Center for Teaching Old Models New Tricks (TOMNET)

A USDOT Tier 1 University Transportation Center

PROJECT PROPOSAL: 2018-2019

Title: Latent Vehicle Type Propensity Segments: Considering the Influence of Household Vehicle Fleet Structure

Principal Investigator: *Patricia Mokhtarian*, Susan G. and Christopher D. Pappas Professor, School of Civil and Environmental Engineering, Georgia Institute of Technology

Introduction/Problem Statement

Understanding vehicle type propensities and choices is of interest to academics and practitioners in a wide array of fields. For example, market researchers may study vehicle type choices to predict consumer purchase behaviors and future market shares (Train and Winston, 2007), while energy researchers study individuals' vehicle type preferences and corresponding driving habits to calculate energy consumption and emissions (Gao et al., 2019). Transportation scholars traditionally study vehicle type to understand and forecast individual and household travel behaviors (Bhat and Sen, 2006), while in recent times, there has been a proliferation of vehicle type studies intended to model the adoption of emerging transport technologies such as electric and automated vehicles (Higgins et al., 2017; Mocanu, 2018). In this study, we propose to investigate vehicle type from a travel behavior perspective, identifying segments with the aim of understanding how personal and household mobility needs, along with a novel range of individual- and household-level characteristics, attitudes, and behaviors, influence vehicle type propensities. Based on the developed model, we will further examine the relationships between vehicle type propensities, gender roles, attitudes, and current and future travel behavior choices/interests, focus areas that can have policy implications in transportation.

A substantial body of literature has classified vehicle type, using a variety of deterministic schemes. Examples of individual attributes used for vehicle type classification include vehicle size (Lave and Train, 1979), body type (Cao et al., 2006), fuel type (Hoen and Koetse, 2014), and make/model (Østli et al., 2017). Other studies have combined attributes and developed mixed classification schemes (Baltas and Saridakis, 2013). Typically, individuals are then deterministically classified on the basis of the type of the vehicle they drive most often.

The proposed project focuses on classifying people, based on the types of vehicles for which they are the main driver, but it (1) also takes into account the entire household fleet of vehicles; (2) draws on a wide range of covariates to portray the kinds of people in each segment; and (3) uses a probabilistic clustering approach, latent class cluster analysis (LCCA). LCCA offers some potential advantages over deterministic approaches. For one thing, statistical criteria to aid in identifying an optimal number of clusters are built into the method (Vermunt and Magidson, 2002). Further, due to the structure of the model, the resultant latent clusters may be more homogeneous than deterministic classifications. For all of the above reasons, we believe that the LCCA model developed in this study could provide new insights into vehicle type propensity segments in the population.

1. Project Objective

The primary objective of the proposed project is to apply latent class cluster analysis to a sample of survey respondents in Georgia, identifying naturally occurring vehicle type segments based on the influence of both individual vehicle type choices and household vehicle fleet structures. Vehicle ownership is a key behavioral indicator, of which *vehicle type choice* is an important subarea of interest due to its important role in an array of fields, ranging from consumer forecasting to energy consumption, emission modeling, and travel behavior analysis, among others. Previous studies have shown that users choosing the same vehicle type have a discernible tendency to share similar characteristics (e.g., age, gender, income). In transportation, generating and understanding vehicle type user profiles can provide key insights for transport supply and demand modeling, as well as urban planning and policy making processes.

A secondary objective is to demonstrate the value of fusing multiple datasets to capitalize on the unique information offered by each.

2. Proposed Methodology and Data

The dataset to be used in this study will be a novel compilation of multiple data sources. Specifically, the study sample will comprise Georgia residents who responded to both the 2017 National Household Travel Survey (NHTS) and the Georgia Department of Transportation Emerging Technologies Survey (GDOT survey). The NHTS is a nationwide travel survey (2017 National Household Travel Survey, U.S. Department of Transportation, Federal Highway Administration), with rich behavioral data that includes a travel diary capturing respondents' activity patterns during a randomly selected day (including both weekdays and weekends across the whole sample). The GDOT survey is a statewide survey conducted on behalf of the Georgia Department of Transportation in 2017-18 (Kim et al., 2019), and is attitudinally-rich with an emphasis on the impacts of emerging technologies on travel behavior in Georgia. In addition to the survey data sources, we will augment each individual record in the dataset with targeted marketing data (TMD) purchased from a commercial data compiler, as well as with land use variables derived from respondents' residential locations. TMD includes variables such as sociodemographic characteristics, consumer behaviors and propensities, financial information, technology usage, and transport-related attributes (Shaw et al., 2020).

Methodologically, the proposed study will use latent class cluster analysis (LCCA) to identify vehicle type profiles. **FIGURE 1** shows the graphical representation of the LCCA model framework, which includes two sub-models: the membership model and the measurement model. In LCCA, the membership and measurement models are estimated simultaneously, enabling both the indicators and the covariates to influence cluster development. Eqs. 1-3 are the mathematical representation of the LCCA model (following the notation in Vermunt and Magidson, 2016).

$$P(\boldsymbol{y}_i|\boldsymbol{z}_i) = \sum_{k=1}^{K} P(k|\boldsymbol{z}_i)\, P(\boldsymbol{y}_i|k) \tag{1}$$

Eq. 1 represents the probability of observing a vector of indicators $\boldsymbol{y}_i$ for individual i, given a particular vector of covariates $\boldsymbol{z}_i$. The equation shows how unobserved latent class membership k, which has K categories, intervenes between the observed $\boldsymbol{y}_i$ and $\boldsymbol{z}_i$. Specifically, $P(k|\boldsymbol{z}_i)$ is the membership probability for a certain latent class k given the observed covariates $\boldsymbol{z}_i$, and $P(\boldsymbol{y}_i|k)$ is the conditional probability of $\boldsymbol{y}_i$ given the latent class k. The next two equations respectively define the constituent probability models of Eq. 1.

$$P(k|\boldsymbol{z}_i) = \frac{\exp\left(\gamma_{k0} + \sum_{r=1}^{R} \gamma_{kr}\, z_{ir}\right)}{\sum_{k'=1}^{K} \exp\left(\gamma_{k'0} + \sum_{r=1}^{R} \gamma_{k'r}\, z_{ir}\right)} \tag{2}$$

Eq. 2 represents the probability that individual i belongs to latent class k given the covariates $\boldsymbol{z}_i$, which is parameterized using the multinomial logit formula. For each latent class, LCCA estimates one intercept γ_{k0} and a set of parameters γ_{kr} corresponding to the R active covariates.

$$P(\boldsymbol{y}_i = \boldsymbol{m}|k) = \frac{\exp\left[\sum_{t=1}^{6}\left(\beta_{m_t 0}^{t} + \beta_{m_t k}^{t}\right) + \sum_{t<t'} \beta_{m_t m_{t'}}^{tt'}\right]}{\sum_{\boldsymbol{m'} \in M} \exp\left[\sum_{t=1}^{6}\left(\beta_{m'_t 0}^{t} + \beta_{m'_t k}^{t}\right) + \sum_{t<t'} \beta_{m'_t m'_{t'}}^{tt'}\right]} \tag{3}$$

Eq. 3 represents the joint probability of the six dichotomous indicators, also parameterized using the multinomial logit formula. Vectors $\boldsymbol{m}$ and $\boldsymbol{m'}$ represent specific combinations of indicators taking on the values 0 and 1, both of which belong to the set M, which contains all possible indicator value combinations (specifically, $2^6 = 64$ possible combinations). Thus, the numerator pertains to a single particular combination of six 0s and 1s (the vector $\boldsymbol{m}$), while the denominator sums over all 64 such combinations. Finally, for each latent class k, $\beta_{m_t k}^{t}$ represents the class-specific deviation from the average propensity. In this study, t belongs to {1=main: car, 2=main: SUV/van, 3=main: truck, 4=other: car, 5=other: SUV/van, 6=other: truck}.

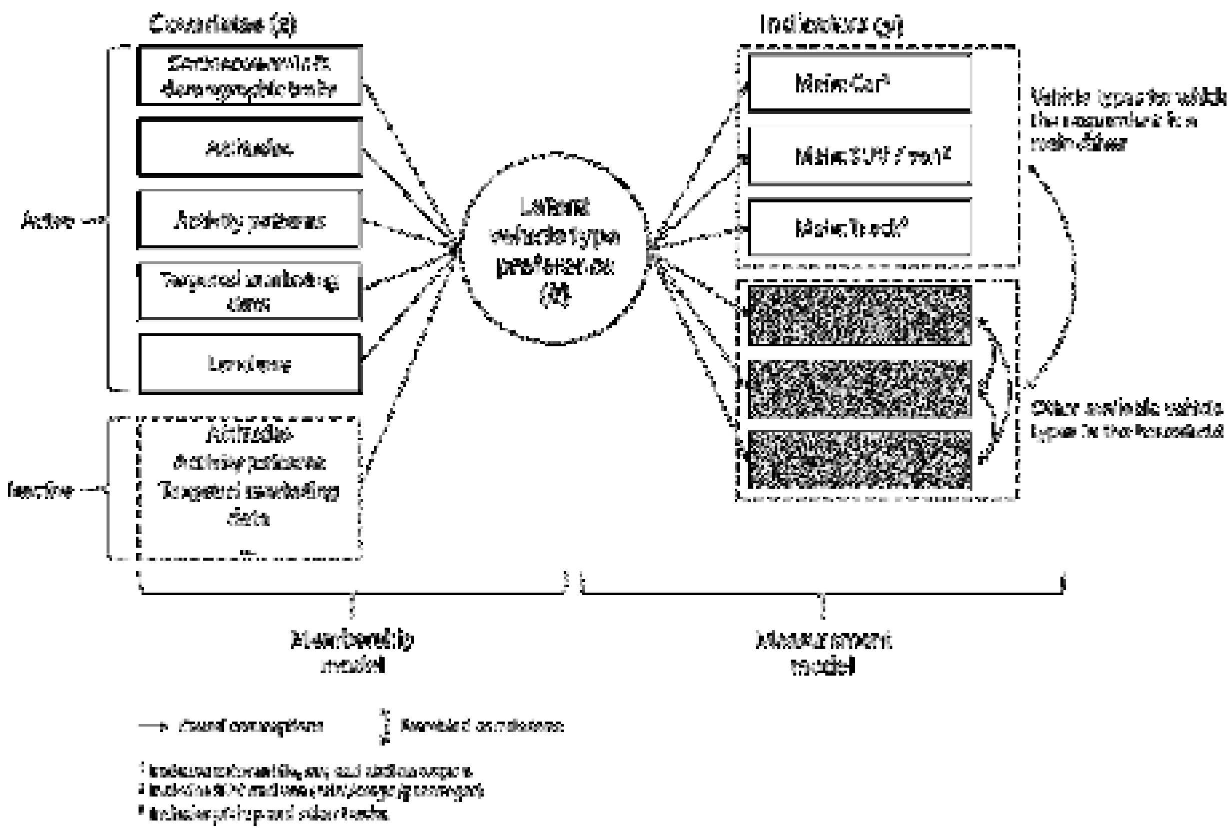

FIGURE 1. Model framework of the latent class cluster analysis (LCCA)

3. Work Plan (Project Tasks)

The proposed project can be divided into the following tasks.

Task 1: Literature review.

We will monitor the literature on vehicle type choice and latent class cluster analysis for the duration of the project. In addition, we will follow literature in related areas, including the use of targeted marketing data in academic research and the incorporation of attitudes in travel demand models.

Task 2: Assemble the dataset.
As mentioned above, the study sample will represent the fusion of a number of sources. The individuals in the sample consist of Georgia drivers who responded to both the 2017 National Household Travel Survey (NHTS) and the Georgia Department of Transportation Emerging Technologies Survey (GDOT survey). Accordingly, responses to both surveys can readily be joined. In addition, however, we will augment each individual record in the dataset with targeted marketing data (TMD) purchased from a commercial data compiler, as well as with land use variables derived from respondents' residential locations. Joining these two sources of data with the survey responses is a non-trivial task. Once the data has been assembled, we will develop case weights to obtain results more representative of the population of Georgia drivers.

Task 3: Find and interpret the best latent class cluster structure for the data.
We will investigate LCCA models with varying numbers of classes, and use the standard Bayesian Information Criterion (BIC) to identify the best number of classes. We will analyze the segment-specific distributions of *indicators* to reveal the latent vehicle type propensities, and based on these we will interpret and name each segment. Then, we will develop and discuss segment profiles based on the segment-specific distributions of *covariates*.

Task 4: Document the results for dissemination.
We will prepare a paper for submission to a high-quality peer-reviewed journal. In addition, we plan to present the paper at one or more professional conferences (see Section 7).

4. Project Schedule

		Month											
	Task Name	1	2	3	4	5	6	7	8	9	10	11	12
1	Literature review												
2	Assemble the dataset												
3	Find and interpret the best latent class cluster structure for the data												
4	Document the results for dissemination												

5. Relevance to the Center Theme/Mission

The proposed project effectively contributes to the TOMNET mission, by exploring ways of fusing data from multiple sources to enrich the insight that can be obtained from any single source. Also very much in keeping with the TOMNET mission, the project will be exploring the use of machine learning methods to reduce the dimensionality of the set of variables available through fusion. The experience gained through this project will inform other TOMNET projects as well.

6. Anticipated Outcomes and Deliverables

In terms of ***research outcomes/benefits*** of this study, the novel compilation of data sources and the application of a data-driven vehicle type classification approach will provide new insights into relationships between vehicle type propensities and a wide range of general and travel-related attributes that have heretofore not been simultaneously studied. For example, we will investigate how attitudes toward battery electric vehicles (BEVs) differ by latent class.

With respect to ***tangible outcomes/deliverables***, we expect to produce a paper to be submitted to a peer-reviewed journal, and also to present this work at one or more conferences, such as the Women's Issues in Transportation conference in September 2019, and the 2020 Annual Meeting of the Transportation Research Board.

7. Research Team and Management Plan

Principal Investigator (PI) ***Patricia Mokhtarian*** is an internationally-known travel behavior scholar, who has specialized in measuring and modeling attitudes and incorporating them into models of travel-related behaviors. She will be responsible for the overall direction of the project, and will be directly engaged with its ongoing progress. A one-page CV for her appears after the budget. In addition, ***one PhD student*** will be responsible for the day-to-day execution of substantive project tasks, occasionally assisted by another PhD student working on related projects.

The project team will meet weekly for in-depth reports on progress and tactical planning. All members are local, so communication will be straightforward, of course supplemented by e-mail during inevitable absences. An internal project website will be set up as a working repository for literature, presentations prepared by the project, data, and analyses. Milestone products, including papers, presentations, and reports will be provided to the central TOMNET site.

8. Technology Transfer Plan

The project PI has a proven track record of scholarly productivity and research dissemination. In July 2019, we will prepare a paper to be submitted for presentation at the Annual Meeting of the Transportation Research Board in January 2020, and for publication in a peer-reviewed journal. Based on past history, we expect multiple opportunities to present project findings throughout the life of the study and beyond, and we will seek out and volunteer for such opportunities as appropriate.

In addition, to disseminate the work among practitioners, we will present the study at the annual Georgia Transportation Institute Research Expo, held at a central Atlanta location of the Georgia Department of Transportation.

9. Workforce Development and Outreach Plan

Regarding the future of our workforce, the proposed project (whose PI is a woman) will contribute heavily to the professional development of at least two PhD students (both women). The PI is devoted to the careful mentoring of female graduate students, including with respect to career-life balance, a major reason why female PhD students do not choose academia (Mason et al., 2009). Research has shown that mentoring and positive role models can make a big difference in the attraction of women to STEM fields (Hill et al., 2010).

Regarding the development of the current workforce, as described in Section 9, we will present the work at a practitioner-oriented meeting in Georgia. This presentation will inform agency staff who are predominantly engineers, about the importance of attitudinal variables to our understanding of travel behavior, as well as demonstrate how to obtain and incorporate such variables..

10. References

Baltas, G., & Saridakis, C. (2013). An empirical investigation of the impact of behavioural and psychographic consumer characteristics on car preferences: An integrated model of car type choice. *Transportation Research Part A, 54*, 92-110.

Bhat, C.R., & Sen, S. (2006). Household vehicle type holdings and usage: an application of the multiple discrete-continuous extreme value (MDCEV) model. *Transportation Research Part B, 40*(1), 35-53.

Cao, X., Mokhtarian, P.L., & Handy, S.L. (2006). Neighborhood design and vehicle type choice: Evidence from Northern California. *Transportation Research Part D, 11*, 133-145.

Gao, Z., Laclair, T., Ou, S., Huff, S., Wu, G., Hao, P., Boriboonsomsin, K. & Barth, M. (2019). Evaluation of electric vehicle component performance over eco-driving cycles. *Energy, 172*, 823-839.

Higgins, C.D., Mohamed, M., & Ferguson, M.R. (2017). Size matters: How vehicle body type affects consumer preferences for electric vehicles. *Transportation Research Part A, 100*, 182-201.

Hill, C., Corbett, C., & St. Rose, A. (2010). *Why so few? Women in science, technology engineering and mathematics*. American Association of University Women. http://www.aauw.org/files/2013/02/Why-So-Few-Women-in-Science-Technology-Engineering-and-Mathematics.pdf

Hoen, A., & Koetse, M.J. (2014). A choice experiment on alternative fuel vehicle preferences of private car owners in the Netherlands. *Transportation Research Part A, 61*, 199-215.

Kim, S.H., Mokhtarian, P.L., & Circella, G. (2019). *The Impact of Emerging Technologies and Trends on Travel Demand in Georgia*. Final Report for project 16-31 of the Georgia Department of Transportation, Atlanta, GA, October. Available at http://g92018.eos-intl.net/G92018/OPAC/Index.aspx

Lave, C.A., & Train, K. (1979). A disaggregate model of auto-type choice. *Transportation Research Part A, 13*, 1-9.

Mason, M.A., Goulden, M., & Frasch, K. (2009). Why graduate students reject the fast track. *Academe 95*(1), 11-16.

Mocanu, T. (2018). What types of cars will we be driving? Methods of forecasting car travel demand by vehicle type. *Transportation Research Record*, *2672*(49), 125-134.

Østli, V., Fridstrøm, L., Johansen, K.W., & Tseng, Y.-Y. (2017). A generic discrete choice model of automobile purchase. *European Transport Research Review, 9*(2), 16.

Shaw, F.A., Wang, X., Mokhtarian, P.L., & Watkins, K.E. (2020, January 15). *Targeted marketing data as a transportation data source: Applications, integration, and validation.* Paper #20-01237 presented at the 99th Annual Meeting of the Transportation Research Board, Washington DC.

Train, K. E., & Winston, C. (2007). Vehicle choice behavior and the declining market share of US automakers. *International Economic Review, 48*, 1469-1496.

U.S. Department Of Transportation, Federal Highway Administration, *2017 National Household Travel Survey*. https://nhts.ornl.gov

Vermunt, J.K., & Magidson, J. (2002). Latent class cluster analysis. *Applied Latent Class Analysis, 11*, 89-106.

Vermunt, J.K., & Magidson, J. (2016). *Technical Guide for Latent GOLD 5.1: Basic, Advanced, and Syntax,* Belmont, MA, Statistical Innovations Inc.

11. Qualifications of Investigator

PATRICIA L. MOKHTARIAN

Susan G and Christopher D Pappas Professor, School of Civil & Environmental Engineering
Georgia Institute of Technology, Atlanta, GA 30332-0355 Email: patmokh@gatech.edu

Education

PhD, Industrial Engineering/Management Sciences, Northwestern University, 1981
MS, Industrial Engineering/Management Sciences, Northwestern University, 1977
BA (summa cum laude), Mathematics, Florida State University, 1975

Employment and Professional Experience (last 25 years)

Susan G and Christopher D Pappas Professor (2016-present) / Professor (2013-2016), School of Civil & Environmental Engineering, Georgia Institute of Technology

Full (1999-2013)/Associate (1996-1999)/Assistant (1990-1996) Professor, Department of Civil & Environmental Engineering, University of California, Davis

Chair and Graduate Adviser (1997-2013), Interdisciplinary Graduate Group in Transportation Technology and Policy, University of California, Davis

Acting Director (1999-2000) / Associate Director for Education (2001-2013), Institute of Transportation Studies, University of California, Davis

Fields of Interest and Expertise

(1) Attitude measurement and survey design; (2) Statistical/econometric analysis of transportation data; (3) Impacts of information/communications technology on travel; (4) Attitudes toward travel; (5) Activities conducted while traveling; (6) Impacts of the built environment on travel behavior

5 Recent Relevant Publications (not already cited in the proposal)

Mokhtarian, P.L. (in press). The times they are a-changin': What do the expanding uses of travel time portend for policy, planning, and life? *Transportation Research Record*.

Kim, S.H., & Mokhtarian, P.L. (2018). Taste heterogeneity as an alternative form of endogeneity bias: Investigating the attitude-moderated effects of built environment and socio-demographics on vehicle ownership using latent class modeling. *Transportation Research A, 116*, 2018, 130-150.

Mishra, G.S., Mokhtarian, P.L., Clewlow, R.R., & Widaman, K.F. (in press). Addressing the joint occurrence of self-selection and simultaneity biases in the estimation of program effects based on cross-sectional observational surveys - Case study of travel behavior effects in carsharing. Online First, *Transportation*. **http://dx.doi.org/10.1007/s11116-017-9791-1**

Lee, R.J., Sener, I.N., Mokhtarian, P.L., & Handy, S.L. (2017). Relationships between the online and in-store shopping frequency of Davis, California residents. *Transportation Research A, 100*, 40-52.

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*(5), 558-584.

Mokhtarian, P.L., & van Herick, D. (2016). Quantifying residential self-selection effects: A review of methods and findings from applications of propensity score and sample selection approaches. *Journal of Transport and Land Use, 9*(1), 7-26.

Graduate Student Supervision/Advising

Graduated: 12 PhDs (including 2 women), 23 MSs (6); **Current (Co-)Supervision**: 6 PhDs (2)

Recent Honors and Awards

Invited speaker, endowed or distinguished/eminent lecture series, 7 occasions (2014-2018)
Invited keynote speaker at 6 international conferences (2014-2017)
Sustained Research Award, School of Civil and Environmental Engineering, Georgia Tech (2015)

12.Budget Including Non-Federal Matching Funds

Institution: Georgia Institute of Technology

Project Title: Latent Vehicle Type Propensity Segments: Considering the Influence of Household Vehicle Fleet Structure

Principal Investigator: Patricia L. Mokhtarian

Budget Period: 8/1/2018 - 07/31/2019

CATEGORY	Budgeted Amount from Federal Share	Budgeted Amount from Matching Funds	Explanatory Notes; Identify Source of Matching Funds
Faculty Salaries	$9,818	$21,500	Georgia Tech faculty salary
Other Staff Salaries	–		
Student Salaries	$27,240		
Fringe Benefits	$4,696	$6,407	Georgia Tech faculty overhead
Total Salaries & Benefits	$41,754	$27,907	
Student Tuition Remission	$19,224		
Operating Services and Supplies	$250		
Domestic Travel	$1,500		
Other Direct Costs (specify)			
Other Direct Costs (specify)			
Total Direct Costs	$62,728	$27,907	
F&A (Indirect) Costs	$25,145	$16,130	Georgia Tech faculty salary
TOTAL COSTS	$87,873	$44,037	

Grant Deliverables and Reporting Requirements for UTC Grants (November 2016)
Exhibit F

UTC Project Information	
Project Title	Latent Vehicle Type Propensity Segments: Considering the Influence of Household Vehicle Fleet Structure
University	Georgia Institute of Technology
Principal Investigator	Patricia L. Mokhtarian
PI Contact Information	patmokh@gatech.edu, 404-385-1443
Funding Source(s) and Amounts Provided (by each agency or organization)	TOMNET, $87,873 Georgia Tech, $44,037
Total Project Cost	$131,910
Agency ID or Contract Number	
Start and End Dates	August 1, 2018 - July 31, 2019
Brief Description of Research Project	Fuse multiple datasets, including those pertaining to two surveys (the National Household Travel Survey) completed by the same Georgia respondents, together with targeted marketing and land use data associated with those respondents. Apply latent class cluster analysis to the data, to identify naturally occurring vehicle type segments based on the influence of both individual vehicle type choices and household vehicle fleet structures.
Describe Implementation of Research Outcomes (or why not implemented)	TBD
Impacts/Benefits of Implementation	TBD
Web Links • Reports • Project Website	TBD