a vector of predictors $X_{REG,s}$ and error term $\epsilon_{REG,s}$:

$$Y_{REG,s} = X_{REG,s}^T\beta_{REG,s} + \varepsilon_{REG,s}, \varepsilon_{REG,s} \sim N\left(0, \sigma_{REG,s}^2\right). \tag{5.3}$$

Usually, regressions are solved by the Ordinary Least Squares (OLS) estimator (Weisberg, 2005). Alternatively, the problem can also be expressed in the form of a likelihood function to be maximized. The two methods are equivalent under the assumption that errors are normally distributed (McCulloch and Neuhaus, 2001, p.117).

For multi-leisure-activity participation, it can then be expressed by the generic equation:

$$Y_{REG,\eta} = X_{REG,\eta}^{T}\beta_{REG,\eta} + \varepsilon_{REG,\eta}, \varepsilon_{REG,\eta} \sim MVN\left(0, \Sigma_{REG,\eta}\right), \tag{5.4}$$

where $Y_{REG,\eta}$ is a set of observed time usages of given leisure activities subset $\eta \subseteq \{\mathrm{LPC}, \mathrm{LH}, \mathrm{LOH}\}$. The likelihood of observing $Y_{REG,\eta}$ is given by the normal density function:

$$P(Y_{REG,\eta}) = \phi(err|\mu, \sigma^2), \tag{5.5}$$

where *err*= $Y_{REG,\eta} - \hat{Y}_{REG,\eta}$. Correspondingly, the time usage of an individual on a single leisure activity s follows a normal distribution with mean μ =0 and variance$\sigma^2 = \sigma_{REG,s}^2$. For those individuals who are involved in m $(m > 1)$ leisure activity types, the time usage follows a multivariate normal distribution with variance:

$$\Sigma_{REG,\eta} = \begin{bmatrix} \sigma_{REG,1}^2 & \cdots & \sigma_{REG,1}, \sigma_{REG,m} \\ \vdots & \ddots & \vdots \\ \sigma_{REG,n}, \sigma_{REG,1} & \cdots & \sigma_{REG,m}^2 \end{bmatrix}. \tag{5.6}$$

5.3.3 The Integrated Discrete-Continuous Choice Model

The integrated discrete-continuous choice framework jointly formulates $Y_{REG,\eta}$ (Eqn. 5.5) and Y_{DIS} (Eqn. 5.2) in order to capture the correlation between discrete and continuous decision variables. Therefore, the integrated framework accounts for the following decisions:

- Discrete variable: choices of leisure activities participated by the individuals

(NL, LH, LPC, LOH, LH&LPC, LH&LOH);

- Continuous variable: time spent on each participated activity (LH, LPC, LOH).

In particular, the model accounts for the correlation between leisure activity choices i and time spent on associated activity set η_i. Taking advantage of the fact that error terms of the regressions and the Probit model follow normal distributions, the combination of error term from the two parts will follow a multivariate normal distribution.

$$(\tilde{\varepsilon}_{LH-NL}, \tilde{\varepsilon}_{LPC-NL}, \tilde{\varepsilon}_{LOH-NL}, \tilde{\varepsilon}_{LH\&LPC-NL}, \tilde{\varepsilon}_{LH\&LOH-NL}, \varepsilon_{REG,\eta}) \sim \mathrm{MVN}(0, \Sigma)$$

$\tilde{\varepsilon}_{LH-NL}, \tilde{\varepsilon}_{LPC-NL}, \tilde{\varepsilon}_{LOH-NL}, \tilde{\varepsilon}_{LH\&LPC-NL}, \tilde{\varepsilon}_{LH\&LOH-NL}$ represent error terms in difference of the Probit model respective to NL activity. $\varepsilon_{REG,\eta}$ is a vector of error terms of regressions on given leisure activities subset $\eta \subseteq \{\mathrm{LPC}, \mathrm{LH}, \mathrm{LOH}\}$.

The joint probability of activity choice and time usage can be derived as

$$\mathrm{P}\left(Y_{REG,\eta_i}, Y_{DIS}\right) = \mathrm{P}\left(Y_{REG,\eta_i}\right) P\left(\, Y_{DIS} \middle| Y_{REG,\eta_i}\right), \tag{5.7}$$

or

$$\mathrm{P}\left(Y_{REG,\eta_i}, Y_{DIS}\right) = \mathrm{P}(Y_{DIS}) P\left(\, Y_{REG,\eta_i} \middle| Y_{DIS}\right). \tag{5.8}$$

The likelihood of observing $Y_{DIS} = i$ conditional on Y_{REG} is

$$\widehat{\mathrm{P}}(Y_{DIS} = i | Y_{REG}) = \frac{1}{K} \sum_{k=1}^{K} \mathrm{B}(\widetilde{\mathrm{V}}_{i-j} + \tilde{\varepsilon}_{i-j}{}^{(k)} > 0, \forall j \neq i), \tag{5.9}$$

Where K is the number of simulations, $\tilde{\varepsilon}_{i-j}{}^{(k)}$ is a draw from a multivariate normal with mean $\mu_{DIS|REG}$ and variance $\Sigma_{DIS|REG}$:

If $\begin{bmatrix} \tilde{\varepsilon}_{i-j} \\ \varepsilon_{REG,\eta} \end{bmatrix} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Sigma_{DIS} & \Sigma_{REG,DIS} \\ \Sigma_{DIS,REG} & \Sigma_{REG,\eta_i} \end{bmatrix}\right)$,

then

$$\mu_{DIS|REG} = 0 + \frac{\Sigma_{DIS,REG}}{\Sigma_{DIS}}(err - 0), \Sigma_{DIS|REG} = \Sigma_{REG,\eta_i} - \frac{\Sigma_{DIS,REG}\Sigma_{REG,DIS}}{\Sigma_{DIS}}) \tag{5.10}$$

Then estimation of the model likelihood reduces to collecting the regression error terms when we compute the Probit. In the simulation, the error term that correspond to regression are always kept to whenever the biggest utility is the observed choice, where the simulated value is averaged by the number of success K_n^*. The Simulated Log Likelihood of the model is given by the following formula:

$$\text{SLL}(\beta, \beta_{REG}, \Sigma, X, X_{REG}) = \sum_{n=1}^{N} \log(\frac{K_n^*}{K} \times \phi(Y_{n,REG}|X_{REG,\eta}^T \beta_{REG,\eta}, \varepsilon_{REG,\eta})) \tag{5.11}$$

where, N is the total number of observations in the data, K_n^* is the number of success in the Probit simulation for the n^{th}observation. Simulation has been executed using 1000 pseudo Monte Carlo draws. Standard errors were calculated using Bootstrap re-sampling techniques.

5.4 Model Estimation Results

Results from the integrated discrete-continuous model are reported in Table 5.3 and Table 5.4, where we present model estimates for weekdays and weekend respectively.

Individual and household socio-demographic variables, travel time to work and to social activities enter the final specifications of the estimated discrete-continuous models.

Table 5.3 Joint discrete-continuous model: estimation results of activity in weekdays (t-stats in parenthesis)

	Activity Choice Sub-model				
Variables	**LPC**	**LH**	**LOH**	**LH&LPC**	**LH&LOH**
Alternative-Specific Constant	-1.769	3.782	-1.627	0.553	0.726
	(-9.744)	(34.297)	(-7.809)	(19.782)	(18.675)
No. of children		-0.290	-0.347	-0.593	-0.394
		(-4.803)	(-4.137)	(-4.441)	(-5.082)
Teen (age ≤18)		-2.163	1.124		
		(-3.892)	(3.280)		
Young (19≤ age ≤25)		-2.493	1.532		
		(-3.585)	(3.526)		
Adult (26≤ age ≤40)		-1.440		-0.408	-0.183
		(-4.500)		(-2.757)	(-5.048)
Unmarried couple (dummy)				-0.166	
				(-2.911)	
Graduate or professional degree (dummy)	0.331	-1.042			-1.203
	(1.638)	(-4.742)			(-6.797)
Full time worker (dummy)	-0.273	-1.483		-2.438	
	(-1.811)	(-14.158)		(-14.458)	

	Time Usage Sub-model		
	Time on LPC	**Time on LH**	**Time on LOH**
Alternative-Specific Constant	1.481	4.515	2.586
	(19.558)	(21.552)	(12.241)
Travel time related to socializing and communicating (hrs.)		-0.656	
		(-6.227)	
Travel time related to work (hrs.)	-0.158	-0.696	-0.264
	(-2.453)	(-8.990)	(-6.218)
No. of children	-0.166	-0.227	-0.302
	(-3.424)	(-6.051)	(-3.728)
Age		0.032	-0.029
		(14.697)	(-1.880)
Full time worker (dummy)	-0.371	-1.580	0.111
	(-1.007)	(-14.506)	(-6.117)
Household income		-0.008	-0.002
		(-11.689)	(-0.946)
Log-likelihood (0)	-23755.88		
Log-likelihood (Final)	-18247.51		

Adjusted Pseudo R-squared	0.24
Number of observations	5594

Household income: scaled with 0.001.

Table 5.4 Joint discrete-continuous model: estimation results (weekends) (t-stats in parenthesis)

	Activity Choice Sub-model				
Variables	**LPC**	**LH**	**LOH**	**LH&LPC**	**LH&LOH**
Alternative-Specific Constant	-1.713 (-10.505)	4.073 (29.870)	-1.244 (-6.234)	1.412 (14.954)	0.261 (13.029)
No. of children	-0.228 (-2.827)	0.148 (-2.890)	-0.584 (-2.914)	-0.419 (-4.681)	-1.007 (-5.246)
Teen (age ≤18)	0.844 (4.101)	-3.126 (-4.375)	2.751 (6.535)		2.414 (2.483)
Young (19≤ age ≤25)		-2.062 (-7.447)	0.746 (2.609)	-0.985 (-1.974)	
Adult (26≤ age ≤40)		-2.171 (-8.436)		-0.115 (-3.397)	-0.393 (-3.397)
Senior (41≤ age ≤65)		-0.687 (-7.732)			0.241 (-4.200)
Graduate or professional degree (dummy)	0.501 (2.311)	-1.085 (-4.021)			-1.705 (-2.851)
Full time worker (dummy)		-1.119 (-4.579)		-1.882 (-7.304)	-1.093 (-2.261)
Female (dummy)		-0.734 (-4.778)		-0.929 (-4.326)	-1.353 (-4.813)

	Time Usage Sub-model		
	Time on LPC	**Time on LH**	**Time on LOH**
Alternative-Specific Constant	1.293 (9.821)	4.137 (23.352)	2.235 (20.466)
Travel time related to socializing and communicating (hrs.)	-0.296 (-1.984)	-1.334 (-10.229)	-0.163 (-1.139)
Travel time related to work (hrs.)		-1.631 (-8.505)	-0.298 (-5.190)
Full time worker (dummy)	-0.297 (-2.829)	-0.475 (-5.841)	-0.147 (-2.936)
Female (dummy)	-0.236 (-2.304)		
No. of children	-0.331 (-3.551)	-0.368 (-8.144)	-0.230 (-1.981)
Age	-0.018 (-2.040)	0.031 (10.752)	
Teen (age ≤18)			1.173 (2.559)

Household income		-0.007 (-8.139)
Log-likelihood (0)	-24101.30	
Log-likelihood (Final)	-20240.24	
Adjusted Pseudo R-squared	0.16	
Number of observations	5612	

Household income: scaled with 0.001.

Results obtained from the weekday model, attest that having a graduate or professional degree increases the probability of using social media, but being a full-time worker has an opposite effect, probably because of time constraints. In general, highly educated people with a demanding job tend not to be involved in leisure activities both in the home and out of the home. Teens and young adults are more likely to spend time outdoor for leisure. Having children significantly reduces the probability of being involved in leisure activities, which may be attributable to mobility constraints imposed by the presence of young children on the out-of-home activities of adults (see Scanzoni and Szinovacz, 1980); however, this variable was found to be not significant for leisure involving the use of a computer that is assumed to have a large portion of social media interaction. In addition to activity involvement, the integrated model provides insights on the amount of time spent on PC, at home, or out of home for leisure. It was found that increasing travel time to work reduces the time spent on social media; full-time working status and the fact to have children has similar effects on PC time for leisure. Travel time to social activities and to work, the number of children and income all have negative signs in the linear regressions used to model time use. A full-time job increases the time spent on leisure out of the home.

For weekend days, teens and professionals are highly involved in social media. People with more kids tend to have leisure at home, while teens and young adults still prefer outdoor activities. All other households and individual characteristics that are considered have negative impacts on leisure activity participation. It should be noted that female, that was found to be not significant in the model for weekdays, turns out to be negative and significant also for social media involvement. Concerning time spent for leisure, all travel time-related variables are negative; in particular, individuals going out to socialize have less time to spend on the internet and on leisure activities in general. Teens consistently prefer out of home leisure activities, and income has a negative effect on leisure at home.

Table 5.5 Integrated discrete-continuous model: covariance of difference matrix

a. *Covariance of difference matrix of weekdays*

$$\hat{\Sigma} = \begin{array}{cc} & \begin{array}{cccccccc} LPC & LH & LOH & LH\&LPC & LH\&LOH & T_{LPC} & T_{LH} & T_{LOH} \end{array} \\ \begin{array}{c} LPC \\ LH \\ LOH \\ LH\&LPC \\ LH\&LOH \\ T_{LPC} \\ T_{LH} \\ T_{LOH} \end{array} & \begin{pmatrix} 2.00 & 0.36 & 1.34 & 0.25 & 0.41 & 0.01 & 0.14 & -0.44 \\ 0.36 & 5.45 & -0.64 & 3.52 & 0.15 & -0.01 & 0.63 & 0.11 \\ 1.34 & -0.64 & 14.61 & -3.00 & -0.14 & 0.02 & -0.66 & -0.16 \\ 0.25 & 3.52 & -3.00 & 3.72 & 0.81 & -0.01 & 0.39 & -0.09 \\ 0.41 & 0.15 & -0.14 & 0.81 & 3.99 & 0.00 & -1.47 & -3.42 \\ 0.01 & -0.01 & 0.02 & -0.01 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.14 & 0.63 & -0.66 & 0.39 & -1.47 & 0.00 & 4.27 & 0.66 \\ -0.44 & 0.11 & -0.16 & -0.09 & -0.34 & 0.00 & 0.66 & 1.20 \end{pmatrix} \end{array}$$

b. *Covariance of difference matrix of weekend*

$$\hat{\Sigma} = \begin{array}{cc} & \begin{array}{cccccccc} LPC & LH & LOH & LH\&LPC & LH\&LOH & T_{LPC} & T_{LH} & T_{LOH} \end{array} \\ \begin{array}{c} LPC \\ LH \\ LOH \\ LH\&LPC \\ LH\&LOH \\ T_{LPC} \\ T_{LH} \\ T_{LOH} \end{array} & \begin{pmatrix} 2.00 & 0.98 & 1.79 & -1.74 & 0.01 & 0.19 & 0.04 & 0.09 \\ 0.98 & 3.00 & -0.72 & -1.46 & -0.36 & 0.60 & 0.24 & -0.05 \\ 1.79 & -0.72 & 3.73 & -3.26 & 0.51 & -0.42 & -0.06 & -0.03 \\ -1.74 & -1.46 & -3.26 & 9.89 & 0.27 & 0.33 & -0.97 & -0.01 \\ 0.01 & -0.36 & 0.51 & 0.27 & 2.11 & -0.20 & -0.04 & -0.05 \\ 0.19 & 0.60 & -0.42 & 0.33 & -0.20 & 0.19 & 0.02 & 0.02 \\ 0.04 & 0.24 & -0.06 & -0.97 & -0.04 & 0.02 & 3.46 & 0.12 \\ 0.09 & -0.05 & -0.06 & -0.01 & -0.05 & 0.02 & 0.12 & 1.65 \end{pmatrix} \end{array}$$

From the analysis of the results, it is possible to conclude that based on our sample and for weekends there is a substitution effect between travel time to social activities and time spent for leisure in the home, out of the home and on PC. With particular reference to the objective of this chapter, it can be said that socializing outside the habitual domicile during the weekends reduces the need to communicate via social networks. The same is not true for weekdays when temporal constraints prevent people from meeting in person relatives and friends. A long commute time reduces the time available for leisure, including the time for social media, especially during the weekdays. Moreover, the long commute time also has a stronger negative effect on in-home leisure activity participation and weaker effects on out-of-home activities during weekends. Activity involvements are also varying among different age groups (see Garikapati 2016, for similar results). Our study also confirms that highly educated people are more likely to be social media users, but in average they do not spend more time than the other population groups with their PC for leisure. The estimation results are also consistent with the findings of previous studies (see, Bhat and Misra 1999; Meloni et al. 2007; Kapur and Bhat 2007), in which similar effects of number of young children and travel time to work were found on in-home and out-of- home leisure activity participations. The covariance of difference matrices presented in Table 5.5 indicate that correlations are well captured by the model across activity participations ($LPC, LH, LOH, LH\&LPC, LH\&LOH$) and time usages ($T_{LPC}, T_{LH}, T_{LOH}$).

5.5 Model Validation and Application

For validation purposes, we re-estimated the model on 80% of the available observations in the dataset and then we applied the model estimates to predict the activity and duration choices of the remaining part of the survey sample. The results show that both models do well in prediction for the discrete part (with errors less than

6% for weekdays and less than 5% for weekends). Concerning, the prediction of the activity duration, the overall error on time spent for leisure activity is about 9% for weekdays; we predict a total duration of leisure activities of about 3.62 hours instead of 3.35 hours. The error on the duration of "time spent on LPC" is due the low number of observations available in the sample. Also, the error for weekends is probably due to the higher variability in activity behaviour over weekends.

In Table 5.6 and Table 5.7, we report the actual relative frequencies of activity choices and time usages, the corresponding values predicted by the model together with the difference between observed and predicted values during both weekdays and weekends.

Table 5.6 Discrete-continuous model: validation results of weekdays

		Actual	Predict	Difference
Activity choice frequencies	NL	10.11%	11.61%	1.50%
	LPC	0.89%	1.76%	0.86%
	LH	63.77%	69.56%	5.78%
	LOH	3.22%	2.23%	-0.99%
	LH&LPC	9.31%	3.33%	-5.98%
	LH&LOH	12.70%	11.52%	-1.18%
Time usage on activity choice	Time spent on LPC (hrs.)	1.68	1.22	-27.69%
	Time spent on LH (hrs.)	3.93	4.38	11.42%
	Time spent on LOH (hrs.)	1.15	0.97	-16.17%
	Average time usage on leisure activity (hrs.)	3.35	3.62	8.26%

Table 5.7 Discrete-continuous model: validation results of weekends

		Actual	Predict	Difference
Activity choice frequencies	NL	10.87%	13.26%	2.39%
	LPC	1.16%	2.37%	1.21%
	LH	67.11%	71.00%	3.88%
	LOH	4.55%	3.66%	-0.89%
	LH&LPC	8.56%	6.76%	-1.79%
	LH&LOH	7.75%	2.95%	-4.80%
Time usage on activity choice	Time spent on LPC (hrs.)	1.57	1.22	-22.22%
	Time spent on LH (hrs.)	4.36	5.28	21.22%
	Time spent on LOH (hrs.)	2.40	1.73	-27.94%
	Average time usage on leisure activity (hrs.)	3.86	4.39	13.86%

The models estimated have been also applied to test substitutions effects across different leisure activity types and the variation on the time budget allocated to each activity. The most significant results are reported in Table 5.8. In general, we calculate small variation effects. A unit increase in the number of children produce significant negative effects on leisure activity involvement that the number of no leisure activity participants increase 33.7% during weekdays and 17.4% during weekends. The increases in children number also reduce the average time spent on leisure activities. While engagements in LPC and LH during weekdays would increase due to the increase number of children. More full-time workers (+ 1 unit) will increase engagement in LPC of about 4.6%. Interestingly, the time usage decisions in young, adult and senior groups are more sensitive to the changes in the travel time during weekdays, while the activity participation decisions are barely affected. Increasing travel time to social activities will reduce time spent on leisure, especially time spent on out of home leisure activities. Time on PC is the least affected by travel time to social activities. For example, when travel time to social activities increases by 25%, time spent on LPC will decrease 1.3% in the young group, 2.5% in adult group and 2.3% in the senior group. Under the same scenario, time spent on out-of-home leisure activity will decrease 5.5% in the young group, 9.6% in the adult group, and 10.5% in the senior group. However, the same choices during weekends are not sensitive the changes in both variables. The results also show that activity participations on leisure activities are barely influenced by the increment or decline of travel time to work. However, the variation in the time usage decisions indicate that travel time to work has different influence during weekends and weekdays. For example, when travel time to work increase 25%, time spent on in-home leisure activity will decrease 1% during weekend, and 0.5% during weekdays. While same variation could reduce time spent on LPC by 0.5% during weekends and increase the time by 0.7% during weekdays.

Table 5.8 Sensitivity of activity choices to changes in socio-demographics and travel times

	NL	LPC	LH	LOH	LH& LPC	LH& LOH
Whole sample						
Actual (weekdays)	542	75	3699	198	435	645
No. of children +1	33.7%	31.9%	0.7%	-6.8%	-33.9%	-10.9%
Fulltime worker +10%	3.4%	-0.2%	-0.8%	2.2%	-2.2%	2.4%
Actual (weekends)	605	87	3604	262	542	512
No. of children +1	17.4%	-16.2%	2.9%	-4.8%	-10.7%	-24.2%
Fulltime worker +10%	1.7%	4.6%	-0.1%	2.6%	-2.0%	-1.6%
Young						
Actual	30	11	127	13	23	55
Travel time to social +25%	-0.2%	-0.3%	-0.1%	0.8%	-0.3%	0.2%
Adult						
Actual (weekdays)	233	49	725	81	103	241
Travel time to social +25%	-0.1%	0.3%	0.0%	0.3%	-0.1%	0.1%
Travel time to work +25%	0.0%	0.5%	0.0%	-0.1%	0.1%	0.0%
Actual (weekends)	222	29	890	74	100	99
Travel time to social +25%	0.0%	0.2%	-0.1%	0.1%	0.1%	0.7%
Travel time to work +25%	0.1%	-0.1%	0.0%	-0.5%	-0.2%	0.1%
Senior						
Actual (weekdays)	285	59	1312	149	175	485
Travel time to social +25%	0.0%	-0.3%	0.0%	-0.2%	0.1%	0.2%
Actual (weekends)	254	37	1601	106	237	225
Travel time to social +25%	0.2%	0.5%	0.0%	-0.1%	0.1%	-0.1%

Table 5.9 Sensitivity of duration choices to changes in socio-demographics and travel times

	Time on LPC	Time on LH	Time on LOH	Average time spent on leisure
Whole sample				
Actual (weekdays)	1.07	3.82	0.90	2.91
No. of children +1	-14.0%	-5%	-34.2%	-6.2%
Fulltime worker +10%	1.0%	-1.7%	0.9%	-1.5%
Actual (weekends)	1.91	4.48	2.02	3.62
No. of children +1	-16.5%	-8.3%	-10.6%	-8.7%

Fulltime worker +10%	-0.6%	-0.5%	-0.1%	-0.5%
Young				
Actual	1.91	3.87	0.98	2.80
Travel time to social +25%	-1.3%	-2.5%	-5.5%	-2.6%
Adult				
Actual (weekdays)	1.26	3.09	0.74	2.26
Travel time to social +25%	-2.4%	-4.9%	-9.6%	-4.9%
Travel time to work +25%	-0.7%	-0.5%	-0.7%	-0.5%
Actual (weekends)	1.95	3.60	1.78	2.83
Travel time to social +25%	-0.5%	-1.1%	-0.1%	-1.0%
Travel time to work +25%	0.5%	-1.0%	-0.3%	-0.9%
Senior				
Actual (weekdays)	1.42	3.43	0.73	2.62
Travel time to social +25%	-2.2%	-4.2%	-10.5%	-4.3%
Actual (weekends)	1.84	4.42	1.94	3.56
Travel time to social +25%	0.0%	-0.9%	-0.2%	-0.8%

5.6 Summary

This chapter proposed an econometric model that assesses the impact of computer/internet usage for leisure and relaxing on activity-travel patterns, including social and commute trips. Such internet usage for leisure can be considered as a proxy of social media involvement since activities such as listening to music or watching videos are excluded in our empirical setting. The framework further expands previous analyses that categorize leisure activities into in-home and out-of-home leisure activities, and explicitly models the time spent on each of the activity types considered. The analysis is based on data extracted from the American Time Use Survey and it has been separately performed for weekdays and weekends.

We found that in the U.S., an individual uses a computer for leisure about 1.5h on an average weekday and 2.33h per day on weekends, which makes the analysis particularly important in an activity-based travel analysis context. The empirical results provide valuable insights into the determinants of activity choice and time use decisions of individuals, such as household and individual demographics, and travel time to other activities. The presence of children in the households decreases the likelihood of being involved in leisure activities (including leisure on computer use) except for leisure at home during the weekends. In general, having children also negatively affects the time dedicated to leisure and relax. Individuals with graduate or professional degree are more likely to use the computer for leisure both on weekdays and weekends. Teens and young people are more likely to spend time outside the home for leisure. The time dedicated to leisure activities by young, adult and senior groups are sensitive to the changes in travel time to social activities during weekdays and weekends, except for time spent on out of home leisure during weekends. The model has also been applied to study possible substitution effects. Results attest that the increase in the number of children will decrease participation in leisure activities and an increase in full-time workers will produce more leisure sequences involving the use of PC and social media. More time spent traveling to social activities will decrease participation to leisure activities during weekdays for adults and seniors.

A number of future research avenues are possible. Psychological effects and social interactions play an important role in individual's schedule and activity decisions. Such variables can be included in this model framework to capture the influence of psychological and social changes in leisure activity involvement. Also, if data is available, it would be interesting to study the time spent on cell phone to access social media networks, as there is evidence that *mobile connection has already overtaken fixed internet access (*Chaffey, 2016).

Finally, the same model structure proposed can be applied to model the complete activity-travel pattern and the daily schedule. Mode and destination choice models should be included in the model structure to account for the accessibility to different out of home leisure activities by all the modes available to the individuals in the sample.

Chapter 6: Simultaneous Choice Model on Psychological Well-beings and Activity Involvement

6.1 Problem Description

As a key component, leisure activity provides opportunities to increase the quality of life and to satisfy social needs. At the same time, involvement in leisure activities influences individual's emotions and satisfaction. Decisions on time and locations of leisure activities thus attract great interests of researchers from transportation and economic that out of home activity participation contributes to road traffics and needs higher monetary/time inputs than choosing leisure activity at home.

As a component of subjective well-being, emotions present individual's cognitive and affective evaluation of his/her life (Diener, 1984; Kahneman 1999; Kahneman and Krueger 2006). Diener et al. (1999) pointed out that fulfilment of psychological needs as well as leisure activities may become an important source of individual's well-being. Participation and opportunities for leisure activities that predict well-being could vary across individuals and cultures (Diener et al. 2003; Iwasaki, 2007). Several studies also have examined the relationship between happiness and activity-travel behaviour. Most studies on travel behaviour indicate a positive relationship between emotions and participation in physical leisure (Leung and Lee, 2005) and social activities (Lloyd and Auld, 2002; Robinson and Martin, 2008). As indicated by Kahneman and Krueger (2006), leisure activity has the highest net effects and could bring a higher level of positive emotions (happy, warm, enjoying myself) than any other types of activity. Other than that, studies also found higher level of happiness is associated with out-of-home activities than in-home activities. (Bergstad et al. 2012; Archer et al. 2012). Other stream of studies also has examined how psychological well-being influences our activity choice and travel needs (Ettema et al., 2012; Enam et al., 2017).

However, most studies in the past only focus on the nonreciprocal relationship between activity choice and emotions, a better and extended model is needed to understand the mutual effects between these two choices. The goal of this study is to contribute to the last line of travel behaviour research that investigate the mutual effects with a recursive Probit modelling system. The remaining of the section is organized as follows. First, the study gives an overview of the latest development in well-being and travel behaviour studies, together with related work on modelling. Then modelling framework and method to calculate average treatment effects is proposed in Section 3. Following that, the 2013 American Time Use Survey (ATUS) and associated well-being (WB) module is introduced in the present study. In Section 5, estimation results are presented and discussed in the context of leisure activity involvement and emotions. Finally, concluding remarks and future research directions are indicated.

6.2 Overview of The Study

As well-being is increasingly recognized as an important factor in travel behaviour decisions, the link between travel behaviour and well-being has been identified in many studies (Cantor and Sanderson, 1999; Water et al. 1989; Abou-Zeid & Ben-Akiva, 2012; Diener, 2000). Abou-Zeid and Ben-Akiva (2011) developed an exploratory model to test the presence of correlations between activity/travel well-being and the propensity to participate in activities. They postulated that greater happiness derived from an activity could induce a greater satisfaction with travel to the activity and increase the propensity to participate in the activity. Later, to extent their work, Abou-Zeid and Ben-Akiva (2012) presented the relationship between activity / travel well-being and activity participation for a number of different activity types theoretically and empirically. They also indicated that it could enhance the behaviour realism and forecasting accuracy when take well-being related factors into consideration.

Previous research provides various evidence that shows the impact of activity participations on well-being. Pychyl and Little (1998) measured well-being together with the engaged activities. They found that activities with purposes of personal and social correlated with positively with life satisfaction. Similar findings were reported by Oishi et al. (1999), who investigated the linkage between daily satisfaction and types of activities that people performed. They found that success of achievement and engaging in rewarding social activities influences daily satisfaction. Ravulaparthy et al. (2016) investigated the linkage between activity time use and subjective well-being among different activity involvements. Focusing on the same population group, Enam et al. (2017) investigated the impacts of emotions during activities on individuals' discretionary activity choices in weekends. The study recommended that people with higher level of positive emotions prefer to engage in discretionary activities. Moreover, other streams of studies also contributed to explain the influence of psychological well-being on travel mode choice (Ettema et al., 2016), trip duration (Ettema et al., 2012; Stutzer and Frey, 2008). Studies in the past also found that activities choices also influence our daily well-being. Ettema et al. (2010) revealed cognitive and affective components exist between happiness and daily work commute. A strong connection between the entities are revealed in the study that people feels a great sense of well-being when they engage in activities that are enjoyable or make progress toward achieving goals. Bergstad et al. (2012) assessed the relationship between affects associated with performance of out-of-home activities in a study of Swedish residents. They suggest a strong and significant connection among the performance of routine out-of-home activities and emotions.

In the past, the psychological indicators are constructed as explanatory variables in the RUM based model to study the influence of factors on individual's behaviour (Koppleman and Hause, 1978; Harris and Keane, 1998). However, the indicators in the utility function do not capture all the aspects of the underlying psychological factors and are often associated with measurement errors, which lead to inconsistent and inefficient estimates consequently (Ashok et al., 2002). The hybrid choice modelling framework is thus developed to address the measurement error in the RUM models (Ben-Akiva et al., 2002; Walker and Ben-Akiva, 2002). Bolduc et al. (2008) applied the hybrid choice model to study customer's perceptions and attitudes towards technological innovations. Bolduc and Alvarez-Daziano (2010) considered both a Probit and mixed multinomial logit discrete choice kernel in the HCM by using simulated maximum likelihood method. Later, Bhat and Dubey (2014) proposed a new estimation approach to integrated latent psychological constructs in choice modelling. In the framework, psychological variables are accommodated as ordinal and continuous indicators. These models enable researchers to investigate the influence of psychological variable on the decision-making process. However, mutual effects between activities and well-being could not be captured in the model since it assumes utilities of activities are not constructed as attributes in the function that measures psychological indicators. Zhang and Yen (2017) developed a recursive system that captures the mutual effects between Supplement Nutrition Assistance Program (SNAP) participation and household food insecurity.

6.3 Conceptual Framework

As indicated in previous sections, the goal of this study is to capture the mutual effect between emotions and leisure activity choices. Our empirical specification is motivated by a utility maximization framework such that each individual derives utility of choosing leisure activity (AC) from time saving on other activities (T) and socio-demographics (W).

$$U = U(T, W) \tag{6.1}$$

Two activity choices are categorized as in-home leisure activity (LH) and out-of-home leisure activity (LOH), where AC=1 if choose LH and AC=0 otherwise. Then activity choice can be expressed as

$$P_{AC} = U(T_{AC=1}, W) - U(T_{AC=0}, W) \tag{6.2}$$

Individual will choose in-home leisure if $P_{AC} > 0$ and out-of-home leisure if $P_{AC} \leq 0$.

Assume emotion indicator (EI) is a function of demographic variables (W) and other personal status (Z) such that EI at category k is $EI_k = F(W, Z)$. Then maximizing the utility yields the reduced form equation for individual EI:

$$EI_k^* = F(W, AC^{LH}, Z) \quad if\ P_{AC} > 0 \tag{6.3}$$

$$EI_k^* = F(W, AC^{LOH}, Z) \quad if\ P_{AC} \leq 0 \tag{6.4}$$

Where $AC^{LH} = 1$ and $AC^{LOH} = 0$, and $EI_k = k$ if $\xi_{k-1} < EI_k^* < \xi_k$ where ξ_{k-1} and ξ_k are threshold parameters.

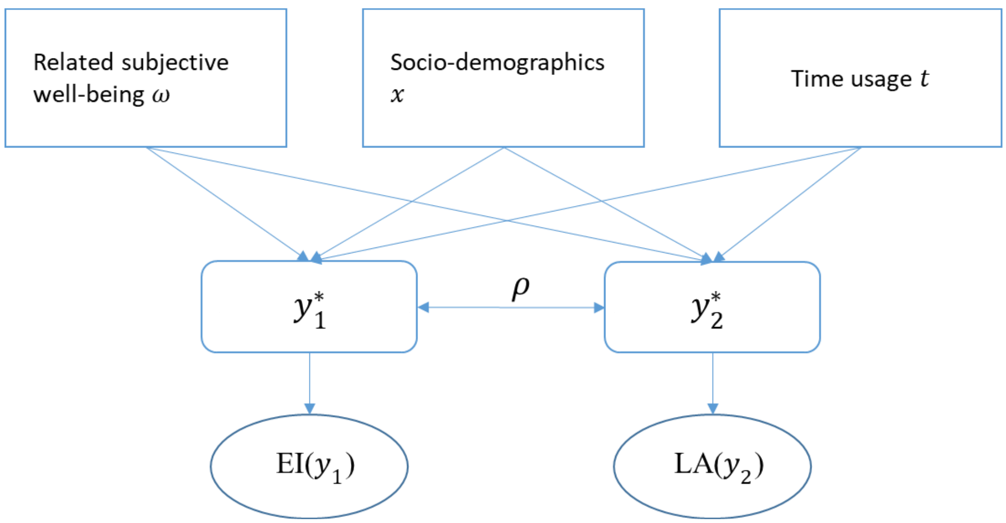

Figure 6.1The recursive modelling framework

6.4 Econometric Procedure

Driven by the theoretical model, we develop a two-equation simultaneous system to deal with the mutual effects of ordinal EI (y_1) and binary AC (y_2). The model is characterized by two structural equations for corresponding latent variables y_1^* and y_2^*

$$y_1^* = \gamma_1 y_2^* + x'\alpha_1 + z'\alpha_2 + \mu_1 \tag{6.5}$$

$$y_2^* = \gamma_2 y_1^* + x'\beta_1 + w'\beta_2 + \mu_2 \tag{6.6}$$

where x, z and w are vectors of exogenous variables with comfortable parameters of $\alpha_1, \beta_1, \alpha_2$ and β_2; γ_1 and γ_2 are scalar parameters, and the error terms are assumed to be bivariate normal distributed with zeros means and unitary variance, correlation ρ and covariance matrix: $\begin{bmatrix}\mu_1\\ \mu_2\end{bmatrix} \sim N(\begin{bmatrix}0\\ 0\end{bmatrix}, \begin{bmatrix}1 & \rho\\ \rho & 1\end{bmatrix})$.

The variance of μ_1 and μ_2 are assumed to be unitary because y_1 is ordinal outcome with only unit increment in each category and y_2 is binary variable. The reduced-form equations are

$$y_1^* = x'\Pi_{11} + z'\Pi_{12} + w'\Pi_{13} + v_1 \quad (6.7)$$

$$y_2^* = x'\Pi_{21} + z'\Pi_{22} + w'\Pi_{23} + v_2 \quad (6.8)$$

where $\Pi_{11}, \Pi_{12}, \Pi_{13}, \Pi_{21}, \Pi_{22}$ and Π_{23} are functions of the structural parameters in Eqn. (6.5) and Eqn. (6.6), and the composite error vector $v = [v_1, v_2]'$ is distributed as bivariate normal with zero means, correlation τ, standard deviation w_1 and w_2, and covariance matrix: $\begin{bmatrix}v_1\\ v_2\end{bmatrix} \sim N(\begin{bmatrix}0\\ 0\end{bmatrix}, \begin{bmatrix}w_1^2 & w_1 w_2 \tau\\ w_1 w_2 \tau & w_2^2\end{bmatrix})$.

Be more specific, we have $w_1^2 = \frac{1+\gamma_1^2+2\rho\gamma_1}{(1-\gamma_1\gamma_2)^2}$, $w_2^2 = \frac{1+\gamma_2^2+2\rho\gamma_2}{(1-\gamma_1\gamma_2)^2}$ and $\tau = [\gamma_1 + \gamma_2 + (1+\gamma_1\gamma_2)\rho]/\sqrt{(1+\gamma_1^2+2\rho\gamma_1)(1+\gamma_2^2+2\rho\gamma_2)}$.

Based on the reduced form of Eqn. (6.7) and Eqn. (6.8), the model with ordinal outcome y_1 and binary outcome y_2 is transformed as

$$y_1 = k \text{ } if \text{ } \xi_{k-1} < y_1^* < \xi_k, k = 0 \ldots K,$$

$$y_2 = \begin{cases} 1, if & y_2^* > 0 \\ 0, if & y_2^* \leq 0 \end{cases},$$

Where ξ_k is threshold parameter such that $\xi_0 = -\infty, \xi_1 = 0, \xi_k = \infty,$ and $\xi_2 \ldots \xi_{k-1}$ are estimable.

Maddala (1983) suggest a two-step estimation of such simultaneous equation system. Although estimates of the two-step procedure are consistent, efficiency cannot be guaranteed. To overcome the shortcoming of two-step estimator, we develop a more efficient maximum likelihood (ML) procedure. Before constructing the likelihood contribution for the sample observation, we define $\psi\Pi_1 = x'\Pi_{11} + z'\Pi_{12} + w'\Pi_{13}$ and $\psi\Pi_2 = x'\Pi_{21} + z'\Pi_{22} + w'\Pi_{23}$, where $\psi = [x', z', w']$. The likelihood contribution for an observation with outcomes $(\mathrm{y}_1 = k, \mathrm{y}_2 = 0)$ and $(\mathrm{y}_1 = k, \mathrm{y}_2 = 1)$ are

$$P(\mathrm{y}_1 = k, \mathrm{y}_2 = 0) = \int_{-\infty}^{-\psi\Pi_2} \int_{\xi_{k-1}-\psi\Pi_1}^{\xi_k-\psi\Pi_1} f(v_1, v_2) dv_1 dv_2 \qquad (6.9)$$

$$P(\mathrm{y}_1 = k, \mathrm{y}_2 = 1) = \int_{-\psi\Pi_2}^{\infty} \int_{\xi_{k-1}-\psi\Pi_1}^{\xi_k-\psi\Pi_1} f(v_1, v_2) dv_1 dv_2 \qquad (6.10)$$

The likelihood function for an independent sample of *n* observations is

$$L = \prod_{i=1}^{n} \prod_{k=1}^{K} \left\{ \Phi_2\left(\frac{\xi_k - \psi_i\Pi_1}{w_1}, \frac{(-1)^{j+1}\psi_i\Pi_2}{w_2}; (-1)^j \tau\right) - \Phi_2\left(\frac{\xi_{k-1} - \psi_i\Pi_1}{w_1}, \frac{(-1)^{j+1}\psi_i\Pi_2}{w_2}; (-1)^j \tau\right)\right\}^{g(\mathrm{y}_i, k)} \qquad (6.11)$$

Where $\Phi_2(x, y, \rho) = \Pr(X \le x, Y \le y)$ is a bivariate standard normal cumulative function (CDF) with correlation ρ, $g(\mathrm{y}_i, k)$ is a dichotomous indicator function which equals 1 if $\mathrm{y}_i = k$ and 0 otherwise, and $\mathrm{j} = \mathrm{y}_{2i}$.

To facilitate interpretation of the effects on explanatory variables, marginal effects of explanatory variables on the probabilities of activity participation and EI are calculated. In addition, to better gauge the effect of leisure activity on each emotion indicator category, we also estimate the average treatment effect (ATE) of leisure activity participation. Specially, for each individual, the probability of out-of-home or at home leisure activity choice is

$$\Pr(y_{2i} = j) = \Phi_1\left(\frac{(-1)^{j+1}\psi_i \Pi_2}{w_2}\right), j = 0,1 \tag{6.12}$$

Where $\Phi_1(.)$ is a standard normal cumulative function (CDF). Applying Eqn. (6.11 - 6.12). The joint probability of each EI category and out-of- home leisure activity is

$$\Pr(y_{1i} = k, y_{2i} = j) = \Phi_2\left(\frac{\xi_k - \psi_i \Pi_1}{w_q}, \frac{(-1)^{j+1}\psi_i \Pi_2}{w_2}; (-1)^j \tau\right) - \Phi_2\left(\frac{\xi_{k-1} - \psi_i \Pi_1}{w_q}, \frac{(-1)^{j+1}\psi_i \Pi_2}{w_2}; (-1)^j \tau\right) \tag{6.13}$$

Applying equation (12) and (13), the conditional probability of EI is

$$\Pr(y_{1i} = k | y_{2i} = j) = \frac{\Pr(y_{1i} = k, y_{2i} = j)}{\Phi_1\left(\frac{(-1)^{j+1}\psi_i \Pi_2}{w_2}\right)}, \tag{6.14}$$

and the conditional probability of choosing out-of-home leisure activity is

$$\Pr(y_{2i} = j | y_{1i} = k) = \frac{\Pr(y_{1i} = k, y_{2i} = j)}{\Phi_1\left(\frac{\xi_k - \psi_i \Pi_1}{w_q}\right) - \Phi_1\left(\frac{\xi_{k-1} - \psi_i \Pi_1}{w_1}\right)} \tag{6.15}$$

Marginal effects of each continuous explanatory variable can be derived by differentiating (Eqn. 6.12 - 6.15). In addition, the treatment effect of in-home leisure activity conditional on individual's emotions is

$$TE_k = \Pr(y_{1i} = k|y_{2i} = 1, y_{1i} > 0) - \Pr(y_{1i} = k|y_{2i} = 0, y_{1i} > 0) =$$

$$\frac{\Pr(y_{1i}=k, y_{2i}=1)}{\Pr(y_{2i}=1) - \Pr(y_{1i}=0, y_{2i}=0)} - \frac{\Pr(y_{1i}=k, y_{2i}=0)}{\Pr(y_{2i}=0) - \Pr(y_{1i}=0, y_{2i}=0)}, \ k = 1 \ldots K \tag{6.16}$$

For statistical inference, standard errors of the marginal and treatment effects can be derived by the delta method (Papke and Wooldridge, 2005)

6.5 Data Descriptions

2013 American Time Use Survey (ATUS) and its well-being (WB) module are used as the primary data sources in the study. We selected *"Level of Happiness"* during selected activities as the emotion indicator (EI) to represent the emotions and feelings that respondent experienced. The emotion indicator is classified into 7 levels that ranges from 0 to 6 (0 is the lowest level of happiness, 6 is the highest level of happiness). Moreover, we are interested in investigating how people's feeling is influencing the leisure activity involvement and related in/out-home activity choices (AC). 6484 leisure activities episodes with EI are distinguished in the survey that are then classified as leisure activity at home (LH) and leisure out-of-home (LOH). For the reason that activity participation with emotion status is randomly selected in the module, respondent without leisure activity involvement cannot be observed in the extracted data.

6.6 Model Estimation Results

The model is developed in-house using MATLAB. Results of the recursive modelling system are reported in Table 6.1, where we present model estimates of the emotion sub model and activity choice sub model. Individual and household socio-demographic variables, travel time to work, shopping and well-being related variables enter the final specification of the estimated model. Average treatment effects (Table 6.3) and marginal effects (Table 6.4) are also calculated to test the influence of each happiness level on activity choices.

Table 6.1 ML estimates of endogenous simultaneous equation model

Variable	**Leisure activity choice**	**Happiness Level**
Latent variables		
In–home leisure (IL)		0.361 (0.193) *
Happiness level (HL)	–0.306 (0.073) ***	
Explanatory variables		
Constant	0.867 (0.178) ***	1.772 (0.306) ***
Age/10	0.142 (0.059) **	–0.088 (0.051) *
$Age^2/1000$	–0.041 (0.062)	0.081 (0.045) *
Male	0.071 (0.039) *	–0.196 (0.029) ***
Household income	–0.008 (0.005)	–0.007 (0.004)
Metropolitan	–0.031 (0.050)	–0.005 (0.038)
Higher than Bachelor's degree	–0.023 (0.062)	–0.119 (0.051) **
Fulltime	–0.175 (0.046) ***	0.095 (0.052) *
Married	0.148 (0.043) ***	0.064 (0.047)
No. of children	0.035 (0.020) *	–0.003 (0.018)
Time on working	–0.019 (0.005) ***	
Time on shopping	–0.056 (0.024) **	
ξ_1		0.173 (0.026) ***
ξ_2		0.442 (0.058) ***
ξ_3		0.966 (0.122) ***
ξ_4		1.441 (0.180) ***
ξ_5		2.008 (0.250) ***
ρ		–0.273 (0.020) ***
Log likelihood	–13345.099	
No. of obs.	6483	

Notes: *** $p < .01$, ** $p < .05$, * $p < .10$. Asymptotic standard errors in parentheses.

Most of the estimates are significant and consistent with expectation. The estimation results of the model reveal that male and elder people are more likely to choose in-home leisure activities and less likely to be happier. Household with higher income are less likely to choose in-home leisure activities and less likely to be happier. Same influence could also be observed on individuals with degree higher than Bachelor. Full time workers are more likely to be happier but have a lower preference to choose in-home leisure. Getting married could increase the probability of choosing in-home leisure and being happier. Increasing number of children in household is negatively influence level of happiness but contributes to the probability of choosing in-home leisure activities. Time spend on working and shopping are negatively related with choice of in-home leisure activities.

The advanced modelling structure proposed allows us to study how household and individuals' socio-demographics are influencing our choices and emotions during activities. Moreover, the model estimates reveal a complex mutual effect between activity choices and emotions. Participation in-home leisure activity ($\beta_{IL} = 0.362$) increase the probability of a higher happiness level, while individuals with a higher happiness level are less likely to participate in-home leisure activities ($\beta_{HL} = -0.306$). In order to investigate how choices of in-home leisure activities are influencing individual's emotions, the average treatment effects of activity choices on each happiness level are calculated in the study.

Table 6.2 Average treatment effects of in–home leisure activity on probabilities of varying emotion indicator

Happiness level	**ATE**
Happiness level = 0 (very unhappy)	0.0231 (0.0026) ***
Happiness level = 1	0.0082 (0.0011) ***
Happiness level = 2	0.0159 (0.0019) ***
Happiness level = 3	0.0377 (0.0044) ***
Happiness level = 4	0.0262 (0.0034) ***
Happiness level = 5	–0.0024 (0.0010) **
Happiness level = 6 (very happy)	–0.1087 (0.0128) ***

Notes: *** $p < .01$, ** $p < .05$, * $p < .10$. Asymptotic standard errors in parentheses.

The results in Table 6.2 indicate that when individuals choose in-home leisure activities as a replacement of out-of-home leisure activities, weak but significant influences could be revealed on each level of happiness. Generally, the average treatment effects (ATEs) suggest for a randomly selected individual who is at relative low level of happiness (happiness score <5), participation in in-home leisure activity could increase his/her current happiness feeling by one level at the probability ranging from 0.8% to 3.7%. However, for individuals at a very high current happiness levels (happiness score =5 or 6), participation in in-home leisure activity would decrease their happiness feelings by one level with the probability of 10%. In sum, our preliminary findings suggest the level of happiness feeling and leisure activities are mutually correlated. On one hand, out-of-home activity increase the level of happiness for individuals at relative low level of happiness; on the other hand, happiness feelings can also decrease the out-of-home activity participation for individuals at very high level of happiness.

Table 6.3 presents the marginal effects of explanatory variables on the joint probability of leisure activity choice and varying happiness levels. Variables of age, gender, household income, education, employment, and marriage status play critical roles in affecting leisure activity type and happiness level jointly. In terms of age, it decreases the probabilities of being 'unhappy' and increases the probabilities of being 'happy' for those who choose to stay at home, while it always decreases the joint probabilities of doing out-of-home leisure and happiness at all levels, which suggest aged people are more likely to do in-home leisure, and age is positively associated with the happiness level for people who choose in-home leisure. Comparing to females, males are less likely to be 'happy' regardless of their leisure activity choices. Be more specific, while choosing in-home leisure activity, a male has 1.4% higher probability to be 'very unhappy' (HL=1) comparing to his female counterpart, but he has 3.55% lower probability to be 'very happy' (HL=7) comparing to his female counterpart. This pattern persists while choosing out-of-home leisure activity. Interestingly, our results reveal the fact that money cannot always guarantee happiness. For individuals at low or moderate happiness levels, the increase of income indeed raises the probabilities of feeling 'happy', but for individuals who are the 'happiest' (HL=7), the increase of income will decrease their probabilities of remaining at the 'happiest' level. Surprisingly, education affects happiness level differently across individuals choosing in-home leisure or out-of-home leisure activities. For individuals who choose out-of-home leisure activity, education has little to do with their feeling of happiness. However, for people choosing in-home leisure activity, more education would only increase the probabilities of those who at relative low happiness levels (increase the probability by 1.07%, 0.34%, 0.63%, 1.33%, 0.72% for HL=1 to HL =5 respectively) but decrease the probability for people who stay at the 'happiest' level (decrease the probability of being HL =7 by 3.36%). The employment status has a contradictory effect on the joint probability of happiness level and leisure activity choice. Comparing to individuals who do not work full time, full-time working individuals who choose in-home leisure activity are less likely to feel 'happiness' while who choose out-of-home leisure activity are more likely to feel 'happiness'. Marriage status has small effects on the joint probability among individuals who choose out-of-home leisure activity. This result coincidences with the common sense that outgoing person tends to have more friends thus feel less lonely. Regarding to individuals who choose in-home leisure activity, married individuals has 1.09% and 4.5% higher probabilities to be HL =6 and HL =7 respectively (very happy) comparing to their single counterparts. The number of children in household and times spent on working and shopping may play some roles in affecting the joint probabilities, but their effects are not significant.

Table 6.3 Marginal effects of explanatory variables on the joint probability of leisure activity choice and happiness levels (HL)

	Choose In–home Leisure Activity						
Variables	**HL = 1**	**HL = 2**	**HL = 3**	**HL = 4**	**HL = 5**	**HL = 6**	**HL = 7**
Age	–0.16 (0.07) **	–0.04 (0.02) *	–0.06 (0.04)	–0.01 (0.10)	0.26 (0.08) ***	0.71 (0.08) ***	1.88 (0.26) ***
Male	1.40 (0.22) ***	0.47 (0.08) ***	0.91 (0.14) ***	2.11 (0.31) ***	1.52 (0.24) ***	0.26 (0.25)	–3.55 (0.80) ***
Household income	0.07 (0.03) **	0.02 (0.01) **	0.04 (0.02) **	0.08 (0.04) **	0.03 (0.03)	–0.06 (0.03) *	–0.33 (0.11) ***
Metropolitan	0.11 (0.26)	0.03 (0.09)	0.06 (0.17)	0.09 (0.39)	–0.02 (0.31)	–0.22 (0.33)	–0.74 (1.03)
Higher than Bachelor's	1.07 (0.44) **	0.34 (0.14) **	0.63 (0.25) **	1.33 (0.53) **	0.72 (0.38) *	–0.41 (0.44)	–3.36 (1.33) **
Fulltime	–0.39 (0.25)	–0.15 (0.08) *	–0.31 (0.16) *	–0.90 (0.37) **	–1.03 (0.29) ***	–1.12 (0.31) ***	–1.02 (0.98)
Married	–0.82 (0.23) ***	–0.25 (0.08) ***	–0.46 (0.15) ***	–0.86 (0.34) **	–0.17 (0.27)	1.09 (0.28) ***	4.50 (0.89) ***
No. of children	–0.05 (0.12)	–0.01 (0.04)	–0.02 (0.08)	0.00 (0.18)	0.09 (0.14)	0.24 (0.13) *	0.60 (0.45)
Time on working	0.04 (0.03)	0.01 (0.01)	0.02 (0.02)	0.02 (0.04)	–0.04 (0.03)	–0.13 (0.04) ***	–0.36 (0.11) ***
Time on shopping	0.12 (0.10)	0.03 (0.03)	0.05 (0.06)	0.05 (0.13)	–0.11 (0.11)	–0.39 (0.16) **	–1.08 (0.48) **
	Choose Out–of–home Leisure Activity						
Variables	**HL = 1**	**HL = 2**	**HL = 3**	**HL = 4**	**HL = 5**	**HL = 6**	**HL = 7**
Age	–0.10 (0.01) ***	–0.05 (0.01) ***	–0.11 (0.01) ***	–0.36 (0.04) ***	–0.50 (0.06) ***	–0.66 (0.08) ***	–0.81 (0.16) ***
Male	0.08 (0.04) **	0.03 (0.02) *	0.05 (0.04)	0.06 (0.13)	–0.16 (0.17)	–0.66 (0.24) ***	–2.53 (0.45) ***
Household income	0.01 (0.01) ***	0.01 (0.00) **	0.01 (0.01) **	0.04 (0.02) **	0.05 (0.02) *	0.04 (0.03)	–0.02 (0.06)
Metropolitan	0.04 (0.05)	0.02 (0.02)	0.04 (0.05)	0.12 (0.16)	0.16 (0.23)	0.18 (0.32)	0.15 (0.60)
Higher than Bachelor's	0.13 (0.08) *	0.05 (0.03)	0.11 (0.07)	0.28 (0.23)	0.22 (0.30)	–0.02 (0.39)	–1.09 (0.69)
Fulltime	0.11 (0.05) **	0.06 (0.02) **	0.13 (0.05) ***	0.50 (0.16) ***	0.77 (0.22) ***	1.19 (0.30) ***	2.16 (0.57) ***
Married	–0.20 (0.04) ***	–0.09 (0.02) ***	–0.20 (0.04) ***	–0.64 (0.14) ***	–0.78 (0.19) ***	–0.84 (0.27) ***	–0.28 (0.51)
No. of children	–0.03 (0.02) *	–0.02 (0.01) *	–0.04 (0.02) *	–0.12 (0.07) *	–0.17 (0.09) *	–0.22 (0.13) *	–0.26 (0.25)
Time on working	0.02 (0.01) ***	0.01 (0.00) ***	0.02 (0.01) ***	0.07 (0.02) ***	0.09 (0.02) ***	0.12 (0.03) ***	0.12 (0.06) *
Time on shopping	0.06 (0.02) **	0.03 (0.01) **	0.06 (0.03) **	0.21 (0.09) **	0.27 (0.11) **	0.34 (0.15) **	0.36 (0.22)

Notes: All effects on probabilities are multiplied by 100. Asymptotic standard errors in parentheses. Statistical significance *** at the 1% level, ** at the 5% level, and * at the 10%

6.7 Summary

This section investigates the mutual effects between leisure activity choices and emotions with a recursive system. The study uses the 2013 American Time Use Survey and its well-being module from U.S. Bureau of Labor Statistics (BLS) which enable us develop a comprehensive model of people's feelings of well-being as a function of activity-travel and time use patterns. The results present us complex mutual effects between activity choices and emotions. Participating in-home leisure activity positively contributes to the probability of a higher happiness level, which is consistent with the finding in previous studies. However, based on the results, individuals with a higher happiness level are less likely to participate in-home leisure activities.

In order to gain insight how in-home leisure activities choices are influencing individual's emotions, the average treatment effects of activity choices on each happiness level are calculated in the study. Concerning the average treatment effects, compared to out-of-home leisure activity, participation in in-home leisure activities increases the probability of lower level of happiness (level 0-4). While it reduces the probability of having higher level of happiness (level 5-6).

The study is the first to evaluate the mutual effects between activity choices and emotions with a simultaneously Probit model system. Findings contribute to the last line of studies on travel behaviours and well-being. The average treatment effects found in the study indicate complex results in the travel behaviour when involved with different levels of emotion. Findings of the study on the role of socio-demographics and mutual effects are important for the public sector-by informing policy makers concerned about the social well-being and transport planning. In the future, the modelling system could be expanded to incorporate multinomial Probit models or multiple order Probit models. On the other side, the modelling system could also be extended to account for the interdependence among emotions of multiple continuous activities.

Chapter 7: Extended Hybrid Choice Model on Psychological Well-beings, Activity Involvement and Time Usage

7.1 Problem Description

Daily activities constitute a key part of people's lives, such activities include working, studying, escorting, shopping, and various other forms of pastimes. Among them, leisure activity plays an important role for providing opportunities to improve the quality of life and to satisfy social needs. It also influences individual's daily emotions and satisfaction. Feelings and emotions during and after activity are also critical in determining whether individuals want to maintain their involvement (Biddle et al., 2003). As a component of SWB, together with life satisfaction, emotions represent individuals' cognitive and affective evaluation of their own life (Diener, 1984; Kahneman 1999; Kahneman and Krueger 2006). Happiness, as a part of SWB, has attracted a plethora of cross-disciplinary research in recent year. From a broad context, earlier studies pointed out that SWB could be affected by culture, personality, household and individual socio-demographic. (Diener et al, 2003; Myers and Diener 1995; Ryan and Deci 2001; Dolan et al, 2008).

As a form of leisure activities, activities associated to computer use and online communication become essential parts of people's daily life. It is also expected that intensive online activities change individual's activity needs and activity forms by relaxing spatial and time constraints in activity participations and modify user's activity patterns and their perception of activity duration (Wang and Law, 2007, Carrasco and Miller, 2008, Ohmori, 2008, Ben-Elia, et al., 2014, Dong et al. 2017). Understanding the role of emotions hidden behind such activities could help us better understand users' choice on these emerging activity participation patterns.

The goal of Chapter 7 is to develop a comprehensive model system that is able to incorporate the role of emotions on individual's activity-travel choice and time use decision. Based on the model introduced in Chapter 6, the part incorporates well-being indicators into the discrete continuous modelling framework. More specifically, the study explores the potential influence of subjective well-being on daily leisure activity participation (choice of leisure activity) and time use (duration of each activity). The deriving model jointly estimates decisions related to time use, activity participation, and level of happiness. An iterative simulated maximization likelihood method is also proposed in the model estimation to avoid the bias caused by endogeneity.

7.2 Methodology

7.2.1 Research Design

Researchers in the past have proved that emotions could influence time use and activity participation (Archer et al. 2012; Kahneman and Krueger 2006; Enam et al. 2015). To explore, to what extent, the emotions affect our choices, it is essential to understand the correlation pattern between activity, time use and emotions incurring when performing a given activity. The hybrid choice model proposed in this study include three sub models that are jointly estimated: the activity type model (Section 7.2.4), the time use model (Section 7.2.3), and the latent variable models that account for emotions and

SWB (Section 7.2.2). The framework is described in Figure 7.1 and it is developed in context of leisure activity participation. To capture the influence of subjective well-being on time use decisions, latent variables are formulated as an attribute in both the time use model and the well-being model. Interdependence between activity participation, time use decisions, and individual's emotion status are also captured in the integrated model.

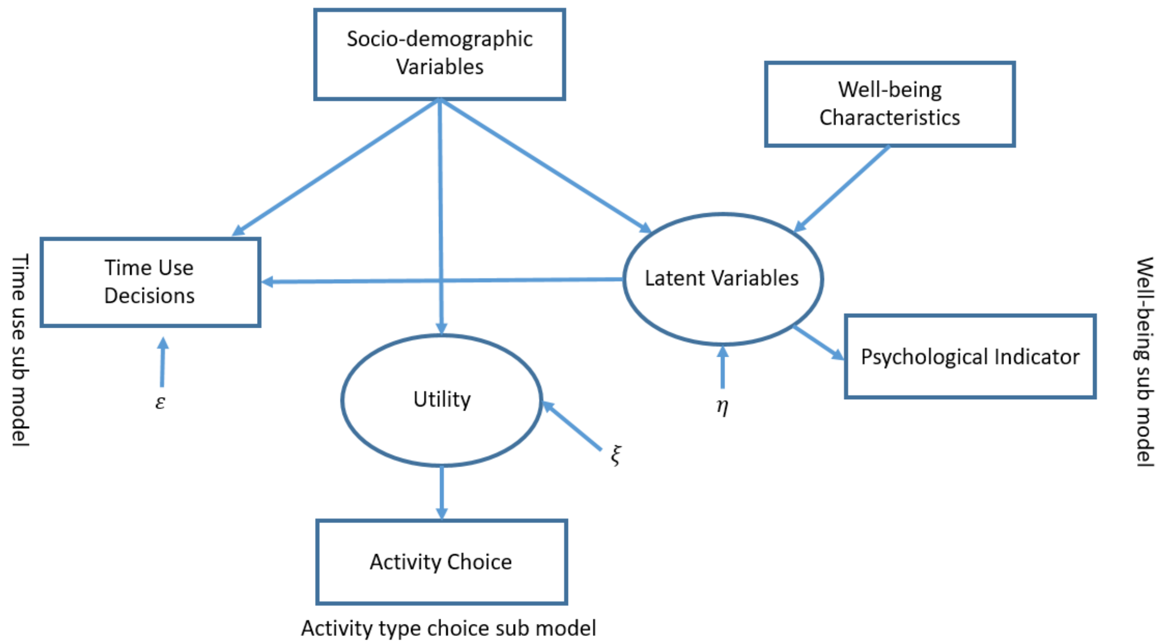

Figure 7.1 Structure of modelling framework

7.2.2 Well-being Measurement Sub-model

In this study, psychological indicators such as the level of happiness during leisure activity is regarded as an ordinal indicator, which is captured through an ordered Probit model. Latent variable Z^*in Eqn. (7.1), which is a vector of continuous latent responses, represents the emotions during leisure activities in the model.

$$Z^* = W^T\omega + X_{Lat}^T\beta_{Lat} + \eta, \eta \sim N\ (0, \sigma_\eta^2), \tag{7.1}$$

where W is a vector of observed variables related to individual well-being status, X_{Lat} is a vector of socio-demographic variables, ω and β_{Lat} are the corresponding coefficients, and η is a random error term assumed to be normally distributed.

The latent variable $\hat{Z}^*$ is then linked to level of happiness *I* through the following ranking scale:

$$I = \begin{cases} 0, & if\ Z^* < 0 \\ 1, & if\ 0 \le Z^* < \gamma_1 \\ \vdots & \\ k-1, & if\ \gamma_{k-2} \le Z^* < \gamma_{k-1} \\ k, & if\ \gamma_{k-1} \le Z^* \end{cases}, \tag{7.2}$$

where k is the indicator of emotion level.$\gamma_{k-1}$is the *k-1* th threshold. It is worth to remind here that the likelihood of the ordered Probit is given by the following equation:

$$P(I) = \prod_k^K \Phi(\frac{\gamma_k - Z^*}{\sigma_\eta^2}) - \Phi(\frac{\gamma_{k-1} - Z^*}{\sigma_\eta^2}) \tag{7.3}$$

7.2.3 Time Use Measurement Sub Model

In time use sub model, a multiple regression model is used to estimate the time spent on leisure activity. Time use decision Y_{REG}, formulated as a continuous dependent variable, is linked to respondent's household and individual socio-demographic variables X and latent variable Z^* through (Eqn. 7.4):

$$Y_{REG} = X_{REG}^T\beta_{REG} + (Z^*)^T\lambda + \varepsilon\ ,\ \ \varepsilon \sim N\ (0, \sigma_\varepsilon^2) \tag{7.4}$$

where β_{REG} is the vector of coefficients corresponding to socio-demographic variables, λ is the loading on latent variable, and ε_{REG} is a normally distributed measurement error term.

As state in Chapter 6, the problem can also be expressed in the form of a likelihood function to be maximized. The likelihood of observing Y_{REG} is given by the normal density function:

$$P(Y_{REG}) = \phi\left(Y_{REG} - \hat{Y}_{REG} \middle| \mu = 0, \sigma^2 = \sigma_{\varepsilon}^2\right). \tag{7.5}$$

7.2.4 Activity Type Choice Sub Model

Discrete choice analysis is used to model participation in leisure activity. Leisure activity types thus constitute the set of alternatives in the model. To each activity type choice j, we associated a utility function:

$$U_j = X_j^T \beta_j + \xi_j \tag{7.6}$$

where β_j are the coefficients associated to socio-demographic attributes X_j^T, and ξ is a normally distributed error term.

A multivariate Probit model is adopted to solve the discrete problem, and therefore the error terms follow a multivariate normal distribution with full, unrestricted covariance matrix. The Probit model is normalized to consider the fact that the level and scale of the utility is irrelevant (Train, 2009). The probability of choosing a given leisure activity *j* can also be expressed in the way of difference:

$$\mathrm{P}(Y = \mathrm{j}) = \int B(\widetilde{\mathrm{V}}_{j-l} + \tilde{\xi}_{j-l} > 0, \forall j \neq l)\phi(\tilde{\xi})d\tilde{\xi} \tag{7.7}$$

where $B()$ is a Boolean indicator of whether the statement in parentheses holds is, $\phi(\tilde{\xi})$ is the density of the error term in differential formation (i.e. $\tilde{\xi}_{j-l} = \xi_j - \xi_l$), $\widetilde{V}_{j-l} = X_j^T \beta_j - X_l^T \beta_l$.

7.2.5 Estimation Process

The decision maker is assumed to be rational and to choose the alternative that maximizes their utilities. Taking advantage of the fact that the error terms of the regression, the ordered Probit and the multinomial Probit follow all a normal distribution, the combination of error terms from the three models follow a multivariate normal distribution:

$$\begin{bmatrix} \tilde{\xi}_{j-l} \\ \eta \\ \varepsilon_{REG} \end{bmatrix} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \Sigma\right)$$

where $\Sigma = \begin{bmatrix} \tilde{\Sigma}_{DIS} & \Sigma_{DIS,LAT} \\ \Sigma_{DIS,LAT} & \Sigma_{LAT} \end{bmatrix}$, $\Sigma_{LAT} = \begin{bmatrix} \Sigma_{REG} & \Sigma_{REG,I} \\ \Sigma_{I,REG} & \Sigma_I \end{bmatrix}$

The joint probability of these models can be derived as

$$\mathrm{P}(Y = j, Y_{REG}, I) = \mathrm{P}(Y = j | Y_{REG}, I) P(I | Y_{REG}) \mathrm{P}(Y_{REG}), \qquad (7.8)$$

The likelihood of *n* th observation with indicator $I = k_i$ conditional on Y_{REG} can then be written as

$$P(I | Y_{REG}) = \Phi\left(\frac{\gamma_{k_n+1} - \mu_{I|REG}}{\sigma_{I|REG}}\right) - \Phi\left(\frac{\gamma_{k_n} - \mu_{I|REG}}{\sigma_{I|REG}}\right), \qquad (7.9)$$

where $\Phi(.)$ is the cumulative distribution function of the standard normal distribution with

$$\mu_{I|REG} = \frac{\sigma_{I,REG}}{\sigma_{REG}}\left(Y_{REG} - \hat{Y}_{REG}\right),$$

$$\sigma_{I|REG} = \sigma_I - \frac{\sigma_{I,REG}\sigma_{REG,I}}{\sigma_{REG}}. \quad (7.10)$$

Since the Probit model ($\mathrm{P}(Y)$) has no closed mathematical form, simulation is applied as described in (Train, 2009, p. 117) to estimate individuals' activity participations. The likelihood of observing $Y_{DIS} = j$ conditional on Y_{REG} and $I = k$ can be expressed as

$$\widehat{\mathrm{P}}(Y_{DIS} = j|Y_{REG}, I = k) = \frac{1}{M}\sum_{m=1}^{M} \mathrm{B}(\widetilde{\mathrm{V}}_{ij} + \tilde{\xi}_{j-l}{}^{(m)} > 0, \forall j \neq l),$$

(7.11)

Where M is the number of simulations, $\tilde{\xi}_{j-l}{}^{(k)}$ is a draw from a multivariate normal with mean $\mu_{DIS|REG}$ and variance $\Sigma_{DIS|REG}$:

$$\mu_{DIS|REG} = 0 + \frac{\Sigma_{DIS,REG}}{\Sigma_{DIS}}\left(Y_{REG} - \hat{Y}_{REG}\right), \Sigma_{DIS|REG} = \Sigma_{REG} - \frac{\Sigma_{DIS,REG}\Sigma_{REG,DIS}}{\Sigma_{DIS}})$$

(7.12)

The parameters that need to be estimated are θ = ($\omega, \beta_{LAT}, \lambda_{REG}, \beta_{REG}, \beta_j, \Sigma$). The general form of Simulated Log Likelihood (SLL) of the integrated model is given by the following formula:

$$\mathrm{SLL}\left(\theta, X_{LAT}, X_{REG}, X_j\right) = \sum_{n=1}^{N} \log \mathrm{P}(Y|Y_{REG}, I)P(I|Y_{REG})\mathrm{P}(Y_{REG}) =$$

$$\sum_{n=1}^{N} \log \frac{1}{M} [\Phi\left(\frac{\gamma_{k_n+1} - \mu_{I|REG}}{\sigma_{I|REG}}\right) - \Phi\left(\frac{\gamma_{k_n} - \mu_{I|REG}}{\sigma_{I|REG}}\right)]\phi\left(Y_{REG} - \hat{Y}_{REG}\middle|\mu = 0, \sigma^2 = \sigma_{\varepsilon}^2\right)$$

(7.13)

where, N is the total number of observations in the data. Simulation has been executed using 5000 Monte Carlo draws. Standard errors were calculated using Bootstrap re-sampling techniques.

7.2.6 Accounting for Endogeneity

In econometrics, an endogeneity problem occurs when the explanatory variables are correlated with the error terms. As anticipated in previous section, the latent variable associated with emotion indicator I becomes an attribute in the time use sub model. Also, we assume that the emotion structural sub model and the time use measurement sub model are correlated, and that the unobserved variable l accounts for the endogeneity caused by the emotions on time use decisions. Then Eqn. (7.1) can be written as

$$\tilde{Z}^* = W^T\omega + X_{Lat}^T\beta_{Lat} + l + \eta, \eta \sim N\ (0, \sigma_\eta^2) \tag{7.14}$$

where $\tilde{Z}^* = Z^* + l$.

Eqn. (7.4) can be written as

$$Y_{REG} = X_{REG}^T\beta_{REG} + (\tilde{Z}^*)^T\lambda + l + \varepsilon\ ,\ \ \varepsilon \sim N\ (0, \sigma_\varepsilon^2) \tag{7.15}$$

An iterative estimation method is applied; at each iteration, expected value $\hat{Z}^*$ is calculated first based on estimates of ω and β_{LAT}. Then coefficients $(\beta_{REG}, \lambda_{REG})$ are estimated based on $\hat{Z}^*$, followed by the calculation of the likelihood of the joint model.

It should be pointed out that the model structure is jointly estimated and that standard errors are calculated on the full information matrix. More specifically, the estimation process can be described by the following procedure. In each iteration:

(1) Update $(\hat{Z}^*)^{(t)}$ using Eqn. (7.1) with $\omega^{(t)}, {\beta_{LAT}}^{(t)}$

(2) Update ${\beta_{REG}}^{(t)}\ and\ {\lambda_{REG}}^{(t)}$ via the least-square estimation of $(X_{REG}^T \beta_{REG} + \left(\hat{Z}^{*(t)}\right)^T \lambda)$ on Y_{REG}.

(3) Update $P(Y_{REG})$ with ${\hat{Y}_{REG}}^{(t)}, {\sigma_{REG}^2}^{(t)}$

(4) Update ${\mu_{I|REG}}^{(t)}\ and\ {\Sigma_{I|REG}}^{(t)}$ using Eqn. (7.10) with ${\Sigma_{LAT}}^{(t)}\ and\ {\hat{Y}_{REG}}^{(t)}$.

(5) Update ${\mu_{DIS|LAT}}^{(t)}\ and\ {\Sigma_{DIS|LAT}}^{(t)}$ using Eqn. (7.12) with $\Sigma^{(t)}, (\hat{Z}^*)^{(t)}, and\ {\hat{Y}_{REG}}^{(t)}$. Z^* in Eqn. (7.12) equals to $(\hat{Z}^*)^{(t-1)}$.

(6) Update $(\lambda^{(t+1)}, \omega^{(t+1)}, {\beta_{LAT}}^{(t+1)}, \beta_j, \Sigma^{(t+1)})$ through

$(\lambda^{(t+1)}, \omega^{(t+1)}, {\beta_{LAT}}^{(t+1)}, {\beta_j}^{(t+1)}, \Sigma^{(t+1)}) = argmax$

$\mathrm{SLL}\left(\theta, X_{LAT}, X_{REG} | {\mu_{DIS|LAT}}^{(t)}, {\Sigma_{DIS|LAT}}^{(t)}, {\mu_{I|REG}}^{(t)}, {\Sigma_{I|REG}}^{(t)}\right)$

The proposed estimation procedure also reduces the dimension of the simulated log-likelihood problem as the variables in the regression are estimated first with least square estimation. This reduces estimation time and saves computational power.

7.2.7 Model Identification

The identification problem of the modelling structure in Figure 1 can be classified as a single factor model with two correlated indicators (O'Brien, 1994; see also Vij and Walker, 2014 for a comprehensive overview of the identification problem that exists in hybrid choice models). In this case, according to O'Brien (1994), if the factor loadings of both indicators are identified then the covariance between the measurement errors of the two indicators is also identified. In the structural equation Eqn. (7.14), loadings of Z^* on I is identified as 1. Thus, the sufficient condition for the model to be identified is that loading of Z^* on Y_{REG} Eqn. (7.11), which is λ, is also identified.

However, any missing measurement of l in the simulated estimation process will increase the correlation $\mathrm{E}\left(\tilde{Z}^*, \varepsilon\right)$, which renders the estimate of $\binom{\hat{\beta}_{REG}}{\hat{\lambda}}$ biased. To avoid this bias, Z^* is imposed as an instrument where $\mathrm{E}(Z^*, \varepsilon) = 0$ and $\mathrm{E}\left(Z^*, \tilde{Z}^*\right) \neq 0$.

Moreover, the estimation of the covariates in the time use sub-model is isolated from the estimation of covariance matrix to ensure that the factor loading is identified. Other covariates in the emotion sub-model and time use sub-model are restricted to be exogenous.

7.3 Synthetic Experiment

The simulation experiment described in this Section tests the ability of the integrated model to produce consistent and unbiased estimators when latent constructs are specified as independent variables in the time use model. The three sub-models are specified by the analyst according to specific assumptions. A total of 3,000 observations are simulated in each dataset; and 100 different datasets are created maintaining the same specification.

In the well-being sub model formulated as in Eqn. (7.1), three variables are included as covariates in the vector w to explain the latent variable Z^*. The vector of coefficients to be estimated is assumed to be $Vech(W) = (W_1 = 0.4, W_2 = 0.6, W_3 = -0.3)$.

Therefore Eqn. (7.1) can be rewritten for this specific case as follows:

$$Z^{*}=[0.4 \quad 0.6 \quad -0.3]\begin{bmatrix}\omega_1\\ \omega_2\\ \omega_3\end{bmatrix}+\eta, \eta \sim N\ (0, \sigma_\eta^2), \tag{7.16}$$

The relation between the latent variable Z^{*} and the observed indicator variable I is given in Eqn. (7.15). In this simulated experiment, the indicator in Eqn. (7.2) is assumed to vary over four levels. Therefore, two threshold parameters need to be estimated in the model; they are assumed to be $\gamma_1 = 1.1$ and $\gamma_2 = 2.4$.

$$I=\begin{cases} 0, & if\ Z^{*}<0\\ 1, & if\ 0 \le Z^{*} < 1.1\\ 2, & if\ 1.1 \le Z^{*} < 2.4\\ 3, & if\ 2.4 \le Z^{*}\end{cases}. \tag{7.17}$$

In the time use sub model, a regression model relates the continuous independent variable to three covariates. The latent variable $\hat{Z}^{*}$ and the independent variable ω_1 that appears also in the well-being sub-model are included as covariates of this sub model. We expect the presence of ω_1 in both the latent variable model and in the time use model to cause endogeneity. The vector of coefficients β_{REG} to be estimated has four elements: $\beta_{REG} = (\beta_{R1} = 1.5, \beta_{R2} = 1.1, \beta_{R3} = 0.7, \lambda_{REG} = -0.5)$. Then Eqn. (7.4) can be written as follows:

$$Y_{REG}=[1.5 \quad 1.1 \quad 0.7 \quad -0.5]\begin{bmatrix}Cons\\ \omega_1\\ \omega_4\\ \hat{Z}^{*}\end{bmatrix}+\varepsilon\ ,\ \ \varepsilon \sim N\ (0, \sigma_\varepsilon^2) \tag{7.18}$$

In the activity choice sub model three alternatives are considered in the choice set and the utilities of the random utility model (Eqn. 7.6) are specified as:

$$\begin{matrix} U_1 \\ U_2 \\ U_3 \end{matrix} = \begin{bmatrix} 0 & 0 & 0 \\ ASC & \omega_5 & \omega_6 \\ ASC & \omega_7 & \omega_8 \end{bmatrix} \begin{bmatrix} 0 & 1.5 & 1.9 \\ 0 & 1.8 & 1.3 \\ 0 & 0.3 & 0.2 \end{bmatrix} + \begin{bmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{bmatrix} \tag{7.19}$$

In the above equation, the parameters to be estimated include the elements of β vector ($\beta_{2,ASC} = 1.5, \beta_{2,\omega_5} = 1.8, \beta_{2,\omega_6} = 0.3, \beta_{3,ASC} = 1.9, \beta_{3,\omega_7} = 1.3, \beta_{3,\omega_8} = 0.2$).

The covariance matrix of the integrated model that is specified according to Σ:

$$\Sigma = \begin{pmatrix} 2.30 & 1.05 & 0.00 & 0.25 \\ 1.05 & 2.50 & 0.00 & 0.55 \\ 0.00 & 0.00 & 1.00 & 0.15 \\ 0.25 & 0.55 & 0.15 & 1.00 \end{pmatrix} = L_\Sigma L'_\Sigma$$

$$= \begin{pmatrix} 1.00 & 0.00 & 0.00 & 0.00 \\ 0.69 & 1.42 & 0.00 & 0.00 \\ 0.00 & 0.00 & 1.00 & 0.00 \\ 0.16 & 0.31 & 0.15 & 0.93 \end{pmatrix} \begin{pmatrix} 1.00 & 0.69 & 0.00 & 0.16 \\ 0.00 & 1.42 & 0.00 & 0.31 \\ 0.00 & 0.00 & 1.00 & 0.15 \\ 0.00 & 0.00 & 0.00 & 0.93 \end{pmatrix} \tag{7.20}$$

Seven Cholesky matrix elements need to be estimated in L_Σ; they are ($L_{21} = 0.692, L_{22} = 1.421, L_{33} = 1.000, L_{41} = 0.165, L_{42} = 0.307, L_{43} = 0.150, L_{44} = 0.925$).

Estimation is performed for each of the 100 synthetic datasets obtained according to the assumptions described. Mean estimated coefficients for each variable of the integrated model are obtained by averaging the results over the 100 estimates obtained. Standard errors and Mean Absolute Error (MAE) are calculated to assess the quality of the estimates. In particular, MAE is calculated according to the following equation:

$$M = \frac{1}{N} \sum_{t=1}^{N} |A_t - F_t|, \tag{7.21}$$

where A_t is the actual value, F_t is the estimated value, and *N* is the total number of synthetic files.

Table 7.1 provides the simulation results, together with the true value of the parameters, followed by the estimated parameter and the standard error of estimates. The model is able to recover most of the parameters remarkably well. MAE value of most variables in the well-being sub model and in the time use sub model are lower than 0.04.

Table 7.1 Estimation results of simulation -study

Parameters	**Actual Value**	**Mean Est.**	**Standard Error**	**Absolute Bias of Estimators**	**Mean Absolute Error (MAE)**
W_1	0.400	0.393	0.063	0.007	0.048
W_2	0.600	0.591	0.091	0.009	0.068
W_3	-0.300	-0.298	0.048	0.002	0.037
γ_1	1.300	1.285	0.197	0.015	0.148
γ_2	2.400	2.373	0.168	0.027	0.020
β_{R1}	1.500	1.501	0.018	0.001	0.014
β_{R3}	0.700	0.699	0.019	0.001	0.014
β_{R2}	1.100	1.101	0.025	0.001	0.020
λ_{REG}	-0.500	-0.519	0.079	0.019	0.063
$\beta_{2,ASC}$	1.500	1.493	0.091	0.007	0.089
β_{2,ω_5}	1.800	1.710	0.095	0.090	0.049
β_{2,ω_6}	0.300	0.353	0.031	0.053	0.099
$\beta_{3,ASC}$	1.900	1.907	0.097	0.007	0.024
β_{3,ω_7}	1.300	1.236	0.087	0.064	0.094
β_{3,ω_8}	0.200	0.204	0.033	0.004	0.065
L_{21}	0.692	0.760	0.090	0.068	0.088
L_{22}	1.421	1.442	0.114	0.021	0.069
L_{33}	1.000	0.982	0.071	0.018	0.064
L_{41}	0.165	0.171	0.046	0.006	0.028
L_{42}	0.307	0.300	0.031	0.007	0.057
L_{43}	0.150	0.132	0.054	0.018	0.036
L_{44}	0.925	0.927	0.018	0.002	0.025

7.4 Real Case Study

So far, our investigation has been based on synthetic data designed specifically for pre-defined parameters. We turn now our attention to a real case study where the the primary data source used in this analysis is extracted from the 2013 American Time Use Survey (ATUS) and its well-being (WB) module. The ATUS is designed and collected by the U.S. Bureau of Labor Statistics and contains detailed information on each activity respondents were involved one day before the interview. Activity related attributes include start and end time of each activity episode participation, activity type, and activity location; individual and household socio-demographic characteristics are also available in the survey. Both at home and out of home activities are reported, which makes ATUS particularly attractive for time use analysis and modelling (Dong et al., 2017). Besides that, the ATUS well-being module contains information relative to three activities that are randomly selected from each respondent's activity diary. Emotions and feelings during selected activity, general health information and satisfaction of life are recorded for each respondent selected.

7.4.1 Data Description

In this study we are interested in investigating how people's feelings affect participation in leisure activities and time spent on leisure. We distinguish between in home and out of home leisure activities and between generic leisure activities and those involving the use of the computer. Using data obtained by combining ATUS general purpose survey and the well-being module, a total of 1,506 observations related to leisure activities performed on weekdays are obtained. For the reason that activity participation with well-being status is randomly selected from individual's activity diary, respondent who did not report any leisure activity are not part of our sample. The resulting set of discrete choices over leisure activity types includes the following three alternatives:

- Leisure activities that involve the use of the computer (LPC);
- In-home other (than computer use) leisure activities (LH);
- Out-of-home other (than computer use) leisure activities (LOH);

The distribution of the alternatives over the selected sample, together with descriptive statistics are reported in Table 7.2.

Table 7.2 Distribution of leisure activity

Category	**Obs.**	**Time spent on leisure (hr.)**			
		Min	**Max**	**Med.**	**Mean**
Out-of-home other leisure activities (LOH)	108	0.08	3.92	0.29	0.70
In-home other leisure activities (LH)	1265	0.08	15.48	1.47	1.94
Computer use for leisure activities (LPC)	133	0.08	12.42	1.00	1.26

Household characteristics, land-use variables, time use and well-being variables for each individual in the sample, are the main variables extracted from the original dataset. Seven levels of happiness are defined in the well-being module ranging from 0 ~ 6, in which 0 indicates the lowest level of happiness and 6 indicates the highest level of happiness. In the study happiness are aggregated to five levels ranging from 0 to 4 (Level 2 and 3 are labeled as 2; Level 4 and 5 are labeled as 3; Level 6 is labeled as 4).

Table 7.3 lists the basic statistics relative to the sample.

Table 7.3 Descriptive statistics

Variables		**By activity type**			**Statistics for all leisure activity cases**				
		LOH	**LH**	**LPC**	**Min**	**Max**	**Med.**	**Mean**	**SD**
Socio-demographic variable:									
Gender (male = 1; otherwise = 0)		0.55	0.46	0.48	0.00	1.00	0.00	0.47	0.50
Metropolitan status (metropolitan = 1; otherwise = 0)		0.90	0.82	0.87	0.00	1.00	1.00	0.83	0.37
Working status (full time = 1; otherwise = 0)		0.75	0.33	0.41	0.00	1.00	0.00	0.37	0.48
No. of people in household		2.24	1.90	2.37	1.00	10.00	2.00	1.97	1.25
Teen (Age<18)		0.00	0.03	0.09	0.00	1.00	0.00	0.03	0.17
Young (18≤Age<25)		0.07	0.04	0.08	0.00	1.00	0.00	0.05	0.21
Adult (25≤Age<46)		0.44	0.19	0.25	0.00	1.00	0.00	0.21	0.41
Senior (46≤Age<65)		0.39	0.36	0.40	0.00	1.00	0.00	0.37	0.48
Older (Age≥65)		0.10	0.38	0.18	0.00	1.00	0.00	0.34	0.47
Household income ($100,000)		0.60	0.52	0.66	0.12	1.50	0.38	0.54	0.39
No. of children in Household		0.58	0.32	0.42	0.00	7.00	0.00	0.35	0.79
Household type	1.Married,	0.32	0.29	0.39	0.00	1.00	0.00	0.30	0.46
	2.Single,	0.37	0.50	0.29	0.00	1.00	0.00	0.47	0.50
Education	1. Bachelor	0.13	0.09	0.11	0.00	1.00	0.00	0.09	0.29
	2. Higher than Bachelor	0.06	0.09	0.09	0.00	1.00	0.00	0.09	0.28
	3. Others	0.81	0.82	0.80	0.00	1.00	1.00	0.82	0.38
Life status variable:									
Level of happiness during the activity (0-4)		2.85	2.84	2.80	0.00	4.00	3.00	2.84	0.91
Level of tired during the activity (0-6)		2.04	2.45	2.13	0.00	6.00	2.00	2.39	2.04
Meaningful of the activity (0-6)		4.14	3.59	3.94	0.00	6.00	4.00	3.66	2.04
Health status 1-5 (best=1; worst=5)		2.49	2.78	2.51	1.00	5.00	3.00	2.74	1.12
Satisfaction of Life 0-10 (best=10; worst=0)		6.80	6.78	6.78	0.00	10.00	7.00	6.78	2.29

Time use variable:								
Time related to working (hrs.)	6.54	2.22	2.96	0.00	23.08	0.00	2.59	4.02
Time related to shopping (hrs.)	0.27	0.26	0.23	0.00	14.00	0.00	0.26	0.66

7.4.2 Model Estimation Results

Estimation results are reported in Table 7.4, where we present model estimates of the integrated time use, well-being and activity type choice model respectively together with initial and final log-likelihood. Individual and household socio-demographic variable, time usage on other activities, and related well-being status enter the final specification.

Table 7.4 Integrated model: estimation results

	Time Usage Sub-model	
Variables	**Coefficient**	***t-value***
Constant	1.14	5.54
Time spent working	-0.09	-8.42
Time spent on shopping	-0.24	-3.98
Income	-0.02	-2.31
Number of children in household	-0.28	-4.86
Latent variable Z^*	0.79	6.06
	Well-being Sub-model	
	Coefficient	***t-value***
Number of people in household	-0.15	22.24
Marital status: single (dummy)	-0.3	583.28
Number of children in household	-0.02	-4.25
Interaction during activity	0.2	20.46
Adult (dummy)	-0.04	2.00
γ_0	0	Fixed

γ_1	1.15	54.13
γ_2	2.29	23.94

	Activity Type Choice Sub-model			
	LH		**LPC**	
	Coefficient	***t-value***	**Coefficient**	***t-value***
Constant	2.89	8.38	0.91	1.48
Teen (dummy)			1.30	1.79
Young (dummy)	-0.98	-2.49		
Adult (dummy)	-0.81	-2.29		
Senior (dummy)	-0.60	-2.12		
Working status (full time)	-1.18	-6.47	-1.08	-4.25
Marital status: single (dummy)			-0.46	-1.76
Log-likelihood (*0*)	-44893.59			
Log-likelihood(*Final*)	-5587.02			
*Adjusted Pseudo R-*squared	0.88			
No. of observations	1506			

Based on the estimation results obtained from the well-being sub model, it can be observed that several individual and household socio-demographic characteristics contribute to explain the level of happiness experienced by individuals during leisure activities. Being single has the most negative effect on the well-being during leisure activities, followed by the increasing number of people living in the household, the dummy variable for being adult and the number of children in the household. Interaction with other people is found to have a positive and significant effect on well-being. These results seem to be consistent with results obtained in the field of psychology. Oishi et al. (1999) noted that more frequent social interactivity has a positive influence on individuals' daily satisfaction, which lead to higher levels of well-being. Our results indicate that if interactions occur between respondents and other people during leisure activity, respondents are more likely to be happier. The integrated model offers the possibility to quantify the effect of emotions on the leisure activity duration through the latent variable construct. The estimate of the latent emotion variable is positive and significant, which indicates that individuals who feel happier during the leisure activity are more likely to spend more time on that activity.

In travel behaviour analysis, it is widely acknowledged that the location of daily activities, their type and duration play an important role on the way individuals plan their agenda and spend their time. Thereby they are all expected to have an effect on the overall quality of life (Spenney et al. 2009; Kitamura et al. 1997). Results obtained from the time use sub model indicate that, as individual's activities are constrained by time, time spent on working and shopping activities have negative impacts on leisure activity duration. Moreover, it is interesting to see that the negative effect of shopping is stronger than the one related to working (-0.24 compared to -0.09). Individuals in households with more children and households with higher income spend less time on leisure. The results from the activity choice sub model reveal that teens are more likely to choose computer related leisure activities comparing to out-of-home leisure activity. While respondents who are single are less likely to participate in computer related leisure activities. Full time workers are less likely to participate in both types of leisure activities during weekdays.

The variance-covariance matrix (Table 7.5) estimated for the joint model captures relations between the three decision variables considered. For the reason that emotions are also affected by activities pre and post chosen leisure activity, such influence could also lead to bias when estimating correlation between chosen activity participation and emotion indicators. Thus, we restrict this correlation to be zero and only try to capture the interrelation between emotions and time use. The values obtained indicate a statistically significant correlation exist among unobserved factors in preferences for activity participation, time use and level of happiness. High correlations exist between LH and LPC, which is expected that these two activities are belong to the same activity type. It is interesting to find a weak but positive covariance between time use and happiness, which indicate that individuals have a stronger tendency to engage in leisure activity of longer duration that lead to higher level of happiness.

Table 7.5 Difference of variance-covariance matrix (t statistic in parenthesis)

$$\Sigma = \begin{array}{c} \\ LH \\ \\ LPC \\ \\ I \\ \\ Time\ use \\ \\ \end{array} \begin{array}{cccc} LH & LPC & I & Time\ use \\ \end{array}$$

$$\Sigma = \begin{array}{c} \\ LH \\ LPC \\ I \\ Time\ use \end{array}\!\!\begin{array}{c} \begin{array}{cccc} LH & LPC & I & Time\ use \end{array} \\ \begin{pmatrix} \begin{matrix}1.00\\(fixed)\end{matrix} & \begin{matrix}0.75\\(3.88)\end{matrix} & \begin{matrix}0.00\\(fixed)\end{matrix} & \begin{matrix}0.69\\(4.77)\end{matrix} \\ \begin{matrix}0.75\\(3.88)\end{matrix} & \begin{matrix}6.46\\(2.13)\end{matrix} & \begin{matrix}0.00\\(fixed)\end{matrix} & \begin{matrix}0.28\\(1.12)\end{matrix} \\ \begin{matrix}0.00\\(fixed)\end{matrix} & \begin{matrix}0.00\\(fixed)\end{matrix} & \begin{matrix}0.79\\(12.97)\end{matrix} & \begin{matrix}0.08\\(0.25)\end{matrix} \\ \begin{matrix}0.69\\(4.77)\end{matrix} & \begin{matrix}0.28\\(1.12)\end{matrix} & \begin{matrix}0.08\\(0.25)\end{matrix} & \begin{matrix}2.38\\(51.35)\end{matrix} \end{pmatrix} \end{array}$$

Based on the estimation results, it is possible to conclude that during weekdays, the amount of time spent on leisure is positively correlated to well-being and that individuals who have higher level of happiness that individuals would like to spend more time on that activity when they feel happy. Besides that, working status also plays an important role in individuals' decision process. In particular, full-time workers and individuals who spend more time at work are less happy and less likely to participate in leisure activities. A substitution effect is found between leisure and shopping; the time spent on shopping have a stronger negative effect than working on time dedicated to leisure during weekdays.

7.5 Summary

Exploring the emotions behind time use and activity participation helps transportation analysts to better understand the hidden schemes that influence individual's decision-making process, which is of paramount importance in estimation activity-based travel demand. The Chapter proposed a model that assesses the impact of emotions on activity participation and time use decisions. Interdependence among the three sub models are captured through a variance covariance matrix. The framework further expands previous hybrid choice model by proposing an iterative procedure in the modelling estimation which handles the bias caused by endogeneity of emotions on time use decisions. Furthermore, our proposed approach easily accommodates psychological ordinal indicators for the latent variables into linear model when simultaneous effects are presented in both models.

A simulation study is developed in the Chapter to test the performance of the model. Estimators in the model are consistent and unbiased. Later, a real case study is also applied based on data extracted from the American Time Use Survey and its associated well-being module. We found that positive emotions (happiness) have positive effects on leisure activity time use decisions. At the same time use decisions are also positively correlated with leisure activity participations. Moreover, the empirical results provide valuable insights into the determinants of activity choice and time use decisions of individuals, such as household and individual demographics, and time spent on other activities. The presence of children in the household decrease the likelihood of spending more time and being happier during leisure activities. Young, adult, and senior groups are less likely to participate in in-home leisure activity when compared with out of home leisure activity when teens are more likely to choose computer related leisure activity. In general, adults perform less happy during leisure. Results also attest that interaction during activities could positively influence the level of happiness.

In closing, several future research avenues are possible based on this modelling formulation and estimation process. The study only considers the condition when endogenous variables are discrete and formulate the relation between continuous independent variable and latent psychological variables. Simultaneous equations between activity participation and emotions could be built in the future to enable a further exploration of the endogeneity between activity participation decisions and emotions. Enrichment of the model specification could also be considered in the future, such as including more structural equations to accommodate more psychological indicators or expanding the study to explore relationship between daily activity participations and overall subjective well-being. Finally, the same proposed model structure and estimation method can be applied to model a wide range of studies in travel behaviour and marketing research as well.

Chapter 8: Integrated Dynamic Activity Scheduling Model

8.1 Problem Description

Models for travel demand have evolved from aggregated trip and tour-based models to activity-based models where activity participation is explicitly estimated on individual travel diaries. All the existing operational activity-based models are developed for a typical weekday travel schedule. However, understanding dynamic processes in travel behaviour, and in particular rhythms and routines, (Axhausen, 2002) requires the observations and the modelling of activity-travel behaviour over longer time periods. The decision to model a typical-day activity-travel is mainly caused by insufficient data to calibrate/validate long time period models and by the high computational requirements of multi-day/multi-week models.

Debate has occurred concerning the appropriate modelling techniques that can capture dynamics in activity-travel behaviour. The traditional approach applies discrete choice models to activity schedule decisions; past studies formulate the problem in a static manner using MNL models (Alder and Ben-Akiva, 1979) or a NL models (Bowman and Ben-Akiva, 2001). Cirillo and Axhausen (2010) present a tour-based nested-logit model for within-day planning and propose different mixed-logit specifications to capture day-to-day variabilities in activity patterns. Arentze and Timmermans (2009) introduced a need-based model for activity scheduling. The utilities of choosing activities depend on the need for other activities. Needs evolve over time depending on the needs of certain activities. Transitions happen when activities are needed.

The goal of Chapter 8 is to develop an integrated dynamic discrete choice model that accounts for dynamics in activity scheduling over time. Following previous studies, activity patterns are divided into tours and stops. Tour choice sub-model and stop choice sub-model are jointly modelled in a hierarchical structure, where stop choices are assumed to be conditional on the tour choices and the related time constraints.

8.2 Activity Scheduling Modelling framework

Activity-based demand models decompose activity chains into tours and stops. A tour is a series of trip beginning and ending at home or at the work location; the elements needed to identify a tour are tour origins and destinations, and intermediate stops. In this study we assume that decisions on having a stop within a tour are conditional on tour choices. Driven by this consideration, the model proposed is formed by tour type sub-models and stop choice sub-models according to the hierarchical structure proposed in Figure 8.1.

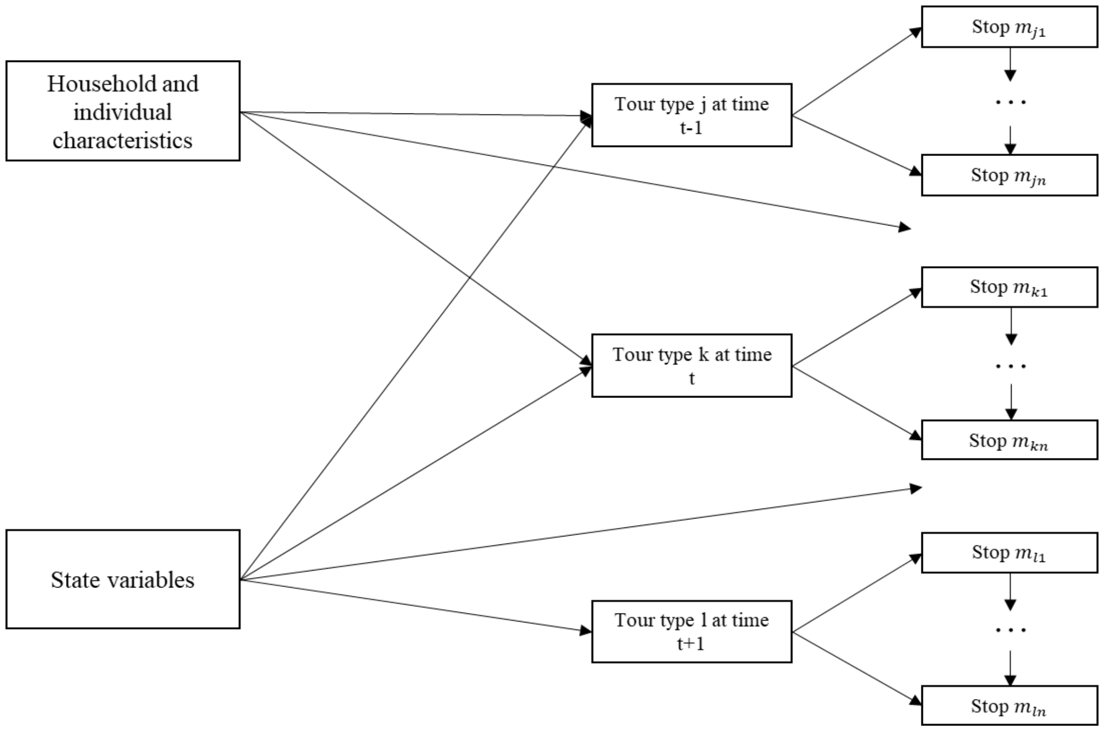

Figure 8.1 Structure of modelling framework

8.2.1 Tour Choice Model

In this chapter Markov decision processes (MDPs) are introduced to support the mathematical framework needed to model tours and stops formation. MDPs model sequence of possible events in which the probability of each event depends only on the state attained in the previous event. In particular, MDP is used to model individual's tour choices, in which we define tours as a sequence of events where the origin of the current tour is the destination of the previous tour. The model is developed based on two main assumptions:

- At time t, individuals' decisions about the current tour $t + 1$ depend on the destination of the previous tour s_t,

  $P(S_{t+1} = s_{t+1}|S_1 = s_1,\ S_2 = s_2, \dots, S_t = s_t) = P(S_{t+1} = s_{t+1}|S_t = s_t)$

- For any give states s_{t+1} and s_t, the transition probability from origin s_t to destination s_{t+1} is stationary. $P(S_{t+1} = s_{t+1}|S_t = s_t)$ is same $\forall\ t \geq 0$.

Tours are formed according to the type of origin s_t and destination s_{t+1}. The probability of traveller *i*'s tour choice at time *t* could be presented with the state transition probability from origins to destinations, $\rho_{it,t+1} = \begin{bmatrix} \rho_{it,t+1}^{(H,H)} & \rho_{it,t+1}^{(H,W)} \\ \rho_{it,t+1}^{(W,H)} & \rho_{it,t+1}^{(W,W)} \end{bmatrix}$.

The probability of choosing certain tour is conditional on the origin of the tour and on individual's socio-demographics. The probability, $\rho_i(s_t, s_{t+1})$, is formulated with a multinomial logit model that

$$\rho_i(s_t, s_{t+1}) = \frac{e^{U_{s_t,s_{t+1}}}}{\sum_{s_t \epsilon (H,W)} e^{U_{s_t,s_{t+1}}}}, \quad \forall s_{t+1} \epsilon (H, W) \tag{8.1}$$

Given that individual's tour choices are influenced by socio-demographic variables and state variables, we assume that the utility of choosing tour *j* is expressed as follows:

$$U_{s_t,s_{t+1}} = x_i\theta_i + \boldsymbol{y}_{it}\lambda_i + \boldsymbol{\epsilon_{it}} \tag{8.2}$$

where

x_i denotes the personal/household socio-demographic covariates. θ_i is the corresponding coefficients.

$\boldsymbol{y_{it}}$ is a random vector of dynamic attributes conditioned on time *t*. Time constraints and times of day related attributes are the dynamic variables to be included in the study. λ_i is a vector of parameters related to $\boldsymbol{y}_{it}$.

$\boldsymbol{\epsilon}_{it}$ is an individual-specific random term, whose components are independently and identically generalized extreme value (GEV) distributed among individuals and periods. We assume $\boldsymbol{\epsilon}_{it}$ to be independent of y_{it}.

8.2.2 Stop/Activity Choice Model with DDCM

Activity choices are not simply random. Rather, they are part of a deliberate process in which individuals trade off the worth of one alternative activity course of action versus another and chooses the alternative that is most likely to maximize his or her welfare. Therefore, individual's decisions on activity schedule consist of a sequence of actions that maximize his or her overall welfare. Factors that influence individual decisions are changing over time; the model proposed aims at including those factors into dynamic models of individuals' decisions over a limited time horizon.

Contrary to the dynamic choice model proposed in Xu et al. (2015), the stop choice model is formulated as an optimal stopping model. Consider a set of travellers i={1,…,I} and time periods t=0,1,…, T, conditional on the chosen tour type j, traveler i has two choices at time t:

- Stay in the tour and choose a stop m other than destination. Then obtain a one-period payoff, c_{it}.
- End current tour and obtain a final payoff b_{it} ;

Contrary to the static model that can be found in most previous studies, decision makers make a sequence of decisions subjected to time constraints, time of day, and perceptions of the consequences from future activity decisions.

We assume that the stop choices are consistent at each time point t. The payoff is expressed as a random utility function as follows

$$\boldsymbol{b}_{\boldsymbol{it}} = u\left(x_i, \boldsymbol{y}_{\boldsymbol{it}}, \theta_i, \lambda_i, \epsilon_{\boldsymbol{it}}\right), \tag{8.3}$$

where bold font indicates random variables.

Specifically, if individual i decides to choose a stop *m* other than to end the tour at a given destination at time t, stop type choice is estimated by a multinomial logit model with an error component of type I extreme value distribution. Correspondingly, for each individual i, $\boldsymbol{v}_{\boldsymbol{it}} = max\ b_{imt}$ follows EV I distribution with cumulative (F_v) and probability density function (f_v) as follows:

$$F_v(u, r_{it}) = exp(-e^{-(u-r_{it})}) \tag{8.4}$$

$$f_v(u, r_{it}) = e^{r_{it}} exp(-e^{-(u-r_{it})} - u) \tag{8.5}$$

Where r_{it} is the mode of this distribution, here we denote that r_{it} to be presented as follows:

$$r_{it} = \mathrm{E}_t[v_{it}], \tag{8.6}$$

where $\mathrm{E}_t[*]$ is the expect payoff of the tour. We consider r_{it} because it is a scalar-valued sufficient statistic for the distribution of future payoffs (Melnikov, 2013), and it contains the information available to the individual i at time t.

Then we describe the decision process as: at each decision-making point *t*, traveller *i* decides whether to end current tour or choosing an intermediate stop that maximize utility (Eqn. 8.3). Denoting the time the traveller decides to end the tour by τ, the optimization problem can be formulated as:

$$D_t\left(\boldsymbol{b_{it}}, c_{it}\right) = \max_{\boldsymbol{\tau}}\left\{\sum_{k=t}^{\boldsymbol{\tau}-1}\beta^{k-t}c_{it} + \beta^{\boldsymbol{\tau}-t}E_{y_\tau}\left[max\,\boldsymbol{b_{i\tau}} \mid y_{it}\right]\right\}, \tag{8.7}$$

where

β is a discount factor in [0,1];

c_{it} is the payoff function of individual i's attributes and the characteristics of traveler i when choose a type of stop, as defined above.

It is important to note that the expectation in Eqn. (8.7) is taken with respect to the state variable $\boldsymbol{y_{it}}$, ρ_i. While D_t remains a random function because of the terms $\boldsymbol{\epsilon_{it}}$ present in the random utility functions. We can rewrite Eqn. (8.7) as

$$D_t(\boldsymbol{v_{it}}, c_{it}) = \max_{\boldsymbol{\tau}}\left\{\sum_{k=t}^{\boldsymbol{\tau}-1}\beta^{k-t}c_{it} + \ \beta^{\boldsymbol{\tau}-t}E_{\widehat{\boldsymbol{T}}_{\tau}}[\boldsymbol{v_{it}}]\right\} \tag{8.8}$$

According to the previously described assumption about $\boldsymbol{\epsilon_{it}}$, $\boldsymbol{v_{it}}$ is Gumbel distributed with a scale factor equal to 1 and r_{it} is the mode of distribution of v_t.We also stress that if $\boldsymbol{\tau} = t$, the right-hand term in Eqn. (8.7) reduces to $\boldsymbol{v_{it}}$. The equation can be transformed from Eqn. (8.7) into

$$D_t(\boldsymbol{v_{it}}, c_{it}) = \max_{\boldsymbol{\tau}}\left\{\boldsymbol{v_{it}}, c_{it} + \ \beta E_{\widehat{\boldsymbol{T}}_{\tau}}[\boldsymbol{v_{it}}]\right\} \tag{8.9}$$

Based on Eqn. (8.9), we see that the decision process simply consists of ending current tour at time t, or keep in current state over one period, taking the payoff c_{it} plus the discounted future return. This is a standard optimal stopping problem, with a stopping set given by

$$\Gamma\left(\hat{T}_t\right)=\left\{\boldsymbol{v}_{\boldsymbol{it}} \mid \boldsymbol{v}_{\boldsymbol{it}} \rangle \boldsymbol{W}_{\boldsymbol{i}}\left(y_t\right)\right\} \tag{8.10}$$

where $\boldsymbol{W}_i(y_t \mid \rho_{jt})$ is the reservation utility level for driver i, which is defined as

$$\boldsymbol{W}_i\left(y_t\right)=c_{it}+\beta E_{y_{t+1}}\left[\boldsymbol{v}_{\boldsymbol{i},\boldsymbol{t}+1}, c_{i,t+1} \mid y_t\right]. \tag{8.11}$$

Using Eqn. (8.11), Eqn. (8.12) can be simplified as

$$D_t\left(\boldsymbol{v}_{\boldsymbol{it}}\right)=\max\left\{\boldsymbol{v}_{\boldsymbol{it}}, \boldsymbol{W}_i\left(y_t\right)\right\}. \tag{8.12}$$

The traveler *i* will end the tour at time t only when $\boldsymbol{v}_{\boldsymbol{it}} > \boldsymbol{W}_i(y_t)$. If i is randomly drawn from the population, the analyst can compute the probability of staying in current tour and choose stop *m* from the vectors of stops *M* as

$$\begin{aligned}
\pi_{imt}(y_t) &\stackrel{\text{def}}{=} P_{it}[D_t(v_{it}) = \boldsymbol{W}_i(y_t) | y_t] \\
&= P_{it}[\boldsymbol{v}_{\boldsymbol{it}} \leq \boldsymbol{W}_{\boldsymbol{i}}(y_t)] \cdot Pr[\boldsymbol{u}_{\boldsymbol{imt}} \geq \boldsymbol{u}_{\boldsymbol{ikt}}, k \neq m] \\
&= F_v\big(\boldsymbol{W}_i(y_t)\big) \cdot Pr\big[\boldsymbol{u}_{\boldsymbol{ijt}} \geq \boldsymbol{u}_{\boldsymbol{ikt}}, k \neq j\big] \\
&= e^{-e^{-}(\boldsymbol{W}_i(y_t)-r_{it})} \cdot \frac{e^{U_m}}{\sum_{m\in M} e^{U_m}}.
\end{aligned} \tag{8.13}$$

The probability of ending current tour is $\pi_{i0t} = 1 - \sum_{m\in M} \pi_{imt}(y_t)$.

8.2.3 Dynamic Estimation Process

We first summarized the parameters to be estimated in the dynamic problem:

- θ, a vector of parameters related to traveler i's attributes x_{it};

- λ, a vector of parameters related to dynamic attributes at time t, y_{jt};

- β, the discount factor, set to 1 for simplicity.

We could write the joint likelihood function of choosing a sequence of stops in tour j as:

$$\Pr_l(\theta,\lambda,\beta) = \rho_{ijl}\prod_{t=1}^{T_l}\pi_{it}\left[D_t(\boldsymbol{v_{it}})|\theta,\lambda,\beta\right] \tag{8.14}$$

Then the likelihood function of activity schedule is written as:

$$\mathcal{L}(\theta,\lambda,\beta) = \prod_{l=1}^{L}\Pr_l(\theta,\lambda,\beta) = \prod_{l=1}^{L}\rho_{ijl}\left\{\prod_{t=1}^{T_j}\pi_{it}\left[D_t(\boldsymbol{v_{it}})|\theta,\lambda,\beta\right]\right\} \tag{8.15}$$

where the probability π_{it} are following the distribution of the variables $\boldsymbol{\epsilon_{ijt}}$, as in Eqn. (8.12) and Eqn. (8.13), given the values of the parameters, L is the number of tour performed by individuals, T_l is the number of decision points in tour *j*.

Maximum log-likelihood estimation method is used applied in the study. In the estimation process, the probability of π_{ijt} depends on the calculation of W_{it}, as Eqn. (8.11). The key step during the estimation process is to identify the expected utility, $E_{y_{t+1}}[\boldsymbol{v_{i,t+1}}, c_{i,t+1}|y_t]$. At each time, traveller is assumed to account for the possible time budget and transition matrix in future scenarios, which are characterized by the time constraints and transition probabilities changing over time. The time horizon is defined on a limited number of time periods. The decision process is formulated by means of a decision tree (see Figure 8.3). The following steps describe the procedure to calculate $\pi_{i0,0}$and the expectation $E_{\boldsymbol{y_1}}[D(\boldsymbol{v_{i1}}, c_{i1})|i]$, which is denoted by $E\ [D_{\mathbf{1}}]$ for simplification purpose.

Given that the duration of activities and travel time varies over time, the model assumes that individuals only make decisions one period ahead and conditional on each tour decision. Therefore, at time period t=0, the individual only makes decision depending on activity involvement at t=1. $E\ [D_{\mathbf{2}}] = 0$ since the individual knows nothing of time period t=2 when faced with the decisions at t= 0 due to uncertainty of time spent on previous activities.

The process of calculating $E\ [D_{\mathbf{1}}]$ is recursive with known utility at the end of the perspective horizon. $\boldsymbol{W}_i(y_0)$ can be obtained after calculating $E\ [D_{\mathbf{1}}]$. The steps repeated to calculate $\pi_{i0,1}$with the same assumption that driver schedule activity at next time period and $E\ [D_{\mathbf{3}}] = 0$, and so on for the rest of the estimations.

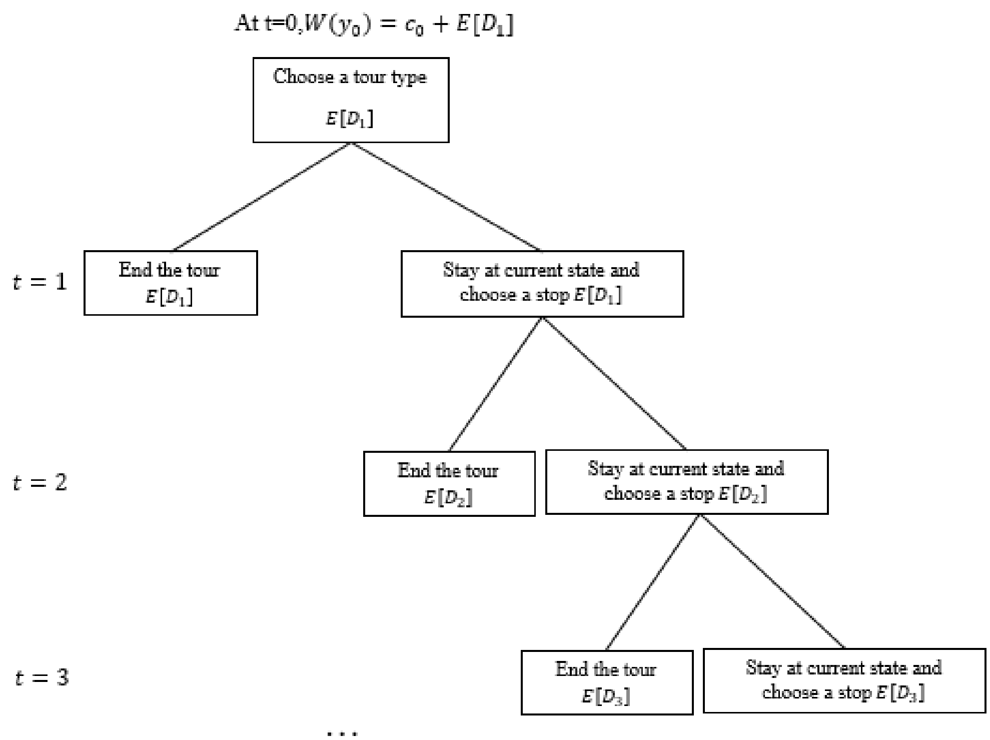

Figure 8.2 Decision tree

8.3 Data Descriptions

The data used for this study is from the "In the Moment" (ITM) Travel Survey. Activity schedules of respondents who completed a seven-day household travel survey are extracted from the data, together with socio-demographics, and mode choices. Activities are classified into three categories (Reichman, 1976): subsistence (work or work-related business), maintenance (grocery shopping, personal and household business, pick-up/drop-off passengers), and leisure (social and recreational purposes). Based on the category and possible duration of activities, six stop types are defined as follows:

- Home: trips end at home;

- Work/school: trip destination is related to work/school
- Maintenance: trip destination is related to maintenance (i.e. eating; shopping)
- Discretionary: trip destination is related to leisure, social, and recreation.
- Short stop: trip destination is related to drop off/pick up, drive through.
- Other: other trip types.

Among respondents completed survey, the average number of trips over seven days is 34.93. The average number of primary stops is 7.29, and intermediate stops is 19.23. The distribution of stops across days of week (Figure 8.3) shows that individuals prefer to perform maintenance trips mainly on Friday, or during weekends. Around 45% of discretionary stops are made during weekends.

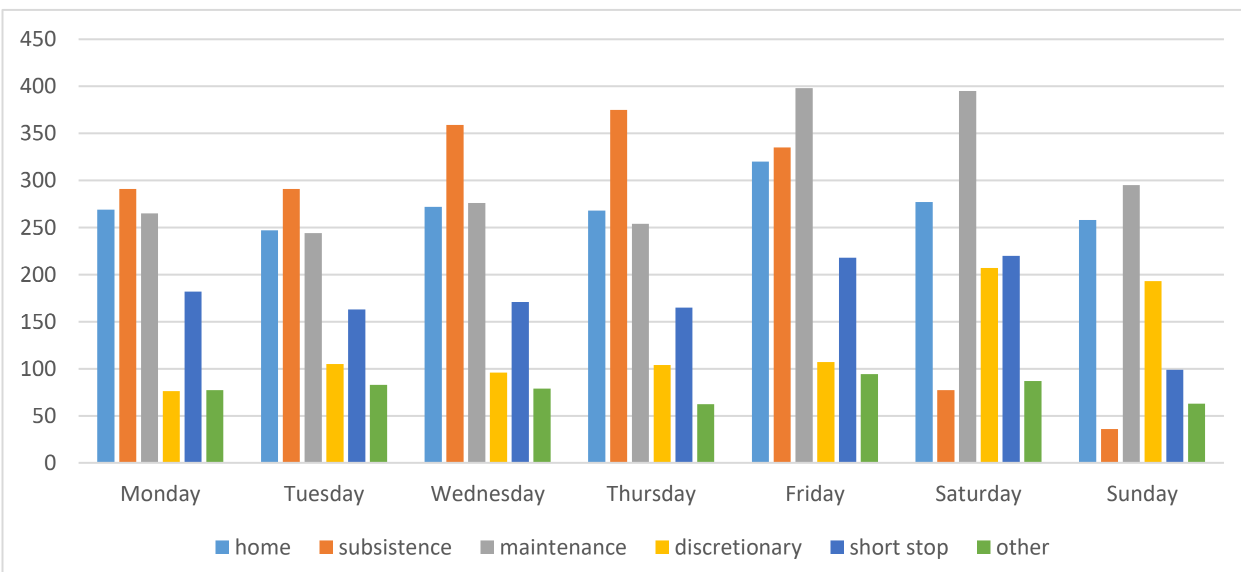

Figure 8.3 Distribution of stops cross seven days

Home can be the origin or the destination of a tour. Subsistence activities are defined as primary stops in the tour. The remaining activity types are intermediate. Tours are generated according to these stated activity types and trip sequences. Four tour types are defined as Home- Work/school (HW), Work/school -Home (WH), Work/school -Work/school (WW), and Home-Home (HH). The distribution of tour types across weekday is shown in Figure 8.4. We observe a higher frequency of HH tour type during weekends.

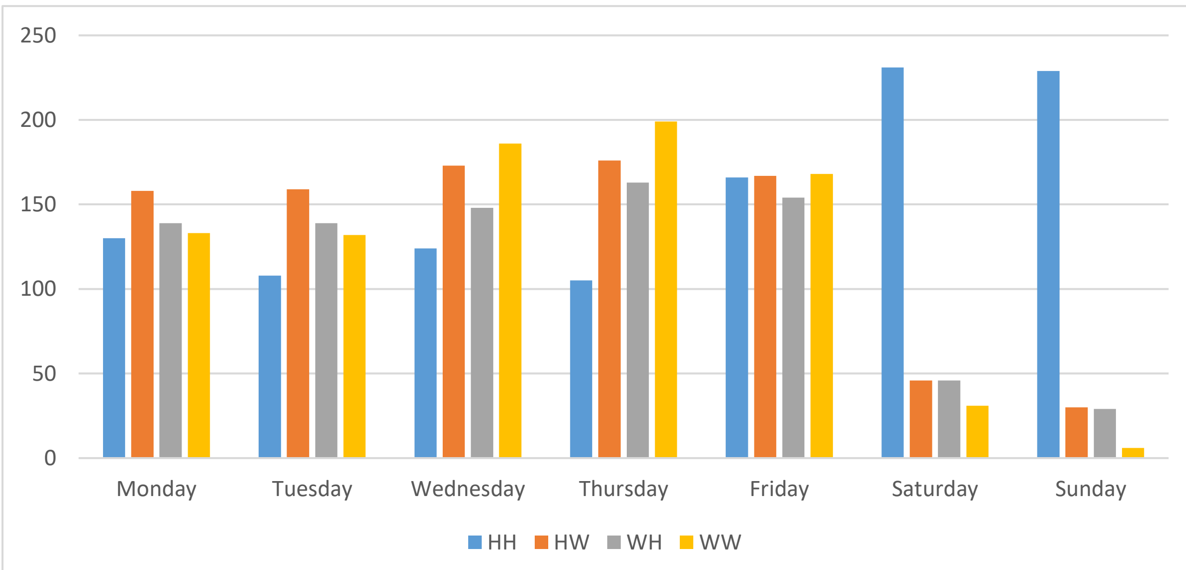

Figure 8.4 Distribution of tours cross seven days

To further explore the relation between tour types and trip types, we cross tabulate in Table 8.1 the type of activity performed in each tour type. Higher proportion of maintenance and discretionary trips occur in HH tours. Individuals also prefer to choose these two types of activities in WH tour rather than HW tour, which may be caused by time constraints occurring during the morning commuting tour.

Table 8.1 Distribution of stops by tour types

OD	Maintenance	Discretionary	Short stop	Others

HH	1282	642	649	312
HW	224	83	176	67
WH	346	101	219	105
WW	219	29	121	39

8.4 Model Estimation Results

In this section, we present results from estimations of the tour choice sub-model and the stop choice sub-model. *HH* in tour choice model and "*short stop*" in stop choice model are set as reference alternative. Constants in the utilities of these two alternatives are set to be zero. Most estimated coefficients are significant and of the expected sign. State variables, such as time left for the day, dummy variables of time of day fall into the final specification of the stop level choice model. In addition to the estimation results, pseudo R-squared value is also calculated based on the log-likelihood with only intercept and the final log-likelihood value, which is 0.39.

The estimation results of tour choice model show determinants of individuals' tour choices. HH is set to be the reference alternative and the relative constant is set be zero. The positive coefficient of "*PM peak hour*" shows that individuals are more likely to choose WH during PM peak. The positive coefficients of "*weekday"* show that individuals prefer to choose WH, HW, and WW during weekdays. Same positive effects are shown on coefficient "*Full-time working status*". The results show that individuals who have more children in household are more likely to choose HH tours. While students and middle age individuals are less likely to choose WW tours.

Stop choice model are conditional on tour choice model and the alternative "*destination*" is the destination of the chosen tour. The results indicate that individuals are less likely to end their tours when there is more time left until the end of the day. The negative coefficients of "*PM peak hour*" indicate that individuals are less likely to choose maintenance, discretionary, and other types of stops during PM peak. Due to time and monetary constraints, the results also indicate that low-income individuals are less likely to choose maintenance and discretionary stops; students are less likely to choose discretionary stops; full-time workers are less likely to choose short stops. The negative coefficients of "*No. of vehicle*" show that individuals with more vehicles in household are less likely to choose maintenance stops.

Table 8.2 Estimation results of integrated dynamic model

	Tour Choice Model			
Variables	**HH**	**HW**	**WH**	**WW**
Constant		-2.09*** (-14.87)	-2.46*** (-15.60)	-3.20*** (15.67)
PM peak hour			1.87*** (14.93)	
Night	-5.71*** (-3.93)			
Weekdays		1.79*** (14.17)	1.81*** (12.66)	2.89*** (15.29)
No. of children	0.13*** (3.77)			
Full-time working status		0.77*** (7.51)	0.69*** (6.27)	1.04*** (9.32)
Student				-0.81*** (-3.37)
Middle age				-0.42*** (-5.01)

	Stop Choice Model				
Variables	**Destination**	**Maintenance**	**Discretionary**	**Short stop**	**Others**
Constant		1.05*** (4.29)	0.72** (2.48)		0.63*** (2.92)
AM peak hour	2.28*** (28.20)			0.14 (0.67)	
PM peak hour	2.73*** (35.02)	-0.19 (-1.20)	-0.75*** (-3.78)		-0.15 (-0.89)
Night	3.07*** (35.64)	0.45*** (3.87)	0.36** (2.52)		

Time left for the day	0.01 (0.93)	0.04*** (3.22)	0.05*** (3.49)		0.01 (0.98)
Bachelor or higher degree					-0.22*** (-3.25)
Student			-0.34* (-1.89)		
Full-time working status				-0.23** (-2.40)	
Low income		-0.26*** (-2.63)	-0.34** (-2.54)		
No. of vehicle		-0.13*** (-3.00)	-0.33*** (-5.54)		
HH	-1.61*** (-15.36)				
HW	-1.19*** (-12.24)				
WH	-1.46*** (-12.6)				
WW	-0.56*** (-5.48)				
Log-likelihood (0)	-18038.82				
Log-likelihood (final)	-10771.12				
pseudo R-squared	0.40				
No. of observation	242				

Notes: *** $p < .01$, ** $p < .05$, * $p < .10$. Asymptotic standard errors in parentheses.

8.5 Summary

As important parts of activity-based modelling, tour level and stop level modelling components are traditionally modelled in a static manner and choice sets are pre-defined. A strong assumption in these models is that individuals' activity patterns must be defined in the choice set. However, individuals schedule their activities in a stochastic manner and they will adjust their decision according to the changes in the environment and time constraints. This section contributes to the last line of current studies by proposing a dynamic model framework for activity choice, where tour and stop choices are modelled in a hierarchical structure.

The results indicate that the integrated model is able to capture the effects of time constraints and socio-demographics on stop and tour choices. Individual's decision process is also investigated by incorporating the perception of future into the model. Results obtained are able to provide valuable insight to future studies. More specific, the results indicate that:

- On tour choices, No. of children contributes to the formation of HH tour. Full-time workers are willing to choose all work-related tours (HW, WH, and WW) during weekdays.

- On stop choices, individuals prefer to end current tour without any additional stops during peak hours. If more time is left, individuals are willing to choose stops other than the final destination. Low income individuals are less likely to choose maintenance and discretionary stops which may be caused by monetary constraints. Household with more vehicles also less likely to choose maintenance stops which may due to the fact that these activities are distributed across different household who have a vehicle available.

Chapter 9: Conclusions and Future Research

9.1 Conclusions and Research Contributions

Understanding the determinants of activities and travel is critical for transportation policy makers, planners, and engineers to design and manage transportation systems. These systems, and their externalities, are interwoven and consisted with human and communities. Discrete choice models, the predominant modelling tool for studying travel behaviour and transportation planning systems, are grounded on theories of individual decision-making. Emerging factors as technology innovations, human well-being, individual's perception of future uncertainty is investigated in this thesis, which enhances the behaviour realism of transportation models and leads to a better understanding of activity and travel behaviour.

The dissertation makes both theoretical and empirical contributions in the field of discrete choice analysis for activity-based modelling. In addition to the results presented, the proposed approaches could serve as a conceptual framework for future research, facilitating the identification of voids within the technical literature. Tools introduced are helpful for different transportation problems and will help decision makers implement more effective policies under the circumstance of innovations and human's recognition. The major contributions on theoretical that this dissertation devotes to the state of art includes:

- Performance of GEV type models and Probit model are investigated when negative correlations are present among activity choices. The results obtained from synthetic experiments and from a real case study indicate that both GEV type model and Probit model produce unbiased estimates of coefficients even when negative correlations exist among alternatives of activity choices. Whereas estimated correlations of GEV type model are still bias.

- Extending the discrete-continuous modelling framework to investigate the impacts of technology innovation on activity choice and time usage decisions. Interdependence between multiple activity choices and related time usage decisions are captured in the model through the covariance matrix. The revealed substitution effects between computer use and physical activities indicate that computer use should be included into future travel demand modelling system.

- Contributing to the existing literature on the relationship between well-being and activity choices, by offering a recursive modelling framework that investigates the simultaneous effects between emotions and activity choices. A Probit model and an order Probit model are jointly estimated. Average treatment effects and marginal effects are also computed to explore the impact of each emotion level on activity choices. The findings indicate that the impacts of emotion on individuals' activity choices vary across different emotion levels.

- An extended hybrid choice model is proposed to investigate the impacts of human emotions on activity choices and time usage decisions. Interdependence among emotion sub-model, activity choice sub-model, and time use sub-model are captured through a covariance matrix. Moreover, an iterative estimation method is proposed to overcome the bias caused by the endogeneity of emotions on activity-time decisions.

- A hierarchical dynamic activity scheduling model framework is proposed to explore the stochastic behaviour of activity-travel pattern. The framework integrates a Markov decision process and a dynamic discrete choice model to approach tour and stop making behaviour. Influences of individuals' perception on time changes are investigated as well as the household and individual characteristics.

The contributions of this dissertation towards empirical applications include:

- The models presented account for the participation in computer related activities, those are not present in current frameworks for travel demand modelling. Including this type of activities into the choice set is essential for city planning and policy analysis, and is especially important in the planning and modelling of smart city. Substitution of computer related activity on out-of-home and in-home leisure activities are quantified, which is not investigated in previous studies. The explored substitutions also give insight into the changes in activity and travel patterns that are caused by technology innovations.
- The consequence of ignoring negative correlations among activity choices when using GEV models are revealed. Failing to account for correlation across observations would result in the inability of GEV models to accurately capture modal shifts and eventually produce biased policy analysis results.

- Demonstrate that endogeneity exists when modelling emotions in activity choice and time use behaviour.
- Explain the substitution effects between in-home and out-of-home leisure activities and that these effects vary across levels of happiness, which is not included in previous studies.
- Explore the marginal effect of explanatory variables across different leisure activity involvement and level of happiness, which is not included in previous studies. The results could further be used for related policy analysis.
- The integrated scheduling models proposed simplify the structure of current activity based model. Joint structure of tour level model, stop level model, and time of day model reduces the number of models to be estimated for real case studies. The model captures tour choices and stop choices over multi-days and accounts for both within-day variations and day-to-day variations in activity scheduling.

9.2 Future Research

The dissertation proposes several innovative discrete choice approaches to enhance current travel demand modelling. However, modelling individuals' and households' behaviour is a complex task and transportation modelling system includes many modules, and the models proposed in this dissertation do not cover all the parts of the system. Real case studies in this dissertation are only showcases that test the performance of proposed methods. To empower travel demand models, the proposed models, especially those on Chapter 5, 6, and 7, could also be applied to activity choices other than just leisure activities. Although the work presented in this dissertation has expanded both the methodology and the empirical analysis related to activity choice and scheduling behaviour, the work can be further extended by considering the following aspects in survey design and choice modelling:

- *Travel survey design:* This work has shown that factors related to subjective well-being influence activity decision making process. However, subjective well-being is not captured in most of current survey and diary design. Collecting emotions related to subjective well-being could considerably enhance the general understanding of factors affecting travel behaviour and activity involvement.

- *Activity Diary Design:* The results of Chapter 5 indicate that substitution effects exist among computer use for leisure and physical leisure activities. However, most activity diaries don't have categories related to computer use. It is essential to include computer related activities in household travel survey and activity diary in order to accommodate models that accounts for the effects that technology has on daily schedules and travel.
- *Additional Indicators:* Previous literature has also investigated the impacts of satisfactions on mode choice and car ownership. Incorporating measures of satisfaction and other related indicators into activity travel behaviour modelling could improve the realism of model representation.

In the area of choice modelling and the application to activity travel demand modeling, future research could explore:

- *Application to Other Activity Types:* First and foremost, activity patterns are formed by various types of activities not limited to leisure activity. As mentioned in previous sections, it would be valuable to investigate the impacts of computer use on activities and time use patterns that include different types of activities other than just leisure activities. It is also worth to extend the model to investigate the relationship between emotion and other activity types.
- *Heterogeneity of Social Groups:* Individuals' activity patterns are heterogeneous across different groups. As an example, individuals belonging to different income groups, and in particular low income population, will have different activity patterns. This could be part of future investigations.
- *Exploring Other Components of Travel Demand Model:* Models developed in the dissertation only cover a limited number of components in the context of an Activity Based framework. More efforts should be directed towards the estimation of a complete system that includes vehicle ownership, mode and destination choice models.

- *Reinforcement Learning:* Individuals activity patterns might exhibit seasonal variability or may vary over the course of the individual life. Reinforcement learning model can also be developed to account for activity pattern recognition over time.

Bibliography

Abbe, E., Bierlaire, M., Toledo, T. (2007). "Normalization and correlation of Cross-Nested Logit models." Tranportation Research B, 41, 795-808.

Achen, C. H. (1986). The statistical analysis of quasi-experiments. Univ of California Press.

Aguiléra, A., Guillot, C., & Rallet, A. (2012). Mobile ICTs and physical mobility: Review and research agenda. Transportation Research Part A: Policy and Practice, 46(4), 664-672.

Akar, G., Clifton, K. J., & Doherty, S. T. (2011). Discretionary activity location choice: in-home or out-of-home? Transportation, 38(1), 101-122.

Akar, G., Clifton, K. J., & Doherty, S. T. (2012). Redefining activity types: Who participates in which leisure activity?. Transportation research part A: policy and practice, 46(8), 1194-1204.

Adler, T., & Ben-Akiva, M. (1979). A theoretical and empirical model of trip chaining behaviour. *Transportation Research Part B: Methodological*, *13*(3), 243-257.

Amemiya, T. (1978). The estimation of a simultaneous equation generalized Probit model. Econometrica: Journal of the Econometric Society, 1193-1205.

Archer, M., Paleti, R., Konduri, K., Pendyala, R., & Bhat, C. (2013). Modelling the connection between activity-travel patterns and subjective well-being. Transportation Research Record: Journal of the Transportation Research Board, (2382), 102-111.

Arentze, T., Hofman, F., van Mourik, H., & Timmermans, H. (2000). ALBATROSS: multiagent, rule-based model of activity pattern decisions. Transportation Research Record: Journal of the Transportation Research Board, (1706), 136-144.

Arentze, T. A., Ettema, D., & Timmermans, H. J. (2011). Estimating a model of dynamic activity generation based on one-day observations: method and results. *Transportation Research Part B: Methodological*, *45*(2), 447-460.

Arentze, T. A., & Timmermans, H. J. (2009). A need-based model of multi-day, multi-person activity generation. *Transportation Research Part B: Methodological*, *43*(2), 251-265.

Balepur, P. N., Varma, K. V., & Mokhtarian, P. L. (1998). Transportation impacts of center-based telecommuting: interim findings from the neighborhood telecenters project. Transportation, 25(3), 287-306.

Becker, G. S. (1965). A Theory of the Allocation of Time. The economic journal, 493-517.

Bekhor, S., Toledo, T., & Prashker. J. N. (2008). "Effects of choice set size and route choice models on path-based traffic assignment." Transportmetrica,4(2), 117-133.

Ben Akiva, M., Bierlaire, M. (1999). "Discrete choice methods and their applications to short term travel decisions." Handbook of Transportation Science, W. Hall, ed., Kluwer, Dordrecht.

Ben-Akiva, M., McFadden, D., Gärling, T., Gopinath, D., Walker, J., Bolduc, D., Börsch-Supan, A., Delquié, P., Larichev, O., Morikawa, T. and Polydoropoulou, A., 1999. Extended framework for modelling choice behavior. Marketing letters, 10(3), pp.187-203.

Ben Akiva, M., Bolduc, D. (1996). "Multinomial Probit with a Logit kernel and a general parametric specification of the covariance structure." Working paper, Massachusetts Institute of Technology.

Ben-Akiva, M., Francois, B. (1983). "μ homogeneous generalized extreme value model." Working paper, Department of Civil Engineering, Massechusets Institute of Technology, Cambridge, MA.

Ben-Akiva, M., McFadden, D., Train, K., Walker, J., Bhat, C., Bierlaire, M., ... & Daly, A. (2002). Hybrid choice models: progress and challenges. Marketing Letters, 13(3), 163-175.

Ben-Akiva, M., Walker, J., Bernardino, A. T., Gopinath, D. A., Morikawa, T., & Polydoropoulou, A. (2002). Integration of choice and latent variable models. Perpetual motion: Travel behaviour research opportunities and application challenges, 431-470.

Ben-Akiva, M., Abou-Zeid, M., & Choudhury, C. (2007, March). Hybrid choice models: from static to dynamic. In *Triennial Symposium on Transportation Analysis VI, Phuket, Thailand*

Ben-Akiva, M. (2010, January). Planning and action in a model of choice. In Choice Modelling: The State-of-the-Art and the State-of-Practice: Proceedings from the Inaugural International Choice Modelling Conference (pp. 19-34). Emerald Group Publishing Limited.

Ben-Elia, E., Alexander, B., Hubers, C., & Ettema, D. (2014). Activity fragmentation, ICT and travel: An exploratory Path Analysis of spatiotemporal interrelationships. Transportation Research Part A: Policy and Practice, 68, 56-74.

Bergstad, C. J., Gamble, A., Hagman, O., Polk, M., Gärling, T., Ettema, D., ... & Olsson, L. E. (2012). Influences of affect associated with routine out-of-home activities on subjective well-being. Applied Research in Quality of Life, 7(1), 49-62.

Bhat, C. R. (2005). A multiple discrete–continuous extreme value model: formulation and application to discretionary time-use decisions.Transportation Research Part B: Methodological, 39(8), 679-707.

Bhat, C. R. (2000). “Incorporating observed and unobserved heterogeneity in urban work mode choice modelling.” Transportation Science, 34, 228-238.

Bhat, C. R. (2001). “Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model.” Transportation Research B, 35, 677-693.

Bhat, C. R. (2003). “Simulation estimation of mixed discrete choice models using randomized and scrambled Halton sequences.” Transportation Research B, 37, 837-855.

Bhat, C. R. (2011). “The maximum approximate composite marginal likelihood (MACML) estimation of multinomial Probit-based unordered response choice models.” Transportation Research Part B: Methodological, 45(7), 923-939.

Bhat, C. R., & Dubey, S. K. (2014). A new estimation approach to integrate latent psychological constructs in choice modelling. Transportation Research Part B: Methodological, 67, 68-85.

Bhat, C.R., (2015). A new generalized heterogeneous data model (GHDM) to jointly model mixed types of dependent variables. Transportation Research Part B: Methodological, 79, pp.50-77.

Bhat, C.R., Pinjari, A.R., Dubey, S.K. and Hamdi, A.S., (2016). On accommodating spatial interactions in a generalized heterogeneous data model (GHDM) of mixed types of dependent variables. Transportation Research Part B: Methodological, 94, pp.240-263.

Bhat, C. R., & Gossen, R. (2004). A mixed multinomial logit model analysis of weekend recreational episode type choice. Transportation Research Part B: Methodological, 38(9), 767-787.

Bhat, C. R., & Koppelman, F. S. (1999). A retrospective and prospective survey of time-use research. Transportation, 26(2), 119-139.

Bhat, C., & Lockwood, A. (2004). On distinguishing between physically active and physically passive episodes and between travel and activity episodes: an analysis of weekend recreational participation in the San Francisco Bay area. Transportation Research Part A: Policy and Practice,38(8), 573-592.

Biddle, S. J., Fox, K., & Boutcher, S. (Eds.). (2003). Physical activity and psychological well-being. Routledge.

Bierlaire, M. (2003). “BIOGEME: a free package for the estimation of discrete choice models.” In Swiss Transport Research Conference (No. TRANSP-OR-CONF-2006-048). Monte-Verita, Ascona, Switzerland, March 19-21.

Bierlaire, M. (2006). "A theoretical analysis of the cross-nested logit model." Annals of operations research, 144, 287-300.

Bolduc, D., & Alvarez-Daziano, R. (2010, January). On estimation of hybrid choice models. In Choice Modelling: The State-of-the-Art and the State-of-Practice: Proceedings from the Inaugural International Choice Modelling Conference (pp. 259-287). Emerald Group Publishing Limited.

Bolduc, D., Boucher, N., & Alvarez-Daziano, R. (2008). Hybrid choice modelling of new technologies for car choice in Canada. Transportation Research Record: Journal of the Transportation Research Board, (2082), 63-71.

Börsch-Supan, A. (1990). “On the compatibility of nested logit models with utility maximization.” Journal of Econometrics, 43, 371-387.

Bowman, J. L., & Ben-Akiva, M. (1997). A travel demand model system for the information era. Transportation, 23, 241-266.

Bowman, J. L., & Ben-Akiva, M. E. (2001). Activity-based disaggregate travel demand model system with activity schedules. *Transportation research part a: policy and practice*, *35*(1), 1-28.

Boyd, J., Mellman, J. (1980). “The effect of fuel economy standards on the U.S. automotive market: A hedonic demand analysis.” Transportation Research A 14, 367–378.

Brownstone, D., Bunch, D., Train , K. (2000). “Joint mixed logit models of stated and revealed preferences for alternative-fuel vehicles.” Transportation Research B, 34, 315-338.

Bureau of Labor Statistics. (2014). 2013 American Time Use Survey [Data file and code book.] Retrieved from https://www.bls.gov/tus/datafiles_2013.htm

Cao, X., Douma, F., & Cleaveland, F. (2010). Influence of E-Shopping on Shopping Travel: Evidence from Minnesota's Twin Cities. Transportation Research Record: Journal of the Transportation Research Board, (2157), 147-154.

Cardell, S., Dunbar, F. (1980). “Measuring the societal impacts of automobile downsizing.” Transportation Research A, 14, 423-434.

Carrasco, J. A., & Miller, E. J. (2009). The social dimension in action: A multilevel, personal networks model of social activity frequency between individuals. Transportation Research Part A: Policy and Practice, 43(1), 90-104.

Carrasco, J.A. and Ortúzar, J. de D. (2002). “A review and assessment of the nested logit model.” Transport Reviews 22, 197–218.

Carranza, J. E. (2010). Product innovation and adoption in market equilibrium: The case of digital cameras. *International Journal of Industrial Organization*, *28*(6), 604-618.

Cascetta, E., Nuzzolo, A., Russo, F., & Vitetta, A. (1996). “A modified logit route choice model overcoming path overlapping problems: specification and some calibration results for interurban networks.” In Proceedings of the 13th International Symposium on Transportation and Traffic Theory (pp. 697-711). Lyon, France: Pergamon, July 24-26.

Cascetta, E., Papola, A., Russo, F. R. A. N. C. E. S. C. O., & Vitetta, A. N. T. O. N. I. N. O. (1999). “Implicit availability/perception logit models for route choice in transportation networks.” In World Transport Research: Selected Proceedings of the 8th World Conference on Transport Research (No. Volume 3).

Chaffey, D. (2016) Mobile Marketing Statistics compilation. http://www.smartinsights.com/mobile-marketing/mobile-marketing-analytics/mobile-marketing-statistics/?new=1. Accessed February, 2017

Chen, Y., Talebpour, A., & Mahmassani, H. S. (2015). Friends Don't Let Friends Drive on Bad Routes: Modelling the Impact of Social Networks on Drivers' Route Choice Behaviour. In Transportation Research Board 94th Annual Meeting (No. 15-4974).

Chu, C. (1989). "A paired combinatorial logit model for travel demand analysis." In Transport Policy, Management & Technology Towards 2001: Selected Proceedings of the Fifth World Conference on Transport Research (Vol. 4).

Cirillo, C., & Axhausen, K. W. (2010). Dynamic model of activity-type choice and scheduling. *Transportation*, *37*(1), 15-38.

Cirillo, C., Xu, R., & Bastin, F. (2015). A dynamic formulation for car ownership modelling. *Transportation Science*, *50*(1), 322-335.

Cirillo, C., Diana, M., & Dong, H. (2015). Time spent on social media and its impact on travel behaviours and activity participation: Preliminary results from a comparative study based on association analysis and discrete choice models.

Connors, R. D., Hess, S., & Daly, A. (2014). "Analytic approximations for computing Probit choice probabilities." Transportmetrica A: Transport Science, 10(2), 119-139.

Copperman, R. B., & Bhat, C. R. (2007). An analysis of the determinants of children's weekend physical activity participation. Transportation, 34(1), 67-87.

D'Andrea, E., Ducange, P., Lazzerini, B., & Marcelloni, F. (2015). Real-time detection of traffic from twitter stream analysis. IEEE Transactions on Intelligent Transportation Systems, 16(4), 2269-2283.

Daganzo, C. F. (2014). Multinomial Probit: the theory and its application to demand forecasting. Elsevier.

Daganzo, C. F., Sheffi, Y. (1977). "On stochastic models of traffic assignment." Transportation Science, 11, 253-274.

Daganzo, C.F. and Kusnic, M., (1993). “Technical Note—Two Properties of the Nested Logit Model.” Transportation Science, 27(4), pp.395-400.

Dagsvik, J. K. (1995). “How large is the class of generalized extreme value random utility models?” Journal of Mathematical Psychology, 39, 90-98.

Dal Fiore, F., Mokhtarian, P. L., Salomon, I., & Singer, M. E. (2014). “Nomads at last”? A set of perspectives on how mobile technology may affect travel. Journal of Transport Geography, 41, 97-106.

Daly, A.J. and Zachary, S. (1978). “Improved multiple choice models.” In D.A. Hensher and M.Q. Dalvi (eds.), Determinants of Travel Choice. Saxon House, Westmead.

Daziano, R. A., & Achtnicht, M. (2013). “Forecasting adoption of ultra-low-emission vehicles using Bayes estimates of a multinomial Probit model and the GHK simulator.” Transportation Science, 48(4), 671-683.

Daziano, R. A., & Bolduc, D. (2013). Incorporating pro-environmental preferences towards green automobile technologies through a Bayesian hybrid choice model. Transportmetrica A: Transport Science, 9(1), 74-106.

Daziano, R. A., & Bolduc, D. (2013). “Incorporating pro-environmental preferences towards green automobile technologies through a Bayesian hybrid choice model.” Transportmetrica A: Transport Science, 9(1), 74-106.

Dey, S. S., Thommana, J., & Dock, S. (2014). Public agency performance management for improved service delivery in the digital age: Case study. Journal of Management in Engineering, 31(5), 05014022.

de Lapparent, M., & Cernicchiaro, G. (2012). How long to own and how much to use a car? A dynamic discrete choice model to explain holding duration and driven mileage. *Economic Modelling*, *29*(5), 1737-1744.

Diener, E.D., Emmons, R.A., Larsen, R.J. and Griffin, S., 1985. The satisfaction with life scale. *Journal of personality assessment*, *49*(1), pp.71-75.

Diener, E. (2009). Subjective well-being. In The science of well-being (pp. 11-58). Springer Netherlands.

Diener, E., Oishi, S., & Lucas, R. E. (2003). Personality, culture, and subjective well-being: Emotional and cognitive evaluations of life. Annual review of psychology, 54(1), 403-425.

Doherty, S.T., Nemeth, E., Roorda, M. and Miller, E.J., (2004). Design and assessment of the Toronto area computerized household activity scheduling survey. *Transportation Research Record*, *1894*, pp.140-149.

Dolan, P., Peasgood, T., & White, M. (2008). Do we really know what makes us happy? A review of the economic literature on the factors associated with subjective well-being. Journal of economic psychology, 29(1), 94-122.

Dong, H., Cirillo, C., & Diana, M. Activity involvement and time spent on computers for leisure: an econometric analysis on the American Time Use Survey dataset. Transportation, 1-21.

Dong, H., Chen, Y. Y., Cirillo, C., & Wong, K. I. (2017). *A Lateral Movement Decision Model for Powered Two-Wheelers in a Developing Country* (No. 17-03581).

Dong, H., & Cirillo, C. (2018). *A Dynamic Approach Toward Stochastic Departure Time Choices Using a Linked Data Set*(No. 18-00621).

Douma, F., Wells, K., Horan, T. A., & Krizek, K. J. (2004). ICT and travel in the twin cities metropolitan area: enacted patterns between Internet use and working and shopping trips. In Proceedings CD-ROM of the 83rd Annual Meeting of the Transportation Research Board, Washington DC.

Ed, D., Suh, E. M., Lucas, R. E., & Smith, H. L. (1999). Subjective well-being: Three decades of progress. Psychological Bulletin, 125(2), 276-302.

Ellison, N. B., Steinfield, C., & Lampe, C. (2007). The benefits of Facebook "friends:" Social capital and college students' use of online social network sites. Journal of Computer-Mediated Communication, 12(4), 1143-1168.

Enam, A., Konduri, K. C., Eluru, N., & Ravulaparthy, S. (2015). Relationship Between Well-Being and Daily Time Use of Elderly: Evidence from Disabilities and Use of Time Survey. In Transportation Research Board 94th Annual Meeting (No. 15-3092).

Ettema, D., Friman, M., Gärling, T., & Olsson, L. E. (2016). Travel mode use, travel mode shift and subjective well-being: Overview of theories, empirical findings and policy implications. In Mobility, Sociability and Well-being of Urban Living (pp. 129-150). Springer Berlin Heidelberg.

Ettema, D., Friman, M., Gärling, T., Olsson, L. E., & Fujii, S. (2012). How in-vehicle activities affect work commuters' satisfaction with public transport. Journal of Transport Geography, 24, 215-222.

Ettema, D., Gärling, T., Olsson, L. E., & Friman, M. (2010). Out-of-home activities, daily travel, and subjective well-being. Transportation Research Part A: Policy and Practice, 44(9), 723-732.

Evans, A. W. (1972). On the theory of the valuation and allocation of time. Scottish Journal of Political Economy, 19(1), 1-17.

Farag, S., Schwanen, T., Dijst, M., & Faber, J. (2007). Shopping online and/or in-store? A structural equation model of the relationships between e-shopping and in-store shopping. Transportation Research Part A: Policy and Practice, 41(2), 125-141.

Farag, S., Weltevreden, J., Van Rietbergen, T., Dijst, M., & van Oort, F. (2006). E-shopping in the Netherlands: does geography matter?.Environment and Planning B: Planning and Design, 33(1), 59-74.

Ferrell, C. (2005). Home-based teleshopping and shopping travel: Where do people find the time? Transportation Research Record: Journal of the Transportation Research Board, (1926), 212-223.

Feil, M. (2010). Choosing the daily schedule: Expanding activity-based travel demand modelling (Doctoral dissertation, ETH).

Fosgerau, M., Frejinger, E., & Karlstrom, A. (2013). A link based network route choice model with unrestricted choice set. *Transportation Research Part B: Methodological*, *56*, 70-80.

Gal-Tzur, A., Grant-Muller, S. M., Kuflik, T., Minkov, E., Nocera, S., & Shoor, I. (2014). The potential of social media in delivering transport policy goals. Transport Policy, 32, 115-123.

Garikapati, V. M., Pendyala, R. M., Morris, E. A., Mokhtarian, P. L., & McDonald, N. (2016). Activity patterns, time use, and travel of millennials: A generation in transition? Transport Reviews, 36, 558-584. doi: 10.1080/01441647.2016.1197337

Gao, S., Frejinger, E., & Ben-Akiva, M. (2010). Adaptive route choices in risky traffic networks: A prospect theory approach. *Transportation research part C: emerging technologies*, *18*(5), 727-740.

Gkiotsalitis, K., Alesiani, F., & Baldessari, R. (2014). Educated rules for the prediction of human mobility patterns based on sparse social media and mobile phone data. In Transportation Research Board 93rd Annual Meeting (No. 14-0745).

Gliebe, J. P., Koppelman, F. S., & Ziliaskopoulos, A. (1999). “Route choice using a paired combinatorial logit model.” In 78th meeting of the Transportation Research Board, Washington DC, January 10-14

Godbey, G. C., Caldwell, L. L., Floyd, M., & Payne, L. L. (2005). Contributions of leisure studies and recreation and park management research to the active living agenda. American journal of preventive medicine, 28(2), 150-158.

Goldfarb, A., Greenstein, S. M., & Tucker, C. E. (Eds.). (2015). Economic Analysis of the Digital Economy. University of Chicago Press.

Goulias, K. G. (1999). Longitudinal analysis of activity and travel pattern dynamics using generalized mixed Markov latent class models. *Transportation Research Part B: Methodological*, *33*(8), 535-558.

Greene, W. H. (2008). Econometric Analysis 6th Edition, Pearson/Prentice Hall, Upper Saddle River, N.J.

Greenstein, S. M., & Tucker, C. E. (2015). Economic analysis of the digital economy. University of Chicago Press.

Gu, Y., Qian, Z. S., & Chen, F. (2016). From Twitter to detector: Real-time traffic incident detection using social media data. Transportation Research Part C: Emerging Technologies, 67, 321-342.

Guin, Angshuman., *Introduction to TRANSIMS*, March 2010, https://t-square.gatech.edu/access/content/group/28974.201002/Introduction_to_TRANSIMS_Part1.pdf

Habib, K.M.N., (2007). Modelling Activity Generation Processes. PhD Thesis, Department of Civil Engineering, University of Toronto.

Habib, K. M. N., Day, N., & Miller, E. J. (2009). An investigation of commuting trip timing and mode choice in the Greater Toronto Area: application of a joint discrete-continuous model. Transportation Research Part A: Policy and Practice, 43(7), 639-653.

Habib, K., Carrasco, J., & Miller, E. (2008). Social context of activity scheduling: discrete-continuous model of relationship between" with whom" and episode start time and duration. Transportation Research Record: Journal of the Transportation Research Board, (2076), 81-87.

Habib, K.M. and Miller, E.J., (2008). Modelling daily activity program generation considering within-day and day-to-day dynamics in activity-travel behaviour. *Transportation*, *35*(4), p.467.

Hamer, R., Kroes, E., & Van Ooststroom, H. (1991). Teleworking in the Netherlands: an evaluation of changes in travel behaviour. Transportation,18(4), 365-382.

Handy, S. L., & Mokhtarian, P. L. (1996). The future of telecommuting.Futures, 28(3), 227-240.

Harris, K. M., & Keane, M. P. (1998). A model of health plan choice:: Inferring preferences and perceptions from a combination of revealed preference and attitudinal data. *Journal of Econometrics*, *89*(1-2), 131-157.

Hasan, S., & Ukkusuri, S. V. (2014). Urban activity pattern classification using topic models from online geo-location data. Transportation Research Part C: Emerging Technologies, 44, 363-381.

Heckman, J. J. (1977). Dummy endogenous variables in a simultaneous equation system.

Heckman, J. J. (1981). The incidental parameters problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process.

Hensher, D., Greene, W. (2001). "The mixed logit model: The state of practice and warnings for the unwary." Working paper, School of Business, The University of Sydney.

Herriges, J., Kling, C. (1996). "Testing the consistency of nested logit models with utility maximization." Economic Letters, 50, 33–39.

Hess, S., Ryley, T., Davison, L., & Adler, T. (2013). "Improving the quality of demand forecasts through cross nested logit: a stated choice case study of airport, airline and access mode choice." Transportmetrica A: Transport Science, 9(4), 358-384.

Hjorthol, R., & Gripsrud, M. (2009). Home as a communication hub: the domestic use of ICT. Journal of Transport Geography, 17(2), 115-123.

Iwasaki, Y. (2007). Leisure and quality of life in an international and multicultural context: What are major pathways linking leisure to quality of life?. Social Indicators Research, 82(2), 233-264.

Kahneman, D., & Krueger, A. B. (2006). Developments in the measurement of subjective well-being. The journal of economic perspectives, 20(1), 3-24.

Kahneman, D., Diener, E., & Schwarz, N. (Eds.). (1999). Well-being: Foundations of hedonic psychology. Russell Sage Foundation.

Kamargianni, M., & Polydoropoulou, A. (2013). Hybrid choice model to investigate effects of teenagers' attitudes toward walking and cycling on mode choice behavior. *Transportation Research Record: Journal of the Transportation Research Board*, (2382), 151-161.

Kamargianni, M., Ben-Akiva, M., & Polydoropoulou, A. (2014). Incorporating social interaction into hybrid choice models. Transportation, 41(6), 1263-1285.

Kamargianni, M., Dubey, S., Polydoropoulou, A., & Bhat, C. (2015). Investigating the subjective and objective factors influencing teenagers' school travel mode choice–An integrated choice and latent variable model. *Transportation Research Part A: Policy and Practice*, *78*, 473-488.

Kapur, A., & Bhat, C. (2007). Modelling adults' weekend day-time use by activity purpose and accompaniment arrangement. Transportation Research Record: Journal of the Transportation Research Board, (2021), 18-27

Karaca-Mandic, P., & Train, K. (2003). "Standard error correction in two-stage estimation with nested samples." The Econometrics Journal, 6(2), 401-407.

Kemperman, A. D., & Timmermans, H. J. (2008). Influence of socio-demographics and residential environment on leisure activity participation.Leisure sciences, 30(4), 306-324.

Kim, J., Bae, Y., & Chung, J. H. (2012). Effects of personal proenvironmental attitudes on mode choice behaviour: new ecofriendly water transit system in Seoul, South Korea. Transportation Research Record: Journal of the Transportation Research Board, (2274), 175-183.

Kitamura, R., van der Hoorn, A. I. J. M., Wijk, F., Timmermans, H., & Ettema, D. (1997). A Comparative Analysis of Daily Time Use and the Development of an Activity-Based Traveller Benefit Measure. Activity-Based Approaches to Travel Analysis., 171-188.

Kitamura, R., Kikuchi, A., Fujii, S., & Yamamoto, T. (2005). An overview of PCATS/DEBNetS micro-simulation system: Its development, extension, and application to demand forecasting. In Simulation Approaches in Transportation Analysis (pp. 371-399). Springer, Boston, MA.

Kitamura, R., Yamamoto, T., Susilo, Y. O., & Axhausen, K. W. (2006). How routine is a routine? An analysis of the day-to-day variability in prism vertex location. Transportation Research Part A: Policy and Practice, 40(3), 259-279.

Kling, C., Herriges, J. (1995). "An empirical investigation of the consistency of nested logit models with utility maximization." American Journal of Agricultural Economics, 77, 875-884.

Koppelman, F., Wen, C. (2000). "The paired combination logit model: Properties, estimation and application." Transportation Research B, 34, 75-89.

Koppelman, F. S., & Hauser, J. R. (1978). Destination choice behaviour for non-grocery-shopping trips. *Transportation Research Record*, (673).

Le Vine, S., Latinopoulos, C., & Polak, J. (2016). Analysis of the relationship between internet usage and allocation of time for personal travel and out of home activities: Case study of Scotland in 2005/6. Travel Behaviour and Society, 4, 49-59.

Lee, J. H., Gao, S., & Goulias, K. G. (2016). Comparing the Origin-Destination Matrices from Travel Demand Model and Social Media Data. In Transportation Research Board 95th Annual Meeting (No. 16-0069).

Leung, L., & Lee, P. S. (2005). Multiple determinants of life quality: The roles of Internet activities, use of new media, social support, and leisure activities. Telematics and Informatics, 22(3), 161-180.

Lin, T., & Wang, D. (2015). Tradeoffs between in-and out-of-residential neighborhood locations for discretionary activities and time use: do social contexts matter? Journal of transport geography, 47, 119-127.

Liu, Y. (2013). *INTEGRATED DISCRETE CONTINUOUS CHOICE MODELS: THEORY AND APPLICATIONS IN ESTIMATING HOUSEHOLD VEHICLE OWNERSHIP, TYPE AND USAGE* (Doctoral dissertation).

Liu, Y., & Cirillo, C. (2015). Model System to Evaluate Impacts of Vehicle Purchase Tax and Fuel Tax on Household Greenhouse Gas Emissions.Transportation Research Record: Journal of the Transportation Research Board, (2503), 51-59.

Liu, Y., Tremblay, J. M., & Cirillo, C. (2014). An integrated model for discrete and continuous decisions with application to vehicle ownership, type and usage choices. Transportation Research Part A: Policy and Practice, 69, 315-328.

Lloyd, K. M., & Auld, C. J. (2002). The role of leisure in determining quality of life: Issues of content and measurement. Social Indicators Research, 57(1), 43-71.

Lorincz, S. (2005). *Persistence Effects in a Dynamic Discrete Choice Model-Application to Low-End Computer Servers* (No. MT-DP-2005/10). IEHAS Discussion Papers.

Luce, R. D., Raiffa, H. (1957). Games and Decisions, John Wiley & Sons, New York.

Mabit, S. L., Cherchi, E., Jensen, A. F., & Jordal-Jørgensen, J. (2014). Hybrid choice modelling allowing for reference-dependent preferences: estimation and validation for the case of alternative-fuel vehicles. In Transportation Research Board 93rd Annual Meeting (No. 14-0705).

Maddala, G. S. (1986). Disequilibrium, self-selection, and switching models. Handbook of econometrics, 3, 1633-1688.

Maghrebi, M., Abbasi, A., Rashidi, T. H., & Waller, S. T. (2015, September). Complementing Travel Diary Surveys with Twitter Data: Application of Text Mining Techniques on Activity Location, Type and Time. In 2015 IEEE 18th International Conference on Intelligent Transportation Systems (pp. 208-213). IEEE.

Maslow, A.H. (1970). *Motivation and Personality*. Second Edition, Harper & Row, New
York.

McCulloch, C. E., & Neuhaus, J. M. (2001). Generalized linear mixed models. John Wiley & Sons, Ltd.

McFadden, D. (1978). “Modelling the choice of residential location.” Spatial Interaction Theory and Residential Location, A. Karlqvist, L. Lundquist, F. Snickkars, and J. Weibull, eds., Amsterdam, 75-96.

McFadden, D., Train, K. (2000). “Mixed MNL models of discrete response.” Journal of Applied Econometrics, 15, 447-470.

McNally, M. G., & Rindt, C. (2007). The activity-based approach. Technical Report n. UCI-ITS-WP-07-1. D. Institute of Transportation Studies University of California, Irvine, U.S.A.

Meloni, I., Spissu, E., & Bez, M. (2007). A model of the dynamic process of time allocation to discretionary activities. Transportation Science, 41(1), 15-28.

Melnikov, O. (2000). Demand for differentiated durable products: The case of the US computer printer market. *Manuscript. Department of Economics, Yale University*.

Meyer, M. D., & Miller, E. J. (2001). Urban transportation planning: a decision-oriented approach.

Miller, E.J. (2005). Introduction. *Transportation*, 319-320

Miller, R. A. (1984). Job matching and occupational choice. *Journal of Political Economy*, *92*(6), 1086-1120.

Mokhtarian, P. L. (1991). Defining telecommuting. Research Report n. UCD-ITS-RR-91-04. Institute of Transportation Studies University of California, Davis, U.S.A.

Mokhtarian, P. L. (1997). The transportation impacts of telecommuting: recent empirical findings. Understanding travel behaviour in an era of change. Edited by P. Stopher and M.Lee Gosselin, 91-106. Oxford, Elsevier.

Mokhtarian, P. L. (1998). A synthetic approach to estimating the impacts of telecommuting on travel. Urban studies, 35(2), 215-241.

Mokhtarian, P. L. (2004). A conceptual analysis of the transportation impacts of B2C e-commerce. Transportation, 31(3), 257-284.

Mokhtarian, P. L., & Salomon, I. (1997). Modelling the desire to telecommute: The importance of attitudinal factors in behavioural models. Transportation Research Part A: Policy and Practice, 31(1), 35-50.

Mokhtarian, P. L., Handy, S. L., & Salomon, I. (1995). Methodological issues in the estimation of the travel, energy, and air quality impacts of telecommuting. Transportation Research Part A: Policy and Practice, 29(4), 283-302.

Mokhtarian, P. L., Salomon, I., & Handy, S. L. (2006). The impacts of ICT on leisure activities and travel: a conceptual exploration. Transportation,33(3), 263-289.

Mokhtarian, P.L. (2002) Telecommunications and travel: The case for Complementarity Journal of Industrial Ecology, 6(2), pp.43-57.

Mondschein, A. (2015). Five-star transportation: using online activity reviews to examine mode choice to non-work destinations. Transportation, 42(4), 707-722.

Munizaga, M. A., Heydecker, B. G., & Ortúzar, J. de D. (2000). “Representation of heteroskedasticity in discrete choice models.” Transportation Research Part B: Methodological, 34(3), 219-240

Myers, D. G., & Diener, E. (1995). Who is happy?. Psychological science, 6(1), 10-19.

Nielsen Holdings, N. V. (2012). State of the media: The social media report 2012. Retrieved May, 10, 2014.

O'Brien, R.M., (1994). Identification of simple measurement models with multiple latent variables and correlated errors. Sociological methodology, 24, pp.137-170.

Ohmori, N. (2008). Application of information on human activity-travel behaviour in urban space and time in the information age. Spatial Data Infrastructure for Urban Regeneration, 127-145.

Oishi, S., Diener, E. F., Lucas, R. E., & Suh, E. M. (1999). Cross-cultural variations in predictors of life satisfaction: Perspectives from needs and values. Personality and social psychology bulletin, 25(8), 980-990.

Olsson, L.E., Gärling, T., Ettema, D., Friman, M. and Fujii, S., (2013). Happiness and satisfaction with work commute. *Social indicators research*, *111*(1), pp.255-263.

Pakes, A. (1984). Patents as options: Some estimates of the value of holding European patent stocks.

Paleti, R., Copperman, R. B., & Bhat, C. R. (2011). An empirical analysis of children’s after school out-of-home activity-location engagement patterns and time allocation. Transportation, 38(2), 273-303.

Papola, A. (2004). “Some developments of the Cross- Nested Logit model.” Tranportation Research B, 38, 833-851.

Pavot, W. and Diener, E., (1993). Review of the satisfaction with life scale. *Psychological assessment*, *5*(2), p.164.

Pender, B., Currie, G., Delbosc, A., & Shiwakoti, N. (2014). Social media use during unplanned transit network disruptions: A review of literature. Transport Reviews, 34(4), 501-521.

Pendyala, R. M., Goulias, K. G., & Kitamura, R. (1991). Impact of telecommuting on spatial and temporal patterns of household travel.Transportation, 18(4), 383-409.

Pendyala, R. M., Kitamura, R., & Reddy, D. P. (1998). Application of an activity-based travel-demand model incorporating a rule-based algorithm. *Environment and Planning B: Planning and Design*, *25*(5), 753-772.

Pendyala, R. M., Yamamoto, T., & Kitamura, R. (2002). On the formulation of time-space prisms to model constraints on personal activity-travel engagement. Transportation, 29(1), 73-94

Prashker, J. N., Bekhor, S. (1998). “Investigation of stochastic network loading procedures.” Transportation Research Record , 1645, 94-102.

Pychyl, T. A., & Little, B. R. (1998). Dimensional specificity in the prediction of subjective well-being: Personal projects in pursuit of the PhD. *Social Indicators Research*, *45*(1-3), 423-473.

Ravulaparthy, S. K., Konduri, K. C., & Goulias, K. G. (2016). Fundamental Linkages Between Activity Time Use and Subjective Well-Being for the Elderly Population: Joint Exploratory Analysis Framework for In-Home and Out-of-Home Activities. Transportation Research Record: Journal of the Transportation Research Board, (2566), 31-40.

Ren, F., & Kwan, M. P. (2009). The impact of the Internet on human activity–travel patterns: analysis of gender differences using multi-group structural equation models. Journal of Transport Geography, 17(6), 440-450.

Revelt, D., Train, K. (2000). “Specific taste parameters and mixed logit.” Working paper, Department of Economics, University of California, Berkeley.

Rice, J. (2006). Mathematical statistics and data analysis. Cengage Learning.

Robinson, J. P., & Martin, S. (2008). What do happy people do?. Social Indicators Research, 89(3), 565-571.

Robinson, J. P., & Martin, S. (2010). IT use and declining social capital? More cold water from the General Social Survey (GSS) and the American Time-Use Survey (ATUS). Social Science Computer Review, 28(1), 45-63.

Rust, J. (1987). Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher. *Econometrica: Journal of the Econometric Society*, 999-1033.

Ryan, R. M., & Deci, E. L. (2001). On happiness and human potentials: A review of research on hedonic and eudaimonic well-being. Annual review of psychology, 52(1), 141-166.

Scanzoni, J. H., & Szinovacz, M. E. (1980). Family decision-making: a developmental sex role model.

Schlich, R. and Axhausen, K.W., (2003). Habitual travel behaviour: evidence from a six-week travel diary. *Transportation*, *30*(1), pp.13-36.

Schwanen, T., & Kwan, M. P. (2008). The Internet, mobile phone and space-time constraints. Geoforum, 39(3), 1362-1377.

Schwanen, T., Kwan, M. P., & Ren, F. (2008). How fixed is fixed? Gendered rigidity of space–time constraints and geographies of everyday activities.Geoforum, 39(6), 2109-2121.

Sener, I. N., Pendyala, R. M., & Bhat, C. R. (2011). “Accommodating spatial correlation across choice alternatives in discrete choice models: an application to modelling residential location choice behaviour.” Journal of Transport Geography, 19(2), 294-303.

Serulle, N. U., & Cirillo, C. (2017). The optimal time to evacuate: A behavioral dynamic model on Louisiana resident data. *Transportation research part B: methodological*, *106*, 447-463.

Shahhoseini, Z., Haghani, M., & Sarvi, M. (2015). “Estimation and application of a multi-class multi-criteria mixed paired combinatorial logit model for transport networks analysis.” Transportmetrica B: Transport Dynamics, 3(1), 59-78.

Spinney, J. E., Scott, D. M., & Newbold, K. B. (2009). Transport mobility benefits and quality of life: A time-use perspective of elderly Canadians. Transport Policy, 16(1), 1-11.

Srinivasan, S., & Bhat, C. R. (2008). An exploratory analysis of joint-activity participation characteristics using the American time use survey. Transportation, 35(3), 301-327.

Stutzer, A., & Frey, B. S. (2008). Stress that doesn't pay: The commuting paradox. The Scandinavian Journal of Economics, 110(2), 339-366.

Train, K. (1986). Qualitative choice analysis: Theory, econometrics, and an application to automobile demand (Vol. 10). MIT press.

Train, K. (1999). "Mixed logit models for recreation demand." Valuing Recreation and the Environment., J. Herriges and C. Kling, eds., Edward Elgar, Northampton, MA.

Train, K. (2000). "Halton sequences for mixed logit." Working paper, Department of Economics, University of California, Berkeley.

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

Train, K., D. McFadden, M. Ben- Akiva. (1987). "The demand for local telephone service: A fully discrete model of residential calling patterns and service choice." Rand Journal of Economics, 18, 109-123.

Tversky, A. (1972). "Elimination by aspects: A theory of choice." Psychological Review, 79, 281– 299.

Västfjäll, D. and Gärling, T., (2007). Validation of a Swedish short self-report measure of core affect. *Scandinavian Journal of Psychology*, *48*(3), pp.233-238.

Veenhof, B. (2006). The Internet: Is it changing the way Canadians spend their time? Statistics Canada

Visser, E. J., & Lanzendorf, M. (2004). Mobility and accessibility effects of b2c e-commerce: a literature review. Tijdschrift voor Economische en Sociale Geografie, 95, 189-205.

Von-Neumann, J., Morgenstern, O. (1944). Theory of Games and Economic Behaviour, Princeton University Press, Princeton.

Vovsha, P. (1997). “The cross-nested logit model: application to mode choice in the Tel- Aviv Metropolitan Area.” Transportation Research Record, 1607, 6-15.

Vovsha, P., Bekhor, S. (1998). “The link-nested logit model of route-choice: overcoming the route overlapping problem.” Transportation Research Record, 1645, 133-148.

Vovsha, P., & Bradley, M. (2004). Hybrid discrete choice departure-time and duration model for scheduling travel tours. *Transportation Research Record: Journal of the Transportation Research Board*, (1894), 46-56.

Walker, J. L. (2001). “Extended Discrete Choice Models: Integrated Framework, Flexible Error Structures, and Latent Variables.” Ph. Dthesis, Massachusetts Institute of Technology.

Walker, J. L., Ben Akiva, M., Bolduc, D. (2004). “Identification of the Logit Kernel (or Mixed Logit) Model.” working paper, Massachusetts Institute of Technology.

Wallsten, S. (2013). What are we not doing when we're online (No. w19549). National Bureau of Economic Research.

Wang, D., & Law, F. Y. T. (2007). Impacts of Information and Communication Technologies (ICT) on time use and travel behaviour: a structural equations analysis. Transportation, 34(4), 513-527.

Wearesocial. (2016). Digital in 2016. http://wearesocial.com/uk/special-reports/digital-in-2016 (2016). Accessed January, 26, 2016.

Weisberg, S. (2005). Applied linear regression (Vol. 528). John Wiley & Sons.

Weltevreden, J. W. (2007). Substitution or complementarity? How the Internet changes city centre shopping. Journal of Retailing and consumer Services, 14(3), 192-207.

Wen, C., Koppelman, F. (2001). “The generalized nested logit model.” Transportation Research B, 35, 627-641.

Williams, H.C.W.L. (1977). “On the formation of travel demand models and economic evaluation measures of user benefit.” Environment and Planning 9A, 285–344.

Williams, H.C.W.L. and Ortúzar, J. de D. (1982). “Behavioural theories of dispersion and the mis-specification of travel demand models.” Transportation Research 16B, 167–219.

Wilson, R., Krizek, K., & Handy, S. (2007). Trends in out-of-home and at-home activities: evidence from repeat cross-sectional surveys. Transportation Research Record, 2014, 76-84.

Wolpin, K. I. (1984). An estimable dynamic stochastic model of fertility and child mortality. *Journal of Political economy*, *92*(5), 852-874.

Yamamoto, Toshiyuki, Ryuichi Kitamura, and Ram M. Pendyala. (2004). "Comparative analysis of time-space prism vertices for out-of-home activity engagement on working and nonworking days." Environment and Planning B: Planning and Design 31.2: 235-250.

Zeelenberg, M., Nelissen, R. M., Breugelmans, S. M., & Pieters, R. (2008). On emotion specificity in decision making: Why feeling is for doing. Judgment and Decision making, 3(1), 18.

Zhang, S., Tang, J., Wang, H., & Wang, Y. (2015). Enhancing Traffic Incident Detection by Using Spatial Point Pattern Analysis on Social Media. Transportation Research Record: Journal of the Transportation Research Board, (2528), 69-77.