University Transportation Research Center - Region 2

Final Report

A Random Utility Based Estimation Framework for the Household Activity Pattern Problem

Performing Organization: State University of New York (SUNY)

June 2016

Sponsor:
University Transportation Research Center - Region 2

University Transportation Research Center - Region 2

The Region 2 University Transportation Research Center (UTRC) is one of ten original University Transportation Centers established in 1987 by the U.S. Congress. These Centers were established with the recognition that transportation plays a key role in the nation's economy and the quality of life of its citizens. University faculty members provide a critical link in resolving our national and regional transportation problems while training the professionals who address our transportation systems and their customers on a daily basis.

The UTRC was established in order to support research, education and the transfer of technology in the field of transportation. The theme of the Center is "Planning and Managing Regional Transportation Systems in a Changing World." Presently, under the direction of Dr. Camille Kamga, the UTRC represents USDOT Region II, including New York, New Jersey, Puerto Rico and the U.S. Virgin Islands. Functioning as a consortium of twelve major Universities throughout the region, UTRC is located at the CUNY Institute for Transportation Systems at The City College of New York, the lead institution of the consortium. The Center, through its consortium, an Agency-Industry Council and its Director and Staff, supports research, education, and technology transfer under its theme. UTRC's three main goals are:

Research

The research program objectives are (1) to develop a theme based transportation research program that is responsive to the needs of regional transportation organizations and stakeholders, and (2) to conduct that program in cooperation with the partners. The program includes both studies that are identified with research partners of projects targeted to the theme, and targeted, short-term projects. The program develops competitive proposals, which are evaluated to insure the mostresponsive UTRC team conducts the work. The research program is responsive to the UTRC theme: "Planning and Managing Regional Transportation Systems in a Changing World." The complex transportation system of transit and infrastructure, and the rapidly changing environment impacts the nation's largest city and metropolitan area. The New York/New Jersey Metropolitan has over 19 million people, 600,000 businesses and 9 million workers. The Region's intermodal and multimodal systems must serve all customers and stakeholders within the region and globally.Under the current grant, the new research projects and the ongoing research projects concentrate the program efforts on the categories of Transportation Systems Performance and Information Infrastructure to provide needed services to the New Jersey Department of Transportation, New York City Department of Transportation, New York Metropolitan Transportation Council , New York State Department of Transportation, and the New York State Energy and Research Development Authorityand others, all while enhancing the center's theme.

Education and Workforce Development

The modern professional must combine the technical skills of engineering and planning with knowledge of economics, environmental science, management, finance, and law as well as negotiation skills, psychology and sociology. And, she/he must be computer literate, wired to the web, and knowledgeable about advances in information technology. UTRC's education and training efforts provide a multidisciplinary program of course work and experiential learning to train students and provide advanced training or retraining of practitioners to plan and manage regional transportation systems. UTRC must meet the need to educate the undergraduate and graduate student with a foundation of transportation fundamentals that allows for solving complex problems in a world much more dynamic than even a decade ago. Simultaneously, the demand for continuing education is growing – either because of professional license requirements or because the workplace demands it – and provides the opportunity to combine State of Practice education with tailored ways of delivering content.

Technology Transfer

UTRC's Technology Transfer Program goes beyond what might be considered "traditional" technology transfer activities. Its main objectives are (1) to increase the awareness and level of information concerning transportation issues facing Region 2; (2) to improve the knowledge base and approach to problem solving of the region's transportation workforce, from those operating the systems to those at the most senior level of managing the system; and by doing so, to improve the overall professional capability of the transportation workforce; (3) to stimulate discussion and debate concerning the integration of new technologies into our culture, our work and our transportation systems; (4) to provide the more traditional but extremely important job of disseminating research and project reports, studies, analysis and use of tools to the education, research and practicing community both nationally and internationally; and (5) to provide unbiased information and testimony to decision-makers concerning regional transportation issues consistent with the UTRC theme.

Project No(s):
UTRC/RF Grant No: 49997-16-25

Project Date: June 2016

Project Title: A Random Utility Based Estimation Framework for the Household Activity Pattern Problem

Project's Website:
http://www.utrc2.org/research/projects/development-of-household-activity-pattern

Principal Investigator(s):
Jee Eun Kang
Assistant Professor
Department of Industrial and Systems Engineering
University at Buffalo
Buffalo, NY 14260-2050
Tel: (716) 645-2357
Email: jeeeunka@buffalo.edu

Performing Organization:
State University of New York (SUNY)

Sponsor(s):
University Transportation Research Center (UTRC)

To request a hard copy of our final reports, please send us an email at utrc@utrc2.org

Mailing Address:

University Transportation Reserch Center
The City College of New York
Marshak Hall, Suite 910
160 Convent Avenue
New York, NY 10031
Tel: 212-650-8051
Fax: 212-650-8374
Web: www.utrc2.org

Board of Directors

The UTRC Board of Directors consists of one or two members from each Consortium school (each school receives two votes regardless of the number of representatives on the board). The Center Director is an ex-officio member of the Board and The Center management team serves as staff to the Board.

City University of New York
Dr. Hongmian Gong - Geography/Hunter College
Dr. Neville A. Parker - Civil Engineering/CCNY

Clarkson University
Dr. Kerop D. Janoyan - Civil Engineering

Columbia University
Dr. Raimondo Betti - Civil Engineering
Dr. Elliott Sclar - Urban and Regional Planning

Cornell University
Dr. Huaizhu (Oliver) Gao - Civil Engineering

Hofstra University
Dr. Jean-Paul Rodrigue - Global Studies and Geography

Manhattan College
Dr. Anirban De - Civil & Environmental Engineering
Dr. Matthew Volovski - Civil & Environmental Engineering

New Jersey Institute of Technology
Dr. Steven I-Jy Chien - Civil Engineering
Dr. Joyoung Lee - Civil & Environmental Engineering

New York University
Dr. Mitchell L. Moss - Urban Policy and Planning
Dr. Rae Zimmerman - Planning and Public Administration

Polytechnic Institute of NYU
Dr. Kaan Ozbay - Civil Engineering
Dr. John C. Falcocchio - Civil Engineering
Dr. Elena Prassas - Civil Engineering

Rensselaer Polytechnic Institute
Dr. José Holguín-Veras - Civil Engineering
Dr. William "Al" Wallace - Systems Engineering

Rochester Institute of Technology
Dr. James Winebrake - Science, Technology and Society/Public Policy
Dr. J. Scott Hawker - Software Engineering

Rowan University
Dr. Yusuf Mehta - Civil Engineering
Dr. Beena Sukumaran - Civil Engineering

State University of New York
Michael M. Fancher - Nanoscience
Dr. Catherine T. Lawson - City & Regional Planning
Dr. Adel W. Sadek - Transportation Systems Engineering
Dr. Shmuel Yahalom - Economics

Stevens Institute of Technology
Dr. Sophia Hassiotis - Civil Engineering
Dr. Thomas H. Wakeman III - Civil Engineering

Syracuse University
Dr. Riyad S. Aboutaha - Civil Engineering
Dr. O. Sam Salem - Construction Engineering and Management

The College of New Jersey
Dr. Thomas M. Brennan Jr - Civil Engineering

University of Puerto Rico - Mayagüez
Dr. Ismael Pagán-Trinidad - Civil Engineering
Dr. Didier M. Valdés-Díaz - Civil Engineering

UTRC Consortium Universities

The following universities/colleges are members of the UTRC consortium.

City University of New York (CUNY)
Clarkson University (Clarkson)
Columbia University (Columbia)
Cornell University (Cornell)
Hofstra University (Hofstra)
Manhattan College (MC)
New Jersey Institute of Technology (NJIT)
New York Institute of Technology (NYIT)
New York University (NYU)
Rensselaer Polytechnic Institute (RPI)
Rochester Institute of Technology (RIT)
Rowan University (Rowan)
State University of New York (SUNY)
Stevens Institute of Technology (Stevens)
Syracuse University (SU)
The College of New Jersey (TCNJ)
University of Puerto Rico - Mayagüez (UPRM)

UTRC Key Staff

Dr. Camille Kamga: *Director, Assistant Professor of Civil Engineering*

Dr. Robert E. Paaswell: *Director Emeritus of UTRC and Distin*guished Professor of Civil Engineering, The City College of New York

Herbert Levinson: *UTRC Icon Mentor, Transportation Consultant and Professor Emeritus of Transportation*

Dr. Ellen Thorson: *Senior Research Fellow, University Transportation Research Center*

Penny Eickemeyer: *Associate Director for Research, UTRC*

Dr. Alison Conway: *Associate Director for Education*

Nadia Aslam: *Assistant Director for Technology Transfer*

Nathalie Martinez: *Research Associate/Budget Analyst*

Tierra Fisher*: Office Assistant*

Bahman Moghimi*: Research Assistant;*
Ph.D. Student, Transportation Program

Wei Hao*: Research Fellow*

Andriy Blagay: *Graphic Intern*

TECHNICAL REPORT STANDARD TITLE PAGE

1. Report No.	2.Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle A Random Utility Based Estimation Framework for the Household Activity Pattern Problem		5. Report Date June 2016
		6. Performing Organization Code
7. Author(s) Jee Eun Kang		8. Performing Organization Report No.
9. Performing Organization Name and Address Department of Industrial and Systems Engineering University at Buffalo 342 Bell Hall Buffalo, NY 14260-2050		10. Work Unit No.
		11. Contract or Grant No. 49997-16-25
12. Sponsoring Agency Name and Address UTRC The City College of New York 137th Street and Convent Avenue New York, NY 10031		13. Type of Report and Period Covered Final Report,
		14. Sponsoring Agency Code
15. Supplementary Notes		

16. Abstract

This paper develops a random utility based estimation framework for the Household Activity Pattern Problem (HAPP). Based on the realization that output of complex activity-travel decisions form a continuous pattern in space-time dimension, the estimation framework is treated as a pattern selection problem. In particular, we define a variant of HAPP that has capabilities of forecasting activity selection and durations in addition to activity sequencing. The framework is comprised of three steps, (i) choice set generation, (ii) choice set individualization and (iii) multinomial logit estimation. The estimation results show that utilities for work, shopping and disuilities for travel time, time outside home, and average tour delay are found to be significant in activity-travel decision making.

17. Key Words	18. Distribution Statement		
Household Activity Pattern Problem, Activity-Travel Patterns, Random Utility Estimation, Derror minimization, goal programming, multinomial logit estimation			
19. Security Classif (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No of Pages 29	22. Price

Form DOT F 1700.7 (8-69)

Disclaimer

A Random Utility Based Estimation Framework for the Household Activity Pattern Problem

June 20, 2016

Abstract

This paper develops a random utility based estimation framework for the Household Activity Pattern Problem (HAPP). Based on the realization that output of complex activity-travel decisions form a continuous pattern in space-time dimension, the estimation framework is treated as a pattern selection problem. In particular, we define a variant of HAPP that has capabilities of forecasting activity selection and durations in addition to activity sequencing. The framework is comprised of three steps, (i) choice set generation, (ii) choice set individualization and (iii) multinomial logit estimation. The estimation results show that utilities for work, shopping and disuilities for travel time, time outside home, and average tour delay are found to be significant in activity-travel decision making.

Keywords

Household Activity Pattern Problem, Activity-Travel Patterns, Random Utility Estimation, D-error minimization, goal programming, multinomial logit estimation

Authors

Name: Zhiheng Xu
Title: Ph.D. student and Graduate Research Assistant
University at Buffalo

Name: Jee Eun Kang (Principal Investigator)
Title: Assistant Professor
University at Buffalo

Name: Roger B Chen (Collaborator)
Title: Assistant Professor
Rochester Institute of Technology

1 Introduction and Literature Review

In the field of transportation planning and engineering, the activity-based approach continues to generate interest in terms of model development and implementation (Pinjari and Bhat, 2011). Activity-based models focus on travelers' participation in *activities* that derive the need to *travel* and therefore account for a more realistic disaggregate view of travel behavior. Instead of focusing on modeling individual trips, activity-travel models describe *why* such trips are derived, and more generally focus on modeling entire travel-activity patterns. Notable models include Comprehensive Econometric Micro-simulator for Daily Activity-travel Patterns (CEMDAP) (Bhat et al., 2004), Florida Activity Mobility Simulator (FAMOS) (Pendyala et al., 2005), Travel Activity Scheduler for Household Agents (TASHA) (Miller and Roorda, 2003), Household Activity Pattern Problem (HAPP) (Recker, 1995), A Learning-Based Transportation Oriented Simulation System (ALBATROSS) (Arentze and Timmermans, 2004), the Forecasting Evolutionary Activity-Travel of Households and their Environmental RepercussionS (FEATHERS) (Bellemans et al., 2010) and etc.

The Household Activity Pattern Problem (HAPP), is a mathematical programming approach to travel analysis under the activity-based framework Recker (1995). According to HAPP, activity travel patterns are the result of a household optimization with respect to a set of space-time and resource constraints, where the objective function represents travel (dis)utility. HAPP serves as a theoretical reasoning behind travel and activity decisions when faced with temporal and spatial constraints, in addition to resource constraints such as travel budget and tour length (Hägerstraand, 1970). Mathematically, HAPP is an interpretation of personal- (household-) level daily travel posed as a variation of the well-known Pickup and Delivery Pattern Problem with Time Windows (PDPTW) formulation.

The structure of HAPP is as follows. For details, refer to Recker (1995).

$$(\text{HAPP}) \min_{X,T} \text{Travel Disutility} \tag{1}$$

$$A \begin{bmatrix} X \\ \hline T \\ \hline Y \end{bmatrix} \leq B \tag{2}$$

Decision variable X explains spatial decisions, T represents temporal decisions of a traveler, and Y is the tour length variable. Constraints, A, B (2) are comprised of (a) spatial constraints and (b) temporal constraints. The objective function is represented as a weighted combination of multiple terms that are used to capture the key essences of travel behavior. The flexible structure of HAPP as a Mixed Integer Linear Programming (MILP) provides a good platform that is easily extended to meet a modeler's specific objectives while keeping the governing rules of travel and activity decisions within the constraints imposed by time-space geography intact. Rescheduling problem (Gan and Recker, 2008, 2013), destination choice (Kang and Recker, 2013), multi-day (Chow and Nurumbetova, 2015), multi-modal extensions (Chow and Djavadian, 2015), activity arrival and duration choice (Yuan, 2014), time-dependent scheduling (Yuan, 2014) have been proposed based on the original concept. In the application side, scenario-based assessment studies were conducted (Kang and Recker, 2014; Chow, 2014; Recker and Parimi, 1999).

This distinct structure of HAPP brings diversity for travel demand models that are mostly based on discrete choice or agent-based approaches. This helps bring explanations for transportation issues that arguably could not have been addressed by other types of travel demand models. For example, it is possible to simulate cases that we have no previously observed travel demand data of, utilizing the property of HAPP that it is built from the governing rules of travel and activity decisions within the constraints imposed by time-space geography. Despite the advantages, HAPP model has not been widely used as a forecasting tool due to the difficulty of estimation the objective function. Forecast activity-travel decisions and activity-travel patterns cannot be generate personalized pattern with constraints alone. The objective function of HAPP needs personal input, specifically people's valuation of things that influence their fundamental activity participation and travel behavior.

One methodological challenge towards implementing HAPP as a forecasting tool is parameter estimation required for the linear-in-parameters objective function given a conventional dataset of observed travel-activity decisions. These parameters reflect the weighting or value households endogenously place on the components of the objective function. Additionally, the constraints in the mathematical program also contain behavioral parameters that require estimation, such as household time and money budgets.

Growing interest in operationalizing the HAPP model has led to three parameter estimation procedures in the literature. Recker et al. (2008) proposed the earliest method which used a generic-algorithm approach to fitting the objective function parameters in order to minimize the string distances of the observed data and optimal solution to the mathematical program. Chow and Recker (2012) developed an inverse optimization formulation to identify individual weights of the objective function. More recently, Regue et al. (2014) proposed a calibration process based on a differential evolution process. Each of these approaches have significantly operationalized the HAPP, but significant challenges require further attention. The above approaches estimate the parameters for each individual household, requiring significant computational resources for a conventional travel dataset which contain many households. Additionally, due to the combinatorial nature of the HAPP, there exists an infinite number of weight combinations given a range of parameter values that will ensure the optimality with respect to a given observed pattern.

In this paper, an estimation procedure based on random utility maximization (RUM) choice theory is developed to propvide parameter estimates for the HAPP objective function. This approach allows for a scaling of these parameters with respect to the sample dataset used through an econometric estimation framework. Given a linear-in-parameters objective function, this work estimates the parameters based on the observed one day activity-travel patterns in found in conventional travel datasets (Ben-Akiva and Lerman, 1985; Train, 2009).

2 Overview of Framework

One of the features of activity-based travel forecasting models is the continuous time frame. The output of HAPP, as well as many other models, includes travelers' complex decisions of activity participation and travel, that can be represented as *a continuous path* in the time-space dimension as seen in Figure (1). On a horizontal time axis, each activity engagement is represented ("H","T","W"). The spatial distance from the home location is represented in vertical dimension. Each letter stands for the following activities: (H)ome; (T)ravel; and

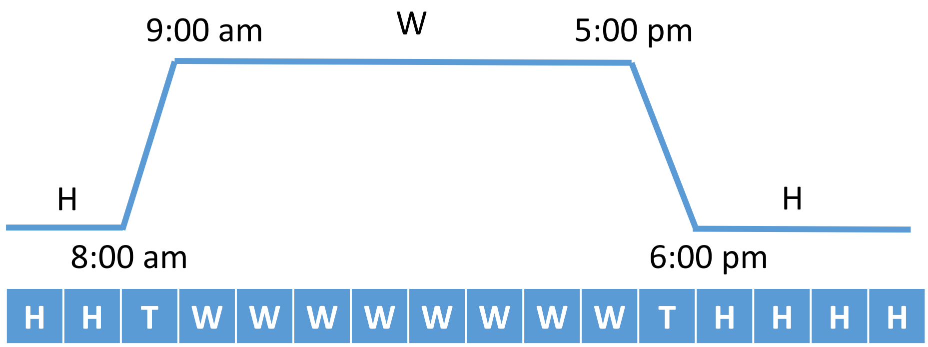

Figure 1: Illustration of a Continuous Path in Space-Time and String Representation of Activity Engagement

This study proposes a framework based on this observation that travelers' activity-travel decision form a continuous path, and this selection of a particular path is based on a random utility theory (RUT). This realization also enables us to connect activity-travel decisions and path selection procedure to existing route choice modeling studies in transportation literature. The framework is built upon two key assumptions. First, the observed activity-travel patterns represent the optimal decisions made for the traveler. Second, a single continuous path in space-time dimension is a single decision output containing complex activity-travel decisions. The observed pattern selection is modeled through a Multinomial Logit model (MNL) (Hensher and Greene, 2003). The functional form of utility functions are assumed universal across choice alternatives, with the attributes of utility functions similar to those in the HAPP objective function. The result of the estimated MNL gives the parameter weights we need for operationalizing and forecasting.

2.1 Estimation Framework

In order to operationalize estimation under an RUM choice framework, the analysis framework developed in this work is comprised of three procedures: (i) choice set generation; (ii) choice set individualization; and (iii) parameter estimation. Figure 3 provides an overview of each of these procedures. All RUM choice models require choice set in which actual decision and alternative choices are known and defined for each decision maker. Utility of each choice is represented by weighted summation of some key attributes of this choice and the choice with highest utility is selected. Based on the assumptions made above, an observation is regarded as the actual decision (daily pattern) and the key attributes come from objective function of HAPP, assuming those terms under HAPP framework capture the real-world decision mechanics. The key challenge is the lack of choice alternatives that are generally unobservable.

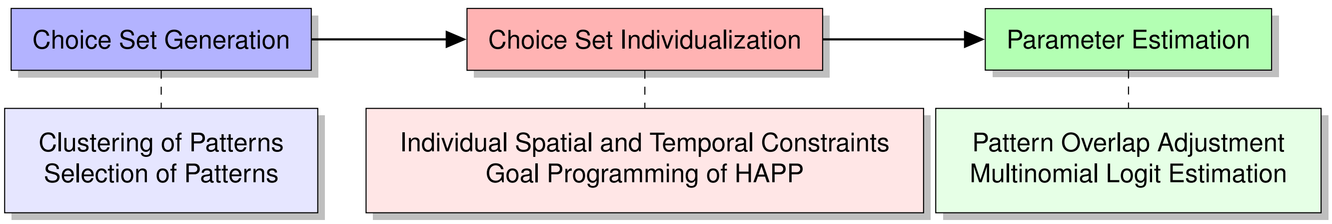

Figure 2: Estimation Framework

Choice Set Generation Under the behavioral framework for this study, travelers' representative activity-travel patterns are modeled as a choice among a choice set of representative patterns. Defining this choice set faced by travelers is necessary for operationalizing the RUM choice framework requires identifying the choice set. To identify the complete set of representative activity-travel patterns, the clustering approach developed by Allahviranloo et al. (2014) is used to identify representative patterns from observed activity-travel patterns in a conventional Household Travel Survey. While a traveler's chosen alternative is easily identified as the observed activity-travel pattern, the alternatives are more difficult to identify. To generate the choice set of alternatives, a pattern is drawn from each of the non-chosen representative pattern clusters. The method developed optimizes the information gain (conceptualized through the inverse Hessian matrix) from this sampling of non-chosen alternatives, more specifically a genetic algorithm that sample non-chosen alternatives and minimizes the D-error of the final sample.

Choice Set Individualization In this study, the choice set generation procedure samples from other observed activity-travel patterns from other travelers in the dataset. However, this leads to the possibility of the non-chosen alternatives being infeasible to the certain travelers. In order to ensure that selected choice alternatives indeed are feasible for each traveler, a goal-programming will adjust the sampled non-chosen alternatives for each traveler, with respect to their personal space-time and resource constraints. We develop the formulation of the goal programming, with the goal of being similar to the sampled activity-pattern, subject to an individual traveler's spatial constraints (through travel time matrix) and temporal constraints (work hour duration in particular).

Parameter Estimation Given a generated choice set of (chosen and non-chosen) alternatives for each traveler in the sample, a choice model is estimated. The alternatives may have significant pattern overlap, for example the timing and duration of in-home occupancy, which would be problematic for choice models assuming the IIA property. This work also addresses methods for dealing with this overlap in choice model estimation. An analogous problem is the path overlap problem from the route choice model estimation literature. These methods will guide the development of these methods for dealing with pattern overlap. The advantage of the proposed methodology is that any choice model formulation can be subsituted for the MNL used in this paper. This is especially advantagous in the travel demand field which has used these choice models extensitvely.

2.2 Utility-Based Activity Participation, Duration, and Travel Decisions

In order to highlight the forecasting capability of HAPP how, we consider the following variation. The key modeling capabilities include, activity participation decision, activity duration decision in addition to various travel decisions and time decisions. This formulation highlights the concepts of utility gain from performing activities and preferred arrival time. Compared to formulation presented by Recker (1995) that only penalize travel related disutility, this formulation includes utility of performing activities so that it introduces a trade-off between pursuing utility by performing more activities and inducing more travel related disutility. That allows the model to accommodate selection of activity participation (i.e., traveler can choose which activity to do and how long the activity should take) to achieve the best personal utility. New terms are introduced to improve the model performance and simplify the estimation of parameters. Similar idea can be found in Yuan (2014) and Chow and Nurumbetova (2015), both introduce utility gain by performing activity. While Yuan (2014) focuses on activity duration choice decision and no activity selection is allowed, Chow and Nurumbetova (2015) further introduces the concept of preferred arrival time to relax hard time window, prism of multi-day arrangement and allows activity selection.

Notation:

- $V = \{1, 2, ..., |V|\}$: The set of vehicles available to the household;
- $P^+ = \{1, 2, ..., n\} = \{W, \hat{P}, S\}$: The set of activity nodes, where W, P, S stand for the sets of work, personal and shopping activities, respectively;
- $P^- = \{n+1, n+2, ..., 2n\}$: The set of return home nodes;
- $P = P^+ \cap P^-$: The set of all activity nodes;
- $N = \{0, P, 2n+1\}$: The set of all nodes, 0 is the start depot, $2n+1$ is the final depot;

- $\hat{\beta} = \{\beta_W, \beta_P, \beta_S, \beta_{TT}, \beta_{TOH}, \beta_{TD}\}$: The set of parameters for each objective term, where the first three parameters stand for utility gain/hour by performing work, personal, shopping activity and the last three stand for travel disutility incurred by total travel time, time outside home and average trip chain delay;

- $t_{u,w}$: The travel time from node u to node w; $u, w \in N$

- $[a_u, b_u]$: Time Window for activity u; $u \in W$

- $L_u \setminus U_u$: The lower\upper limit for the duration of activity u; $u \in W$

- $E_k \setminus F_k$: The earliest departure time\latest arrival time for vehicle k; $k \in V$

- $X^k_{u,w}$: Binary variable, equal to unity if vehicle k travels from node u to node w; $u, w \in N, k \in V$

- T_u: The variable standing for the start time of activity u; $u \in P$

- $T^k_0 \setminus T^k_{n+1}$: The time of vehicle k first departing from home\last returning to home; $k \in V$

- S_u: The variable standing for the duration of activity u; $u \in P^+$

- W_u: The variable standing for waiting time after activity u; $u \in P^+$

- TD: Average Trip Chain Delay.

- I_u: Binary indicator, equal to unity if u activities in P are chosen, $u = 0, 1, 2, ..., n$

- M: Big number

Formulation:

$$
\begin{aligned}
\textbf{Max:} \quad & \beta_W \sum_{u \in W} S_u + \beta_P \sum_{u \in \hat{P}} S_u + \beta_S \sum_{u \in S} S_u + \beta_{TT} \sum_{k \in V} \sum_{u \in N} \sum_{w \in N} t^k_{u,w} X^k_{u,w} \\
& + \beta_{TOH} \Big(\sum_{k \in V} \sum_{u \in N} \sum_{w \in N} t^k_{u,w} X^k_{u,w} + \sum_{u \in P^+} S_u + \sum_{u \in P^+} W_u \Big) + \beta_{TD} TD
\end{aligned} \tag{3}
$$

Subject to:

$$\sum_{k\in V}\sum_{w\in N} X^k_{u,w} \leq 1, u \in P^+ \tag{4}$$

$$\sum_{w\in P^+} X^k_{0,w} \leq 1, k \in V \tag{5}$$

$$\sum_{u\in P^-} X^k_{u,2n+1} \leq 1, k \in V \tag{6}$$

$$\sum_{w\in N} X^k_{u,w} - \sum_{w\in N} X^k_{w,u} = 0, u \in P, k \in V \tag{7}$$

$$\sum_{w\in N} X^k_{w,u} - \sum_{w\in N} X^k_{w,n+u} = 0, u \in P^+, k \in V \tag{8}$$

$$\sum_{k\in V}(\sum_{w\in P^-} X^k_{0,w} + \sum_{u\in P^+} X^k_{u,2n+1} + \sum_{u\in N} X^k_{u,0} + \sum_{w\in N} X^k_{2n+1,w}) = 0 \tag{9}$$

$$T_u + S_u + W_u + t_{u,n+u} \leq T_{n+u}, u \in P^+ \tag{10}$$

$$\sum_{k\in V} X^k_{u,w} = 1 \Longrightarrow T_u + S_u + W_u + t_{u,w} = T_w, u \in P^+, w \in P \tag{11}$$

$$\sum_{k\in V} X^k_{u,w} = 1 \Longrightarrow T_u + t_{u,w} \leq T_w, u \in P^-, w \in P \tag{12}$$

$$X^k_{0,w} = 1 \Longrightarrow T^k_0 + t_{0,w} = T_w, w \in P^+, k \in V \tag{13}$$

$$X^k_{u,w} = 1 \Longrightarrow T_u = T^k_{2n+1}, u \in P^-, k \in V \tag{14}$$

$$E_k \leq T^k_0 \leq T^k_{2n+1} \leq F_k, k \in V \tag{15}$$

$$a_u \leq T_u \leq b_u, u \in P^+ \tag{16}$$

$$\sum_{k\in V}\sum_{w\in P} X^k_{u,w} = 0 \Longrightarrow S_u = 0, u \in P^+ \tag{17}$$

$$L_u + M(\sum_{k\in V}\sum_{w\in P} X^k_{u,w} - 1) \leq S_u \leq U_u, u \in P^+ \tag{18}$$

$$\sum_{u=0}^{n} I_u = 1 \tag{19}$$

$$I_0 = 1 \Longrightarrow TD = 0 \tag{20}$$

$$I_0 = 1 \Longrightarrow \sum_{k\in V}\sum_{u\in P^+}\sum_{w\in P} X^k_{u,w} = 0 \tag{21}$$

$$\tag{22}$$

For m = 1 to n

$$I_m = 1 \Longrightarrow TD = \frac{1}{m} \sum_{u \in P^+} (T_{n+u} - T_u - S_u) \tag{23}$$

$$I_m = 1 \Longrightarrow \sum_{k \in V} \sum_{u \in P^+} \sum_{w \in P} X^k_{u,w} = m \tag{24}$$

End

The first three terms in the objective function stand for utility gain by performing work, personal social and shopping activities. The other terms represent travel disutility incurred by traveling. Total travel time, total time outside of home and average trip chain delay are penalized respectively. Selection of disutility terms will be discussed in detail in section 5. Constraints (4)-(9) focus on constructing the HAPP network. (4) means one activity node can be visited at most once. (5)-(6) mean the flow leaving start depot must head to activity nodes and have to return to final depot through return home nodes. (7) is the balance flow constraints. (8) means an activity node and its corresponding return home node must both be visited or both not. (9) remove unwanted flows from the network. Constraints (10)-(14) enable time transmission through trip chain. (15)-(18) constrain the departure and arrival time for each vehicle; time windows for certain activities; upper and lower limit for duration of each activity. (19)-(24) applied a trick to capture the average tour delay. I variables capture the number of activities chosen and different I variable leads to different format of TD that makes it always stand for the average of tour delays for all chosen activities.

3 Choice Set Generation

The key challenge of estimating HAPP based on RUT is that (1) alternatives are not directly observable and (2) there is an infinitude number of reasonable alternatives given the continuous time dimension. While generating reasonable choice alternatives has been an integral part in route choice modeling (Bekhor et al., 2006; Prato, 2009), the task of generating quality choice set is even more complex and challenging.

In this study, the choice set generation procedure samples from other observed travel patterns from other travelers in the dataset. A choice set will be comprised of five *representative* activity-travel patterns shown in the data set. However, this leads to the possibility of the non-chosen alternatives being infeasible to the certain travelers. In order to ensure that selected choice alternatives indeed are feasible for each traveler, a goal-programming will adjust the sampled non-chosen alternatives for each traveler, with respect to their personal space-time and resource constraints. We develop the formulation of the goal programming that focuses on both spatial constraints (through travel time matrix) and temporal constraints (work hour duration in particular).

3.1 2-Stage Activity-Travel Pattern Clustering

Several works investigated classification and understanding of *representative/typical* activity-travel patterns (Allahviranloo et al., 2014; Recker et al., 1985). We use a 2-stage clustering method that is modified from a single stage clustering proposed in Allahviranloo et al. (2014) to account for both **time allocation, or activity engagement** and **activity sequencing**. Allahviranloo et al. (2014) used differences of two activity-travel patterns (measured through Sequence Alignment Method, SAM) as features of each pattern.

A uni-dimensional string representation of activity-travel decisions is used as the basic representation of the data. Sequence analysis has been widely used in various fields to understand features, functions, structures, or evolution. Sequencing representation was first used for activity patterns by Wilson (1998) to analyze one-dimensional activity patterns. Later, multi-dimensional representation was used to include information of mode choice, location, and accompanying persons (Joh et al., 2002). For this research project, we first define a representation that includes both approaches to account for time allocation and activity sequencing. Given the unit time stamp for 18-hour period (Starting from 6:00 pm and ending at 12:00 am), each time stamp (5 min) labeled as activity purpose such as "Home(H)", "Work(W)", "Shopping(S)", "Recreational(R)", "Personal(P)", "Maintenance(M)", etc., as well as travel as an activity "Travel(T)". An example is shown in Figure (1). Since activity purpose and duration include decisions as well as travel decisions are all captured, Levenshtein distance (Kruskal, 1983) between every pair of observations captures the dissimilarity between them. The longer the distance is, the more dissimilar two patterns are. All to All dissimilarity matrix got from previous process will be features of clustering as in Allahviranloo et al. (2014).

The most widely used clustering technique is k-means. Two key concerns are, how to set K and how to measure both **time allocation** and **activity-travel sequencing**. First, a trial and error method is used to set K as clustering is an unsupervised learning. We initially set K to be 3, 4, 5, 6 and evaluate the clustering result one by one. Large K is not preferred for complexity of the MNL and given the limited size of data. We calculate the average of each attribute for each cluster and evaluate how variant the attributes of one cluster are to another. We then choose K that yields the most distinguished clusters. In other words, clusters have very different level of attributes so that they are not similar to each other.

To accommodate **activity sequencing** decision in addition to **time allocation**, a new string that represents changes in activity types is generated. By simply removing all consecutive duplicate items in the string generated for time allocation, the new string is essentially a sequence of activities. As an example, the string in Figure 1 is reduced to "HTWTH".

A sequential 2-stage method is applied for clustering. The process is presented in the following figure:

Time allocation and activity participation are taken into account in a sequential manner. After clustering based on time allocation, the observations fall into each cluster are similar in terms of time allocation but not necessarily similar in terms of activity sequencing. For example, some activities with short duration may be pretty much ignored since there are not many

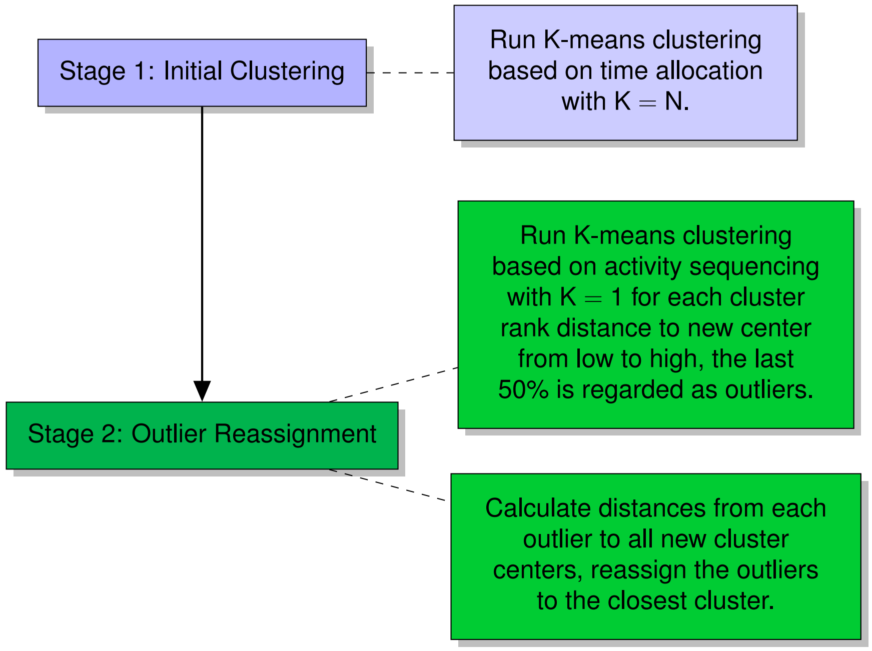

Figure 3: 2-Stage Clustering Method

intervals labeled by such activities. But a pattern with short-time activities may have significant difference with another pattern in terms of activity sequencing. That's the motivation to reassign some of the observations in terms of difference of activity sequencing. After evaluating the how variant attributes are among all clusters, the results indicate that $K = 5$ yields clusters that are more different to each other.

3.2 Representative Activity-Travel Patterns

This study is based on 2000 California Statewide Household Travel Survey. For the sake of simplicity, we only choose samples that are categorized as worker and belong to single member household. Data points with missing information are excluded and all trips start from home and end at home. The sample size of the study is 2183.

The following Figure 4 shows 5 cluster centers we get as representatives of patterns.

Basically we can find that centers of cluster 1 and 2 stand for two different types of full-time workers. Center of cluster 4 is likely to represent part-time workers. Centers of cluster 3 and 5 can be interpreted as two types of patterns when the worker is not on duty. Center of cluster 3 shows that the worker tends to perform few out-of-home activities while for center of cluster 5 the worker prefers performing many selective activities.

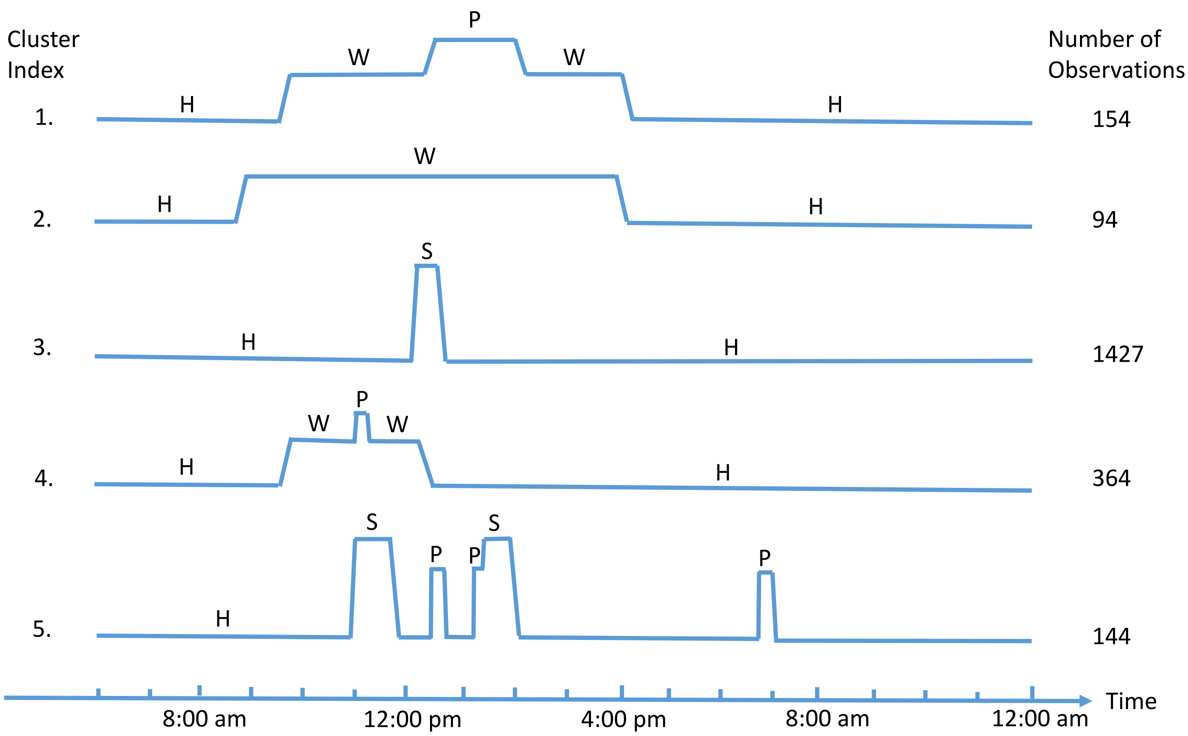

Figure 4: Representative Activity-Travel Patterns

3.3 Choice Set Sampling to Minimize D-error

As discussed above, the choice set of alternatives faced by travelers is typically latent or unobservable to the analyst. We assume that travelers face a choice set of five different patterns identified in the cluster analysis. The individual traveler chooses one type of pattern from this choice set of five patterns. Given a sample of observed daily patterns, for each observed pattern the cluster membership is know. The task faced is to reconstruct the choice sets by sampling from the other clusters of pattern types. We acknowledge that some choices may be actually infeasible because of spatial and temporal constraints (for example, a pattern might have a work stop, but the traveler is retired). This will be addressed in Section 4

While simply sampling an observed pattern from each cluster is one possibility for constructing this choice set, with respect to model estimation, this selection can be more deliberate, minimizing the resulting standard error of parameter estimates. From this methodological standpoint, choice set construction aims to generate a choice set, where the observed pattern is one alternative, and the rest are sampled from the identified clusters, so as to minimize the inverse of the Asymptotic Variance-Covariance (AVC) matrix of parameters.

The swapping algorithm presented in the original paper cannot be adapted to our case. The swapping algorithm works as follows: starting with initial choices, attributes are modified one by one. D-error is calculated at each step and the algorithm terminates if D-error converges to a certain point. The major problem is, for the attributes associated with travel pattern, the values cannot be modified arbitrarily since the intrinsic space and time constraints behind them might be broken. That's why we choose to sample one pattern from each cluster which is valid in nature. The following equation gives the utility function of choice i:

$$U_i = \beta_W W_i + \beta_S S_i + \beta_P P_i + \beta_{TT} TT_i + \beta_{TOH} TOH_i + \beta_{TD} TD_i, i = 1, 2, .., 5; \qquad (25)$$

Where $\beta_W, \beta_S, \beta_P, \beta_{TT}, \beta_{TOH}$ and β_{TD} stand for coefficients of the crucial attributes we present above. They are homogeneous for all choices. W_i, S_i, P_i, T_i, E_i and D_i are the numerical value of attributes calculated based on pattern i.

D-error is calculated based on determinant of the inverse of AVC matrix which can only be got once the choice set is made up. In order to make a choice set, one pattern is sampled within each cluster except for the one observation belongs to and each cluster contains hundreds to thousands items. Enumerating all the possibilities requires evaluation of D-error for each instance and that makes it very computational expensive. Due to the combinatorial nature of this problem, we choose to apply genetic algorithm to get the best available choice set within the time range.

4 Choice Individualization via Goal Programming

The selected samples from last step are from different individuals with different personal, space and time constraints. Although the sampled pattern should be close to the real representative alternative the individual may face, such differences can bring noises to estimation. Both spatial (mainly travel times) and temporal (mainly work activity) noises exists and are adjusted.

The sampled pattern is from an individual living in a totally different geographical location. Thus, the travel time matrix input is replaced with that of the particular traveler. For cases where full travel time information is not available, we create a travel time matrix that is plausible from known travel times (i.e., respecting triangle inequality) as well as distribution of travel times for activity types from the data set. The number of work activities and preferred (or constrained) work duration come from the observation while the sampled pattern controls the preferred arrival time for all activities and existence and duration for activities except for work. If there is also a work activity in sampled pattern, we constraint the work duration of selected alternative pattern by the preferred work duration of actual observation and the work start time of the sampled pattern. Similarly, we constraint start time and duration of other activities based on sampled pattern if existing. If an activity is not bounded by any inferred constraints, we introduce an universal loose bound for duration of such activity in case of a giant activity is generated without any restriction.

Using these individualized spatial and temporal constraints, the goal programming is formulated to derive a feasible pattern for each traveler. The goal is to be as close as the sampled pattern given individual temporal and spatial constraints. The formulation is as follows.

Notations:

- $P^+ = \{1, 2, ..., n\} = \{W^+, O^+\} = \{W^+, \hat{P}^+, S^+\}$: The set of activity nodes where W^+ is the set of work activities nodes and O^+ is the set of optional activity nodes, which contains $\hat{P}^+$ personal social activities and S^+ shopping activities;
- $P^- = \{n+1, n+2, ..., 2n\} = \{W^-, O^-\} = \{W^-, \hat{P}^-, S^-\}$: The set of corresponding return home nodes;
- $\hat{\beta} = \{\beta_W, \beta_P, \beta_S\}$: The set of parameters for each objective term, where β_W is the weight for time difference of work and β_O is the weight for time difference of optional activities.
- g_u: The preferred arrival time for activity u;
- q_u: The preferred duration of activity u;
- $L_u \setminus U_u$: Lower bound \ upper bound multipliers for preferred duration; $u \in P^+$
- T_u: The variable standing for the start time of activity u;
- $\tau_u^e \setminus \tau_u^l$: The early\late arrival deviation with respect to preferred arrival time.

Formulation:

$$\textbf{Min} \quad \beta_W \sum_{u \in W^+ \cup W^-} (\tau_u^e + \tau_u^l) + \beta_P \sum_{u \in \hat{P}^+ \cup \hat{P}^-} (\tau_u^e + \tau_u^l) + \beta_S \sum_{u \in S^+ \cup S^-} (\tau_u^e + \tau_u^l) \tag{26}$$

Subject to:

Constraints (4)-(15)

$$T_u + \tau_u^e - \tau_u^l = g_u, u \in P \tag{27}$$

$$\text{If } q_u \neq null \quad L_u q_u \leq S_u \leq U_u q_u, u \in P^+ \tag{28}$$

The objective function minimizes the deviation between activity start times in the generated and target patterns. Constraint (33) stands for the function form of deviation. Constraint (34) limit the duration of new pattern with respect to the target duration and the activity type, while generally speaking, the bound is tighter for work activities and looser for optional activities, since work activities are usually more restrictive.

The following Figure 5 gives an real-case illustration example of how the goal programming shapes the generated alternatives. The duration of work activities in alternative 1 and 4 is shorten based on observed pattern, while the second work activity in alternative 4 is omitted because there is no such work activity observed. As is presented in previous steps, we generate a travel time matrix in which some cells are directly extracted from observed pattern while others are filled by sampling from distributions. In alternative 1, 2 and 3, we can see that part of the travel times are personalized based on the generated travel time matrix that comes from observation.

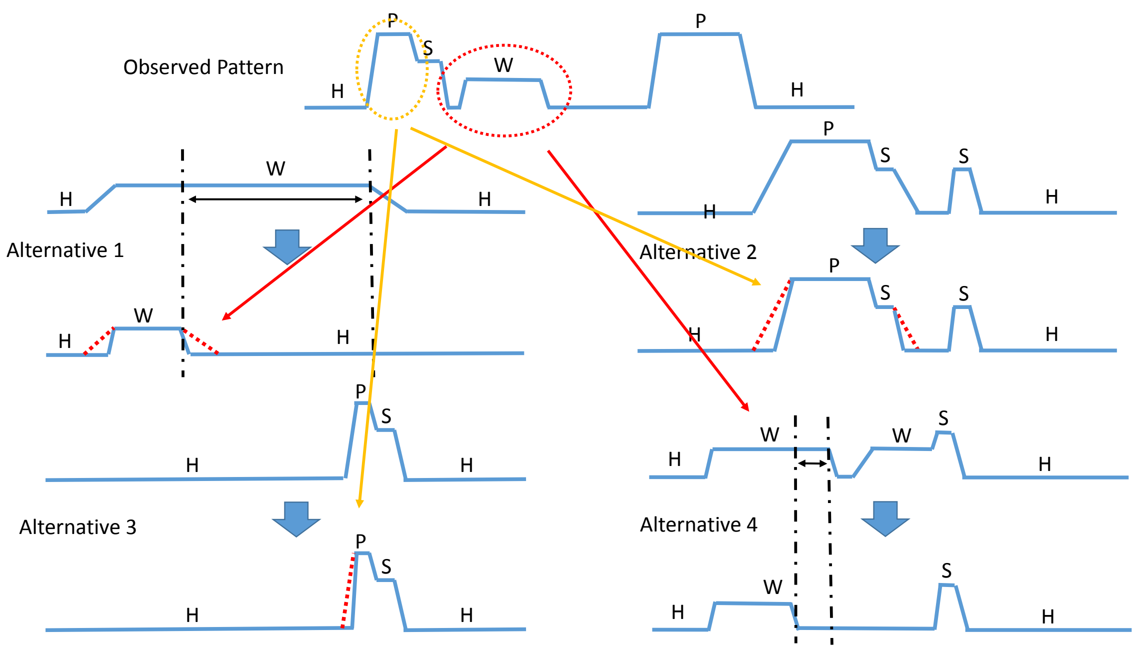

Figure 5: Illustration of Choice Individualization via Goal Programming

5 Multinomial Logit Model Estimation

When selecting travel utility/disutility terms, they must be compatible with both MNL and MILP structure. Objective terms used for previous HAPP literature (Recker, 1995; Recker et al., 2008; Regue et al., 2014; Chow and Recker, 2012) are rather simplified to represent real utility/disutility for the linearity and computational convenience. We introduce new more realistic terms (Table 1) and computational solutions. Not only these terms better represent disutility conceptually, but the improved estimation results confirm such an argument. It is noted that formulations shown in Sections 2.2 and 4 accommodate these terms.

HAPP Literature		**New Proposed Terms**	
Total travel time	$\sum_{k\in V}\sum_{u\in N}\sum_{w\in N} t_{u,w}X^k_{u,w}$	**Total travel time**	$\sum_{k\in V}\sum_{u\in N}\sum_{w\in N} t_{u,w}X^k_{u,w}$
Extent of the day	$\sum_{k\in V}(T^k_{2n+1} - T^k_0)$	**Time outside home**	$\sum_{k\in V}\sum_{u\in N}\sum_{w\in N} t^k_{u,w}X^k_{u,w} + \sum_{u\in P^+} S_u + \sum_{u\in P^+} W_u$
Trip chain delay	$\sum_{u\in P^+}(T_{n+u} - T_u)$	**Average trip chain delay**	$\sum_{u\in P^+}(T_{n+u} - T_u - S_u)/\#activity$

Table 1: Travel Disutility Terms

In the variant version of HAPP we present, the objective function consists of utility of performing activities and disutility incurred by traveling. The trade-off between these two parts makes it possible for activity selection during the decision making process. For the utility gain by doing work (β_W), shopping (β_S) and personal social activities (β_P), they can be regarded as independent with each other. We revised the previously used travel disutility terms to represent more logical and realistic measurements as well as to improve the model performance. Time outside of home excludes the time spent at home during the day that should be penalized. New definition of trip chain delay mitigates the exaggerated effect of the previously used term when number of activities is increased. And it excludes the duration of activities in order to avoid double counting issue. By making these changes, we greatly reduce the correlation among travel disutility terms and make it better fit MNL.

5.1 Pattern Overlap

Overlap among alternatives violates the IIA assumption. Since we can hardly avoid overlap among patterns, it is necessary to capture its effect. An analogous problem is present in route choice models (Bekhor et al., 2006; Prato, 2009). We use a similar structure to path-size logit or C-logit to address this overlap. The expression of probability P_i of choosing pattern i within the choice set C is given as follows:

$$P_i = \frac{V_i + \beta_{CF} CF_i}{\sum_{j \in C}(V_j + \beta_{CF} CF_j)} \tag{29}$$

where V_i is the observed utility.

In this work, a commodity term is defined for each choice to capture the similarity to all other alternatives within the same choice set. The idea is similar to what we did in section 3. A string representation is generated for each pattern. Similarity $S_{i,j}$ between two patterns i and j is defined as the proportion of similarity and calculated as $1 - \frac{\text{Levinstein Distance}}{\text{Maximum Possible Levinstein Distance}}$. As the Levinstein Distance calculates the dissimilarity between two patterns (Section 3.1), the proportion of dissimilarity is calculated by using the denominator of maximum possible dissimilarity score (i.e., two patterns are completely different). As time spent at home is very dominant for most of people as ”Home” for night time, we restrict similarity calculation to the rest portion, from 6AM till 12AM next day. In our data set, the earliest departure and the latest arrival back at home were within 6AM and 12AM time window.

The overlap, which is represented by the commodity factor CF, is the average of similarity to all other patterns in the choice set:

$$CF_i = \sum_j S_{i,j} \tag{30}$$

5.2 Estimation Results

As discussed before, the utility function contains two parts: the utility gain by performing activities and the disutility incurred by traveling. For all 3 models listed, the first part of utility gain remains the same. All activities are categorized into work, personal social activities and shopping, the marginal gain of performing activities within the same category is regarded as the same. The definition of travel disutility is different among 3 models. Details are shown in Table 2.

The model fit is increased while the new definitions of extent and tour delay are introduced. Model 3 is the most preferred in the three models. First, most intercepts of utility functions are insignificant. Since HAPP is an optimization model, intercepts have no impact on its decision process (Constant can be removed from objective function in optimization), but does have impact on the decision process of MNL, which causes bias on prediction. Secondly, overlap is insignificant, which is also preferred because at the time of predicting one’s pattern, the alternative choices are not available. Thirdly, the signs of variables are okay. The sign of coefficients are generally good in theory. When interpreting the results and implications, it is noted that the ”Time Outside Home” measure will relatively negate all utility gains. For example the travel

Model 1 HAPP	$\beta_W \sum_{u\in W} S_u + \beta_P \sum_{u\in \hat{P}} S_u + \beta_S \sum_{u\in S} S_u + \beta_{TT} \sum_{k\in V} \sum_{u\in N} \sum_{w\in N} t^k_{u,w} X^k_{u,w} + \beta_{ED} \sum_{k\in V}(T^k_{2n+1} - T^k_0) + \beta_{TC} \sum_{u\in P^+}(T_{n+u} - T_u)$
Utility Function	$V_i = \beta_W W_i + \beta_S S_i + \beta_P P_i + \beta_{TT} TT_i + \beta_{ED} ED_i + \beta_{TC} TC_i$
Model 2 HAPP	$\beta_W \sum_{u\in W} S_u + \beta_P \sum_{u\in \hat{P}} S_u + \beta_S \sum_{u\in S} S_u + \beta_{TT} \sum_{k\in V} \sum_{u\in N} \sum_{w\in N} t^k_{u,w} X^k_{u,w} + \beta_{ED} \sum_{k\in V}(T^k_{2n+1} - T^k_0) + \beta_{TD} TD$
Utility Function	$V_i = \beta_W W_i + \beta_S S_i + \beta_P P_i + \beta_{TT} TT_i + \beta_{ED} ED_i + \beta_{TD} TD_i$
Model 3 HAPP	$\beta_W \sum_{u\in W} S_u + \beta_P \sum_{u\in \hat{P}} S_u + \beta_S \sum_{u\in S} S_u + \beta_{TT} \sum_{k\in V} \sum_{u\in N} \sum_{w\in N} t^k_{u,w} X^k_{u,w} + \beta_{TOH}(\sum_{k\in V} \sum_{u\in N} \sum_{w\in N} t^k_{u,w} X^k_{u,w} + \sum_{u\in P^+} S_u + \sum_{u\in P^+} W_u) + \beta_{TD} TD$
Utility Function	$V_i = \beta_W W_i + \beta_S S_i + \beta_P P_i + \beta_{TT} TT_i + \beta_{TOH} TOH_i + \beta_{TD} TD_i$

Table 2: Comparison of Three Models

time brings, counter-intuitively, positive utility. However, including time outside home, its sign stays in the negative region and will be a disutility. Work, personal, and shopping all present positive utilities after accommodating with time outside home. Personal activities bring the most utility per time unit followed by work and shopping activities.

Index	Attributes	Estimate	St. Error	t-value	Pr($> \lvert t \rvert$)	Significance
Model 1:						
	2:(intercept)	0.2324	0.2315	1.0043	0.3153	
	3:(intercept)	3.2017	0.1526	20.9793	0.0000	***
	4:(intercept)	1.0363	0.1669	6.2094	0.0000	***
	5:(intercept)	0.5297	0.1930	2.7452	0.0060	**
	Work	0.6425	0.0383	16.7740	0.0000	***
	Personal	2.3277	0.0976	23.8462	0.0000	***
	Shopping	0.6086	0.0913	6.6684	0.0000	***
	TravelTime	-0.0387	0.0593	-0.6533	0.5135	
	Extent	-0.4862	0.0337	-14.4259	0.0000	***
	TourDelay (Original)	-0.0109	0.0103	-1.0572	0.2904	
	Overlap	16.3095	1.1121	14.6659	0.0000	***
	Log-Likelihood:	-972.56				
	McFadden R^2:	0.58687				
	Likelihood ratio test :	chisq = 2763.2				
Model 2:						
	2:(intercept)	-0.4553	0.3634	-1.2530	0.2102	
	3:(intercept)	2.0047	0.2349	8.5335	0.0000	***
	4:(intercept)	0.2354	0.3475	0.6775	0.4981	
	5:(intercept)	-0.8860	0.3534	-2.5074	0.0122	*
	WorkDuration	1.1051	0.0826	13.3815	0.0000	***
	Personal	2.4920	0.1846	13.4975	0.0000	***
	Shopping	0.9702	0.1612	6.0168	0.0000	***
	TravelTime	0.2878	0.0944	3.0494	0.0023	**
	Extent	-0.3587	0.0569	-6.3020	0.0000	***
	TourDelay (Modified)	-1.0332	0.0585	-17.6697	0.0000	***
	Overlap	6.7425	2.0765	3.2471	0.0012	**
	Log-Likelihood:	-259.4				
	McFadden R^2:	0.89129				
	Likelihood ratio test :	chisq = 4253.4				

Index	Attributes	Estimate	St. Error	t-value	Pr(> $|t|$)	Significance
Model 3:						
	2:(intercept)	-0.6810	0.5513	-1.2354	0.2167	
	3:(intercept)	1.5622	0.3575	4.3704	0.0000	***
	4:(intercept)	-0.5104	0.6071	-0.8407	0.4005	
	5:(intercept)	-0.3323	0.4751	-0.6996	0.4842	
	Work	2.0113	0.1632	12.3251	0.0000	***
	Personal	3.4870	0.3433	10.1567	0.0000	***
	Shopping	1.8509	0.2650	6.9840	0.0000	***
	TravelTime	0.7610	0.1325	5.7442	0.0000	***
	TimeOutsideHome	-0.8638	0.0873	-9.8996	0.0000	***
	TourDelay (Modified)	-1.2284	0.0994	-12.3620	0.0000	***
	Overlap	4.4151	2.8172	1.5672	0.1171	
	Log-Likelihood:	-121.96				
	McFadden R^2:	0.94953				
	Likelihood ratio test :	chisq = 4588.8				

Table 3: Selected Model Results. Significance code(p-value threshold, label): 0 ”***”, 0.001 ”**”, 0.01 ”*”, 0.05 ”.”, 0.1 ” ”

5.3 Effects of Choice Individualization via Goal Programming

Goal programming plays a crucial role of improving the performance for the estimation. Theoretically the initial choice set sampled from others' reported patterns is the best bet we have for minimizing D-error. However, it is far from perfect because essentially the alternatives are patterns of others. It does give a rough idea how the pattern of this particular person may look like but it doesn't give much implication on the details (Link travel time, social economic attributes, etc). The following and graph give the experimental results of the MNL performance before and after goal programming process:

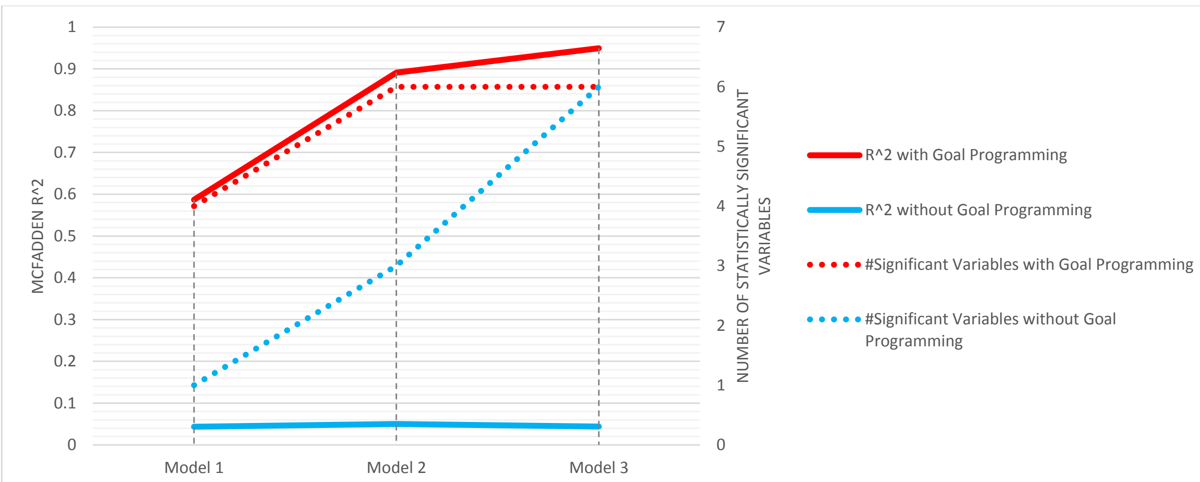

Figure 6: Examples for Discussion of Attributes

The choice individualization improves the estimation results significantly, particularly for Model where new disutility terms of time outside home and average tour delay are introduced. Details of estimation results are available in Appendix.

6 Conclusion

This paper develops an estimation framework based on Random Utility Theory for the well-known Household Activity Pattern Problem (HAPP). Based on the realization that travelers' activity-travel decisions form a continuous line in space and time, the HAPP is treated as a pattern selection procedure. In addition to providing theoretical basis of heterogeneous individuals' decision making, RUT based estimation generates an unique identified set of parameters which is more suited for forecasting capability of HAPP.

The proposed framework includes three components. In **Choice Set Generation** we select one alternative pattern from each pattern cluster found in the data set that are distinctively different. We select patterns that will minimize the D-error based on a genetic algorithm. Then we individualize choice set alternatives in **Choice Set Individualization** based on a goal programming that will create a feasible pattern as close to the selected sample pattern. The goal programming formulation contains constraints of individuals' temporal and spatial constraints. This choice individualization step is found to be significant in increasing the fit of the MNL estimation. For the **MNL Estimation**, overlap among alternatives is treated as the commodity factor (CF) as it is in C-logit or path size logit used in route choice set generation problems. Both travel disuitliy terms and activity participation utility gains are found to be significant.

7 Acknowledgment

This work, in part, has been funded by the Region II University Transportation Research Center (UTRC) and the National Science Foundation (NSF). Authors are grateful for their generous support.

References

Allahviranloo, M., R. Regue, and W. Recker (2014). Pattern clustering and activity inference. In *Transportation Research Board 93rd Annual Meeting*, Number 14-1274.

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

Bekhor, S., M. E. Ben-Akiva, and M. S. Ramming (2006). Evaluation of choice set generation algorithms for route choice models. *Annals of Operations Research 144*(1), 235–247.

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

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

Bhat, C. R., J. Y. Guo, S. Srinivasan, and A. Sivakumar (2004). Comprehensive econometric microsimulator for daily activity-travel patterns. *Transportation Research Record: Journal of the Transportation Research Board 1894*(1), 57–66.

Chow, J. Y. (2014). Activity-based travel scenario analysis with routing problem reoptimization. *Computer-Aided Civil and Infrastructure Engineering 29*(2), 91–106.

Chow, J. Y. and S. Djavadian (2015). Activity-based market equilibrium for capacitated multimodal transport systems. *Transportation Research Part C: Emerging Technologies 59*, 2–18.

Chow, J. Y. and A. E. Nurumbetova (2015). A multi-day activity-based inventory routing model with space–time–needs constraints. *Transportmetrica A: Transport Science 11*(3), 243–269.

Chow, J. Y. and W. W. Recker (2012). Inverse optimization with endogenous arrival time constraints to calibrate the household activity pattern problem. *Transportation Research Part B: Methodological 46*(3), 463–479.

Gan, L. P. and W. Recker (2008). A mathematical programming formulation of the household activity rescheduling problem. *Transportation Research Part B: Methodological 42*(6), 571–606.

Gan, L. P. and W. Recker (2013). Stochastic preplanned household activity pattern problem with uncertain activity participation (shapp). *Transportation Science 47*(3), 439–454.

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

Hensher, D. A. and W. H. Greene (2003). The mixed logit model: the state of practice. *Transportation 30*(2), 133–176.

Joh, C.-H., T. Arentze, F. Hofman, and H. Timmermans (2002). Activity pattern similarity: a multidimensional sequence alignment method. *Transportation Research Part B: Methodological 36*(5), 385–403.

Kang, J. E. and W. Recker (2013). The location selection problem for the household activity pattern problem. *Transportation Research Part B: Methodological 55*, 75–97.

Kang, J. E. and W. W. Recker (2014). Measuring the inconvenience of operating an alternative fuel vehicle. *Transportation Research Part D: Transport and Environment 27*, 30–40.

Kruskal, J. B. (1983). An overview of sequence comparison: Time warps, string edits, and macromolecules. *SIAM review 25*(2), 201–237.

Miller, E. J. and M. J. Roorda (2003). Prototype model of household activity-travel scheduling. *Transportation Research Record: Journal of the Transportation Research Board 1831*(1), 114–121.

Pendyala, R. M., R. Kitamura, A. Kikuchi, T. Yamamoto, and S. Fujii (2005). Florida activity mobility simulator: overview and preliminary validation results. *Transportation Research Record: Journal of the Transportation Research Board 1921*(1), 123–130.

Pinjari, A. R. and C. R. Bhat (2011). Activity-based travel demand analysis. *A Handbook of Transport Economics 10*, 213–248.

Prato, C. G. (2009). Route choice modeling: past, present and future research directions. *Journal of Choice Modelling 2*(1), 65–100.

Recker, W., J. Duan, and H. Wang (2008). Development of an estimation procedure for an activity-based travel demand model. *Computer-Aided Civil and Infrastructure Engineering 23*(7), 483–501.

Recker, W. W. (1995). The household activity pattern problem: General formulation and solution. *Transportation Research Part B: Methodological 29*(1), 61–77.

Recker, W. W., M. G. McNally, and G. S. Root (1985). Travel/activity analysis: pattern recognition, classification and interpretation. *Transportation Research Part A: General 19*(4), 279–296.

Recker, W. W. and A. Parimi (1999). Development of a microscopic activity-based framework for analyzing the potential impacts of transportation control measures on vehicle emissions. *Transportation Research Part D: Transport and Environment 4*(6), 357–378.

Regue, R., M. Allahviranloo, and W. Recker (2014). Understanding household priorities when scheduling activities.

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

Wilson, W. C. (1998). Activity pattern analysis by means of sequence-alignment methods. *Environment and Planning A 30*(6), 1017–1038.

Yuan, D. (2014). *Incorporating Individual Activity Arrival and Duration Preferences within a Time-of-day Travel Disutility Formulation of the Household Activity Pattern Problem (HAPP)*. UNIVERSITY OF CALIFORNIA, IRVINE.

8 Appendix

Effect of Goal Programming Stage on Estimation Results

As we keep revising the travel disutility terms, the performance of MNL tends to become better in terms of model fit (McFadden R^2) and t statistics (Number of statistically significant variables), which gives an illustration of the necessity of making these changes. We can see that the model fit is generally low before goal programming. It is expected so because of the lack of implication on personal details. The model fit as well as the number of significant variables tends to increase dramatically after goal programming.

Index	Attributes	Estimate	St. Error	t-value	Pr($>\|t\|$)	Significance
Model 1:						
	2:(intercept)	-0.2645	0.1409	-1.8764	0.0606	.
	3:(intercept)	2.3370	0.0932	25.0519	0.0000	***
	4:(intercept)	0.8611	0.1271	6.7705	0.0000	***
	5:(intercept)	-0.4220	0.1456	-2.8982	0.0037	**
	Work	0.0276	0.0238	1.1568	0.2473	
	Personal	0.0357	0.0313	1.1390	0.2547	
	Shopping	0.0341	0.0512	0.6662	0.5052	
	TravelTime	0.0032	0.0387	0.0841	0.9329	
	Extent	-0.0233	0.0187	-1.2477	0.2121	
	TourDelay (Original)	0.0138	0.0055	2.4870	0.0128	*
	Overlap	8.1533	0.6515	12.5146	0.0000	***
	Log-Likelihood:	-2251.6				
	McFadden R^2:	0.04355				
	Likelihood ratio test :	chisq = 205.05				
Model 2:						
	2:(intercept)	-0.1417	0.1405	-1.0090	0.3129	
	3:(intercept)	2.3336	0.0920	25.3386	0.0000	***
	4:(intercept)	1.0348	0.1285	8.0507	0.0000	***
	5:(intercept)	-0.2871	0.1450	-1.9797	0.0477	*
	WorkDuration	0.0171	0.0221	0.7738	0.4390	
	Personal	0.0284	0.0287	0.9910	0.3216	
	Shopping	0.0080	0.0486	0.1663	0.8679	
	TravelTime	-0.0675	0.0395	-1.7066	0.0878	.
	Extent	-0.0273	0.0189	-1.4438	0.1487	
	TourDelay (Modified)	0.1241	0.0254	4.8752	0.0000	***
	Overlap	8.3338	0.6329	13.1668	0.0000	***
	Log-Likelihood:	-2266.4				
	McFadden R^2:	0.050173				
	Likelihood ratio test :	chisq = 239.44				

Index	Attributes	Estimate	St. Error	t-value	Pr($>\|t\|$)	Significance
Model 3:						
	2:(intercept)	-0.2265	0.1354	-1.6725	0.0944	.
	3:(intercept)	2.3366	0.0907	25.7372	0.0000	***
	4:(intercept)	0.9521	0.1273	7.4783	0.0000	***
	5:(intercept)	-0.2027	0.1396	-1.4516	0.1466	
	Work	-0.0617	0.0163	-3.7781	0.0001	***
	Personal	-0.0570	0.0221	-2.5725	0.0100	*
	Shopping	-0.0461	0.0453	-1.0179	0.3087	
	TravelTime	-0.0675	0.0358	-1.8870	0.0591	.
	TimeOutsideHome	0.0641	0.0126	5.0537	0.0000	***
	TourDelay (Modified)	0.0772	0.0207	3.7204	0.0001	***
	Overlap	7.7558	0.6255	12.3979	0.0000	***
	Log-Likelihood:	-2309.8				
	McFadden R^2:	0.044114				
	Likelihood ratio test :	chisq = 213.19				

Table 4: Model Results before Goal Programming Process. Significance code(p-value threshold, label): 0 ”***”, 0.001 ”**”, 0.01 ”*”, 0.05 ”.”, 0.1 ” ”

University Transportation Research Center - Region 2
Funded by the U.S. Department of Transportation

Region 2 - University Transportation Research Center
The City College of New York
Marshak Hall, Suite 910
160 Convent Avenue
New York, NY 10031
Tel: (212) 650-8050
Fax: (212) 650-8374
Website: www.utrc2.org