NATIONAL CENTER FOR UNDERSTANDING FUTURE
TRAVEL BEHAVIOR AND DEMAND

Final Project Report

How Much Do Attitudinal Variables Improve Travel Demand Models? Evaluation Using an Overlap Sample from an Attitude-rich Survey and the 2017 National Household Travel Survey

BY

Ilsu Kim
Email: ikim302@gatech.edu

Patricia L. Mokhtarian
Email: patmokh@gatech.edu

School of Civil and Environmental Engineering
Georgia Institute of Technology
790 Atlantic Drive
Atlanta, GA 30332

April 2025

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
N/A	N/A	N/A

4. Title and Subtitle	5. Report Date
How much do attitudinal variables improve travel demand models? Evaluation using an overlap sample from an attitude-rich survey and the 2017 National Household Travel Survey	April 30, 2025
	6. Performing Organization Code N/A
7. Author(s) Ilsu Kim, https://orcid.org/0000-0001-6517-363X Patricia L. Mokhtarian, PhD, NAE, https://orcid.org/0000-0001-7104-499X	**8. Performing Organization Report No.** N/A
9. Performing Organization Name and Address School of Civil and Environmental Engineering Georgia Institute of Technology 790 Atlantic Drive, Atlanta, GA 30332	**10. Work Unit No. (TRAIS)** N/A
	11. Contract or Grant No. 69A3552344815 and 69A3552348320
12. Sponsoring Agency Name and Address U.S. Department of Transportation, University Transportation Centers Program, 1200 New Jersey Ave, SE, Washington, DC 20590	**13. Type of Report and Period Covered** Final Report, 2024-2025
	14. Sponsoring Agency Code USDOT OST-R

15. Supplementary Notes
This report is currently undergoing peer review at a journal, under the same title.

16. Abstract
This study aims to evaluate the effectiveness of adding a handful of attitudinal marker statements to transportation surveys (instead of designing, deploying, and factor-analyzing a full set of attitudinal variables). We exploit the rare opportunity offered by the 2017 Georgia Department of Transportation (GDOT) Emerging Technologies (ET) survey and the 2017 Georgia add-on to the National Household Travel Survey (NHTS) having 1,245 respondents in common. The non-overlap GDOT ET survey dataset (N = 2,043) is selected as the *donor* sample, based on which elastic net regression (ENR) models are trained for imputation of attitudinal factor scores using marker variables (MVs). The overlap NHTS dataset (i.e., the *recipient* sample) (N = 1,245) is treated as if it has only MVs, with attitude scores needing to be imputed using the ENR models trained on the *donor* sample. The ENR models display high prediction performance in both the *donor* and *recipient* datasets, while MVs present excellent performance as well. Three travel behavior variables in the *recipient* dataset are modeled with no attitudes, predicted attitude scores, and MVs: household vehicle count, (personal yearly) vehicle miles driven, and hybrid/electric vehicle adoption. For each dependent variable, several attitudes show statistical significance, although their contributions to model fit vary. The results indicate that including attitudes leads to (a) better prediction of less-common alternatives (zero vehicles and hybrid/electric vehicle adoption), primarily by improving the prediction of the groups most likely to select such alternatives, and (b) discovery of additional *non-attitude* variables that would have been considered insignificant otherwise.

17. Key Words	18. Distribution Statement
Marker variables, attitude imputation, overlap sample, machine learning, travel behavior models	No restrictions.

19. Security Classif.(of this report)	20. Security Classif.(of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	37	N/A

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

DISCLAIMER

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated in the interest of information exchange. The report is funded, partially or entirely, under Grant No. 69A3552344815 and 69A3552348320 from the U.S. Department of Transportation's University Transportation Centers Program. The U.S. Government assumes no liability for the contents or use thereof.

ACKNOWLEDGMENTS

This research was partially supported by the National Center for Understanding Future Travel Behavior and Demand (TBD), a National University Transportation Center sponsored by the U.S. Department of Transportation (USDOT) under grant numbers 69A3552344815 and 69A3552348320. The authors would like to thank the TBD National Center and USDOT for their support of university-based research in transportation, particularly for the funding provided for this project. Conversations with and support from Drs. Jason Soria, Atiyya Shaw, and Sung Hoo Kim have provided considerable input to the analysis.

TABLE OF CONTENTS

EXECUTIVE SUMMARY 1
INTRODUCTION 2
DATA AND METHODS 4
Terminology 4
Survey data 4
Exploratory factor analysis (EFA) 5
1.1 Attitude imputation 6
Evaluation of the predicted attitudes and MVs 8
RESULTS 8
Internal evaluation 8
External evaluation 11
Model estimation 11
No-attitude and predicted-attitude models 11
Marker-variable models 15
1.1.1 3.2.4. Segmentation analysis 19
DISCUSSION AND CONCLUSIONS 23
REFERENCES 26
Appendix A Descriptive statistics of dependent and explanatory variables 30
Appendix B Distributions of vehicle count by household income and education level 32
Appendix C Hybrid/electric vehicle adoption by household income and education level 32
Appendix D Crosstabulation of household income and education 32

LIST OF TABLES

Table 1 Attitudinal factors and associated items 7
Table 2 Vehicle ownership MNL models (in R 4.3.3 with *mlogit* package) 17
Table 3 Log-linear VMD models (estimated in R 4.3.3) 18
Table 4 Hybrid/electric vehicle adoption BL models (in R 4.3.3 with *mlogit* package)....................18

LIST OF FIGURES

Figure 1 Donor and recipient datasets 5
Figure 2 Correlations of observed attitudes (a) with predicted attitudes and (b) with marker variables..10
Figure 3 Comparison of log-likelihoods of hybrid/electric vehicle adoption models 15
Figure 4 Vehicle ownership models: percent correctly classified values 21
Figure 5 Hybrid/electric vehicle adoption models: percent correctly classified values....................22

EXECUTIVE SUMMARY

This study aims to evaluate the effectiveness of adding a handful of attitudinal marker statements to transportation surveys (instead of designing, deploying, and factor-analyzing a full set of attitudinal variables). We exploit the rare opportunity offered by the 2017 Georgia Department of Transportation (GDOT) Emerging Technologies (ET) survey and the 2017 Georgia add-on to the National Household Travel Survey (NHTS) having 1,245 respondents in common. The non-overlap GDOT ET survey dataset (N = 2,043) is selected as the *donor* sample, based on which elastic net regression (ENR) models are trained for imputation of attitudinal factor scores using marker variables (MVs). The overlap NHTS dataset (i.e., the *recipient* sample) (N = 1,245) is treated as if it has only MVs, with attitude scores needing to be imputed using the ENR models trained on the *donor* sample. The ENR models display high prediction performance in both the *donor* and *recipient* datasets, while MVs present excellent performance as well. Three travel behavior variables in the *recipient* dataset are modeled with no attitudes, predicted attitude scores, and MVs: household vehicle count, (personal yearly) vehicle miles driven, and hybrid/electric vehicle adoption. For each dependent variable, several attitudes show statistical significance, although their contributions to model fit vary. The results indicate that including attitudes leads to (a) better prediction of less-common alternatives (zero vehicles and hybrid/electric vehicle adoption), primarily by improving the prediction of the groups most likely to select such alternatives, and (b) discovery of additional *non-attitude* variables that would have been considered insignificant otherwise.

INTRODUCTION

Facing various factors affecting travel behavior, including altered work/commute patterns and advancements in transportation technologies, travel demand models are, and will continue to be, in need of enhanced prediction performance. One well-established way (in the *academic* literature) to better understand and predict travel behavior is to include attitudes when modeling transportation-related choices (e.g., travel mode choice (Domarchi et al., 2008; Popuri et al., 2011), eating out (Haddad et al., 2023), car-sharing membership (Becker et al., 2017), and car ownership (Cao et al., 2007; Wu et al., 1999)). This enables explaining behavior more completely and meaningfully, as well as simulating scenarios involving changes in attitudes. However, *practice-oriented* travel demand models have not incorporated attitudes because of the challenges associated with measuring and forecasting them (Mokhtarian, 2024). If in-practice models are to incorporate attitudes, it is imperative to develop practical approaches for doing so, and to clearly demonstrate the effectiveness of such approaches.

A major challenge, however, is that historically, measuring attitudes has typically involved designing and deploying dozens of attitudinal statements in a survey (generally three or more items per construct), and then extracting attitudinal factors or constructs using techniques such as exploratory factor analysis (EFA) (Rummel, 1970; Hatcher, 1994). This is naturally a non-starter for the already burdensome household travel surveys that underpin most practice-oriented travel demand forecasting models. Thus, one clear way to lighten the burden is to use an abbreviated set of attitudinal statements instead of the full set. The basic idea is to use only *key (or marker)* items (one or two per construct) while retaining an acceptable level of measurement fidelity. An extensive body of research has explored and validated the idea of reducing the number of items per construct across various types of measurements: personality traits (Gosling et al., 2003; Rammstedt & John, 2007), depression scales (Burisch, 1997), materialism scales (Richins, 2004), love attitudes (Hendrick et al., 1998; Morrow et al., 1995; Thompson & Borrello, 1987), health conditions (Goldberg & Hillier, 1979; Tambs & Moum, 1993; Ware et al., 1996), difficulties in emotion regulation (Kaufman et al., 2016), pedestrian streetscapes (Cain et al., 2017; Sallis et al., 2015), and neighborhood walkability (Cerin et al., 2006, 2009; Silveira & Motl, 2020). Stepping even farther away from the central tendency of conventional practice, some researchers have adopted and advocated using single-item measures due to their efficiency in measurement as well as in data cleaning and analysis (see Section 1.1 of Gosling et al. (2003) for examples such as Robins et al. (2001) and Sandvik et al. (1993)). This is the approach taken by the present study.

This investigation is prompted by a line of research that has recently been launched on attitude imputation into household travel surveys, using machine learning (ML) functions trained on variables common to two survey datasets (Mokhtarian, 2024). It was discovered that using *selected* attitudinal variables (i.e., *marker* variables, or MVs) as the common variables for imputation (Shaw, 2021; Soria, 2023; Soria & Mokhtarian, 2024) far outperforms using socio-economic and demographic (SED) and land-use variables (Malokin et al., 2019), and targeted marketing variables (Shaw, 2021). The central idea is to use one survey dataset (the "*donor* sample", containing *a full set* of attitudinal variables as well as attitudinal factor scores from EFA) to train ML functions that predict attitudinal factor scores using a skeletal set of MVs as explanatory variables or "features" (see Section 2.1), and then apply those functions to another dataset (the "*recipient* sample") that contains the same MVs, to impute attitude scores into it. This allows attitudinal information to be attached to the respondents in the recipient dataset without measuring the *whole* set of attitudinal variables used to reveal the attitudinal constructs in the donor dataset.

Using MVs for attitude imputation has produced high correlations between *observed* and *predicted (or imputed)* attitudes (an assessment referred to as *internal* evaluation) (Shaw, 2021; Soria, 2023; Soria & Mokhtarian, 2024). However, examination of the contributions of such predicted attitudes to travel behavior modeling (i.e., *external* evaluation) has so far been limited. Prior external evaluation efforts of MV-imputed attitudes primarily focused on the model "lift" obtained from the inclusion of imputed attitudes, with less attention given to the statistical significance of the included attitudes and other explanatory variables (Shaw, 2021; Soria & Mokhtarian, 2024).

To address this research gap, for each travel behavior variable selected from the *recipient* dataset for *external* evaluation, this study compares the best models with and without predicted attitudes. The best models are obtained after exploring various model specifications to investigate changes in the statistical significance of the impacts of other explanatory variables when predicted attitudes are included. In addition, further exploring the potential of MVs themselves as explanatory variables (briefly tested by Soria and Mokhtarian (2024)), the efficacy of using MVs directly (instead of *predicted* attitudes) is examined. Lastly, the model fit improvements are evaluated across various SED status groups to better understand the sources of overall improvement.

In addition to more systematic *external* evaluation efforts, this study has at least two additional distinctions from previous research. First, it is the first (to our knowledge) to employ a separate recipient dataset from another survey for the external evaluation of MV-imputed attitudes[1], whereas prior related studies utilized only one survey for evaluation (Shaw, 2021, Section 5.3; Soria, 2023; Soria & Mokhtarian, 2024). Second, moving one step forward from the external evaluation in Section 5.3 of Shaw (2021), which used only one model form (i.e., linear regression) to model a few travel behavior variables and included all attitudinal factors in the models regardless of their significance (for simplicity, given that the study was for a proof of concept), this study adopts the model form suitable for each of the travel behavior variables to be modeled and prunes model specifications so that the final best model contains only statistically significant explanatory variables.

We expect that the results of this study will help governments and regional planning agencies assess the potential of adding MVs to their transportation surveys and employing attitudinal variables in practice-oriented travel behavior modeling for more effective transportation planning and decision-making. Also, this study will contribute to increasing our knowledge with respect to attitudinal measurement and the use of ML to inform causal models.

The remainder of this paper consists of three sections. Section 2 explains the data and methods used in this study, while Section 3 presents the main findings, first with respect to internal evaluations, and then regarding external evaluations. Section 4 provides some additional discussion and offers suggestions for future research.

[1] Such *out-of-sample* external evaluations have been conducted previously by Malokin et al. (2019) and Shaw (2021, Sections 5.1 and 5.2). However, in those cases the common variables for attitude imputation were not attitudinal markers but SED, residential location, and targeted marketing variables.

DATA AND METHODS

This section provides a comprehensive overview of the study's key concepts, data, and analytical methods, with five subsections. Section 2.1 defines key terms related to the measurement, identification, prediction, and evaluation of attitudes. Section 2.2 describes the datasets used in this study, emphasizing how this study leverages the unique opportunity of having many respondents who responded to two surveys. Section 2.3 details the process of identifying attitudinal factors and determining MVs, while Section 2.4 outlines the process of training elastic net regression (ENR) models for imputing attitudes. Section 2.5 illustrates how imputed attitudes are evaluated.

Terminology

In this study, "attitudinal *items*" indicate statements to which respondents react on a 5-point Likert-type scale ("strongly disagree" to "strongly agree"). The responses to the statements are referred to as "attitudinal *variables*" (ranging from -2 to 2, with 0 indicating "neutral"). A few dozen statements are strategically designed and deployed to measure a smaller number of attitudinal *constructs* (i.e., *factors*), which EFA empirically identifies based on the correlation patterns of responses (Rummel, 1970). *Marker* variables (MVs) are the one or two attitudinal variables that are most strongly associated with each factor and, thus, play a key role in predicting the scores on the factor. The attitudinal factor scores acquired directly from the factor solution (i.e., in our case by applying the Bartlett factor score coefficient matrix to the full set of attitudinal variables) are referred to as *observed* attitudes[2], while the scores imputed using the ML functions with MVs as inputs are termed *predicted* attitudes. The predicted attitudes and MVs are *internally* evaluated with respect to how well they reproduce the observed attitudes, and then *externally* evaluated by being tested in the downstream travel behavior models.

Survey data

This study employs datasets from two surveys: the 2017 Georgia Department of Transportation (GDOT) Emerging Technologies (ET) survey (an attitude-rich survey) (Kim et al., 2019) and the 2017 Georgia add-on to the National Household Travel Survey (NHTS) (Kash et al., 2021). However, it further exploits a distinctive feature not found in previous research along these lines (see Figure 1). Specifically, 1,245 people (in the *overlap* sample) responded to *both* surveys. Because those individuals responded to the former survey, we have responses to all 46 attitudinal items for each of them. By joining the attitudinal MVs from the overlap GDOT ET dataset to the overlap NHTS dataset, we can treat the latter dataset as the *recipient* dataset (N = 1,245) for which attitudinal factor scores are to be imputed. On the other hand, the non-overlap GDOT ET dataset (N = 2,043) becomes the *donor* dataset that is used to develop the EFA solution and train attitude imputation ML functions (see Sections 2.3 and 2.4 for full details). With this setup, we are essentially simulating a situation in which the (*non-overlap*) GDOT ET dataset is collected with a full set of attitudinal variables, and then a handful of MVs (identified in the donor dataset through EFA) are included in the NHTS to obtain the (*overlap*) NHTS dataset. However, the fact that we

[2] In reality, these scores are only estimates of the true, but unobserved, scores on the factor (also termed a "latent construct", for this reason). We refer to them as "observed" to avoid confusion with the "predicted" scores obtained from the ML process.

also have the full set of attitudinal variables (and thus can compute the “observed” factor scores) for the overlap sample means that in this instance we can additionally assess how well the transferred attitudes reproduce the “true” ones. In other words, we can also perform an internal evaluation on the recipient sample.

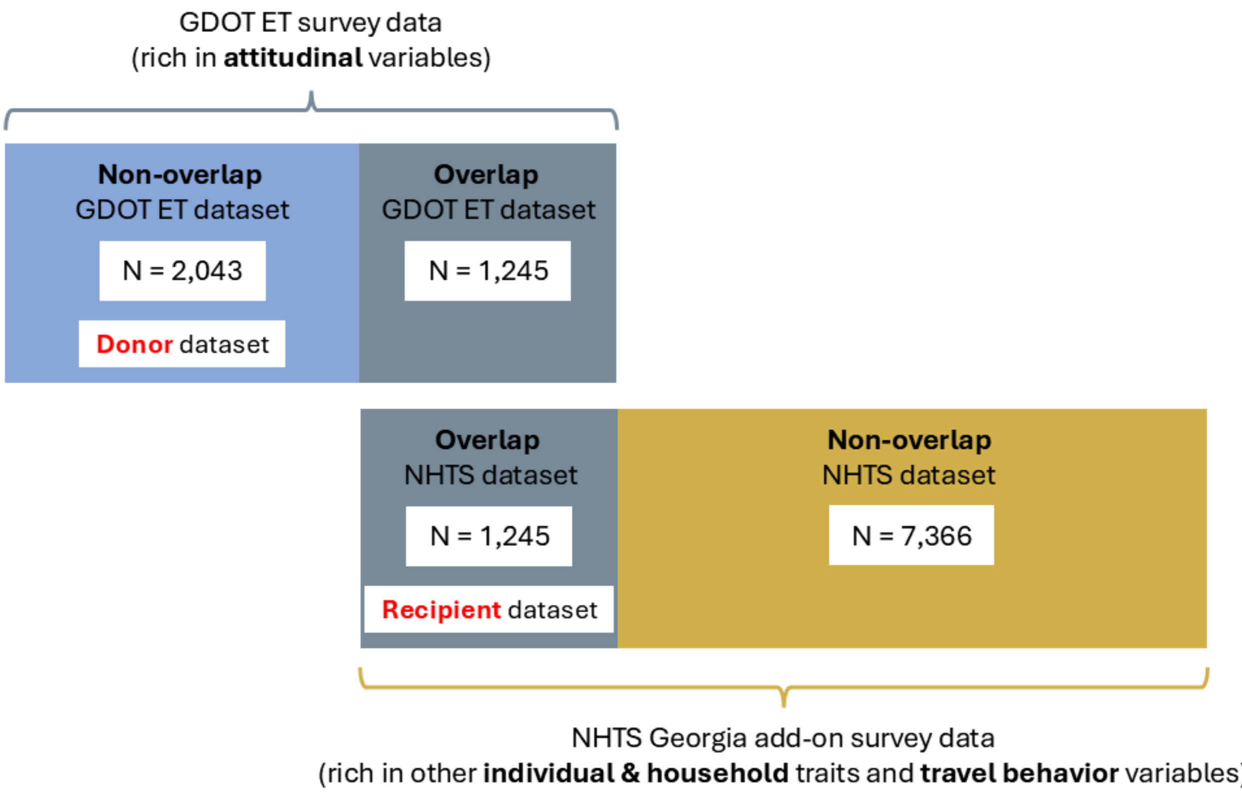


Figure 1 Donor and recipient datasets

Exploratory factor analysis (EFA)

To identify attitudinal factors and determine the MVs using the variables measured in the GDOT ET survey, EFA was implemented with the *donor* dataset. Given that Kim et al. (2019) obtained a common factor analysis solution after extensive exploration using the *full* GDOT ET dataset, the same specifications were adopted with respect to the variables included in the analysis (the 38 variables listed in Table 1), the number of factors (15 factors), and the factor rotation method (oblimin rotation with delta = 0). We reiterate, however, that for this study we performed the factor analysis only on the donor sample, not the full GDOT ET sample, in view of the fact that in the typical expected application of this approach, the donor and recipient samples would not overlap. Had we allowed them to overlap here, the imputation process would have been unfairly advantaged by the fact that 38% (1,245 of 3,288) of the cases used to create the factor score coefficients for the donor sample also constituted the recipient sample into which the imputed factor scores would be transferred.

Confirming the stability of the previous factor solution, we obtained a comfortingly similar solution (see Table 1 in comparison with Table 4-2 of Kim et al. (2019)), which enables giving the factors the same names. The list of statements with pattern loadings of magnitude 0.3 or greater for each factor did not change, with only one exception[3]. The highest-loading statements did not change for 12 out of the 15 factors, while those of the other three factors (*modern urbanite*, *pro-exercise*, and *sociable*) changed without altering the practical interpretations of the factors. Bartlett factor scores were calculated from the final solution, and added to the donor dataset.

1.1 Attitude imputation

For attitude imputation, this study adopts ENR[4]. This approach uses a regularization method of giving penalties to large coefficient estimates to prevent overfitting to the training set (i.e., the *donor* dataset), which is rational given that the prediction performance of imputation functions in the *recipient* dataset is of greater importance. In fact, Soria and Mokhtarian (2024) trained a random forest (RF) and an ENR model for each of four attitudes, and found that the differences in model performance between the training and test datasets were much smaller for ENR and that ENR models performed slightly better in the test dataset (than RF models).

In this study, the MV of an attitude is generally the variable based on the highest-loading item on that attitude (i.e., the bolded and italicized items in Table 1). For *modern urbanite*, however, the second highest-loading item is used because the top two items have practically the same pattern loadings (0.363 and 0.362), and the highest-loading item has a higher loading on another factor (0.460 on *materialistic*, compared to 0.363 on *modern urbanite*). All 15 MVs in Table 1 are used to train an ENR model for each attitude, using the *scikit-learn* library (version 1.1.1) in Python 3.10.14. The equation below governs how ENR coefficients are estimated in this study[5]:

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}}\left(\frac{1}{2n}\|X\beta - y\|_2^2 + \alpha\rho\|\beta\|_1 + \frac{\alpha(1-\rho)}{2}\|\beta\|_2^2\right) \qquad \text{(Equation 1)}$$

where n = donor sample size = 2,043, X = MVs of the donor sample, y = vector of observed attitudes, $\hat{\beta}$ = vector of estimated coefficients, α = hyperparameter determining the overall importance given to penalty terms, and ρ = hyperparameter associated with the relative weights on L1 and L2 penalties (ranging from 0 to 1 with $\rho = 0$ and $\rho = 1$ indicating ridge regression and lasso regression, respectively). To set hyperparameters for each model, a grid search with 10-fold cross-validation was conducted with $\alpha \in \{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1, 10, 10^{2}\}$ and $\rho \in \{0.05, 0.10, 0.15, \ldots, 0.95, 1.00\}$, which is the same grid used by Soria and Mokhtarian (2024).

[3] The pattern loading of the statement "I would/do enjoy having a lot of luxury things" on the *pro-suburban* factor changed from 0.361 to 0.294, which is only slightly below 0.3 (not shown in Table 1).

[4] Proposed by Zou and Hastie (2005), ENR combines the L1 penalty used in lasso regression (Tibshirani, 1996) and L2 penalty used in ridge regression (Hoerl & Kennard, 1970).

[5] Please refer to https://scikit-learn.org/stable/modules/linear_model.html#elastic-net for more details.

Table 1 Attitudinal factors and associated items

Factor	Statement [a]	Pattern loading [b]
Pro-non-car-alternatives	***I like the idea of walking as a means of travel for me.***	***0.691***
	I like the idea of bicycling as a means of travel for me.	0.614
	I like the idea of public transit as a means of travel for me.	0.339
Tech-savvy	***Learning how to use new technologies is often frustrating for me.***	***-0.858***
	I am confident in my ability to use modern technologies.	0.825
Commute benefit	***My commute is a useful transition between home and work (or school).***	***0.740***
	My travel to/from work (or school) is usually pleasant.	0.571
	I wish I could instantly be at work (or school) – the trip itself is a waste of time.	-0.413
Modern urbanite	I would/do enjoy having a lot of luxury things. [d]	0.363
	I like the idea of having stores, restaurants, and offices mixed among the homes in my neighborhood.	***0.362***
	My phone is so important to me, it's almost part of my body.	0.333
Work-oriented	***At this stage of my life, having fun is more important to me than working hard.***	***-0.619***
	I'm too busy to have as much leisure time as I'd like.	0.500
	It's very important to me to achieve success in my work.	0.293
Materialistic [c]	***I usually go for the basic ("no-frills") option rather than paying more money for extras.***	***-0.583***
	I would/do enjoy having a lot of luxury things. [d]	0.460
	The functionality of a car is more important to me than the status of its brand.	-0.454
	I prefer to minimize the amount of things I own.	-0.337
	I like to wait a while rather than being first to buy new products.	-0.336
Polychronic	***I prefer to do one thing at a time.***	***-0.832***
	I like to juggle two or more activities at the same time.	0.694
Pro-environmental [c]	***Cost or convenience takes priority over environmental impacts (e.g. pollution) when I make my daily choices.***	***-0.894***
	I am committed to an environmentally-friendly lifestyle.	0.507
Pro-exercise [c]	***I am committed to exercising regularly.***	***0.736***
	The importance of exercise is overrated.	-0.576
Family/friends-oriented [c]	***Family/friends play a big role in how I schedule my time.***	***0.564***
	It's okay to give up a lot of time with family and friends to achieve other worthy goals.	-0.519
Pro-suburban	***I prefer to live in a spacious home, even if it's farther from public transportation or many places I go to.***	***0.618***
	I see myself living long-term in a suburban or rural setting.	0.420
Waiting-tolerant	***Having to wait is an annoying waste of time.***	***-0.874***
	Having to wait can be a useful pause in a busy day.	0.491
Travel-liking [c]	***I generally enjoy the act of traveling itself.***	***0.714***
	I like exploring new places.	0.543
Sociable	***I'm uncomfortable being around people I don't know.***	***-0.541***
	I consider myself to be a sociable person.	0.491
Pro-car-owning	***I definitely want to own a car.***	***0.750***
	I am fine with not owning a car, as long as I can use/rent one any time I need it.	-0.566
	I like the idea of driving as a means of travel for me.	0.538
	As a general principle, I'd rather own things myself than rent or borrow them from someone else.	0.372

a. Each bolded and italicized item corresponds to the marker variable (MV) of the associated factor.
b. Pattern loadings under 0.300 (in magnitude) are suppressed in this table. One exception is the inclusion of "[i]t's very important to me to achieve success in my work" (0.293) for *Work-oriented*.
c. For easier interpretation, we reversed the direction of the factor by multiplying the original loadings by (−1).
d. This statement has a loading greater than 0.3 (in magnitude) on two constructs (0.363 on *modern urbanite* and 0.460 on *materialistic*); it is the only cross-loading item (using a 0.3 threshold).

Evaluation of the predicted attitudes and MVs

After training ENR models with the donor dataset and imputing attitudinal factor scores into the recipient dataset, we assess the performance of the prediction process by examining the correlations of the observed and predicted attitudes in *both* datasets. Additionally, we investigate the correlations that MVs taken singly have with their associated observed attitudes. We call these assessments the *internal* evaluation. Subsequently, we *externally* evaluate the efficacy of the predicted attitudes and MVs in a few travel behavior models estimated using the recipient dataset. First, after testing various model specifications, for each travel behavior variable the best models respectively without and with predicted attitudes are investigated with respect to the attitudes influencing the travel behavior, improvement in model fit with the inclusion of attitudes, and changes in the statistical significance of other non-attitude variables. Next, we test the inclusion of individual MVs instead of ENR-predicted attitude scores, comparing the best models containing MVs with the other best models. Finally, we appraise model fit improvements (from including predicted attitudes) across different income and education groups.

RESULTS

Internal evaluation

The effectiveness of ENR in predicting attitudes using MVs is confirmed again, with high correlations between observed and predicted attitudes in both donor and recipient datasets[6] (see the top panel of Figure 2). First, although only the non-overlap GDOT ET dataset (i.e., donor sample) is used as the training set, the correlations (in the training set) are close to those of Shaw (2021, Section 4.5), who used the entire GDOT ET dataset as the training set (but with a slightly different factor solution) to train 15 ENR models with 15 MVs. This signifies that the use of the non-overlap GDOT ET dataset (instead of the full sample) as the training set barely harmed the attitude prediction performance of the ENR models. Specifically, six factors have correlations above 0.9 while only two factors (*modern urbanite* and *materialistic*) have correlations lower than 0.8. Second, the ENR models perform as well in the recipient dataset as they do in the donor dataset. The largest gap between the correlations of the two datasets is 0.029 (for *commute benefit*) while the other attitudes have a gap smaller than 0.020. In fact, a few models perform better in the recipient dataset, albeit only barely (*polychronic*, *pro-exercise,* and *sociable*). These observations corroborate the robustness of ENR against overfitting.

One interesting (but also rational) tendency is that smaller-magnitude pattern loadings of MVs are associated with lower prediction performance (of predicted attitudes). For instance, the *modern urbanite* attitude has the smallest correlations (respectively 0.703 and 0.687 in the donor and recipient datasets; the top panel of Figure 2) and also the smallest pattern loading of its MV (0.362; Table 1). Other attitudes whose associated pattern loading magnitudes are smaller than 0.6 also

[6] As mentioned in Section 2.2, in the situation we are *simulating* (i.e., in which the recipient sample consists of NHTS respondents, for which only 15 attitudinal variables are measured instead of the 38 variables used for EFA), the recipient dataset cannot have the *observed* attitudinal factor scores. With the overlap sample of this study, however, the observed attitudes can be computed by applying the Bartlett factor score coefficient matrix obtained from the EFA performed on the donor sample, because we have the full set of attitudinal variables measured for the respondents in the recipient dataset.

present relatively smaller correlations (*materialistic*, *family/friends-oriented*, and *sociable*). In cases where further improvement in prediction of those attitudes is needed, including a few additional MVs can be considered (see Soria (2023) for various approaches to selecting additional MVs).

The bottom panel of Figure 2 presents the correlation between each observed attitude and the corresponding MV (with its sign reversed when the pattern loading on the attitude is negative), which addresses the potential of using MVs instead of predicted attitudes. In both datasets, predicted attitude factor scores have higher correlations with their corresponding observed attitude scores than MVs do (i.e., a given bar in the top panel is higher than the corresponding bar in the bottom panel). This is not surprising, considering that each predicted attitude score is obtained from the linear combination of 15 attitudinal variables including the associated MV. The interesting finding is that the gaps in correlations between the two panels are rather narrow, indicating that an individual MV estimates its respective attitudinal factor score almost as well as the ML-generated score does. For 13 attitudes, the gaps are less than or equal to 0.050 (in both datasets). Two exceptions are *modern urbanite* (donor: 0.200, recipient: 0.175) and *pro-suburban* (0.072 and 0.075).

It is noteworthy that the gaps are the largest for *modern urbanite* and, at the same time, the prediction of its factor scores is worst among all attitudes (correlations: 0.703 and 0.687; see Figure 2). We believe that these observations are closely linked to the pattern loadings of *modern urbanite* (shown in Table 1). First, *modern urbanite* exhibits the smallest pattern loading of its MV (0.362), a value significantly lower than the pattern loadings associated with the MVs of other attitudes (e.g., the next smallest pattern loading of a MV on its associated attitude is 0.541). Second, the three statements with loadings larger than 0.3 (on *modern urbanite*) represent somewhat distinct aspects associated with the latent attitudinal construct we named *modern urbanite*, with pattern loadings similar in magnitude (0.363, 0.362, and 0.333). Therefore, the information from its MV, when excluding the other two variables, is insufficient to produce a high correlation with the observed *modern urbanite* attitude obtained from the full set of attitudinal variables (0.503 and 0.512; see Figure 2). However, the ENR-predicted *modern urbanite* attitude performs substantially better (0.703 and 0.687) with the help of information from the other 14 MVs. In summary, these findings suggest that using all MVs to predict an attitudinal factor score is superior to relying solely on its MV as a surrogate (or partial) measure of the attitude (in terms of internal evaluation), which becomes much more prominent when the attitude is less cohesive and its MV has a relatively weak loading.

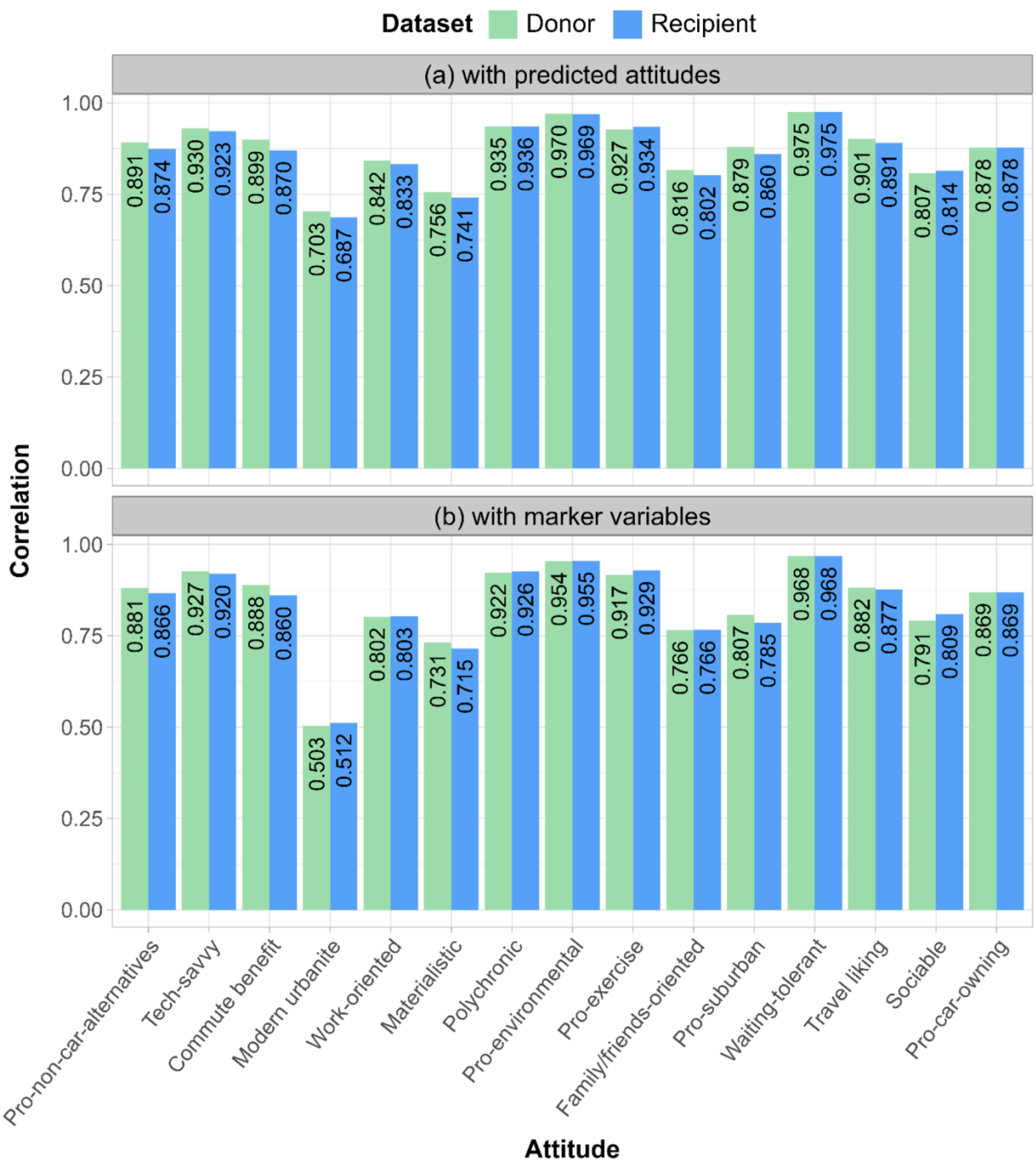


Figure 2 Correlations of observed attitudes (a) with predicted attitudes and (b) with marker variables

External evaluation

Model estimation

Three travel behavior variables are selected from the recipient dataset for external evaluation: household vehicle count, (personal yearly) vehicle miles driven (VMD), and hybrid/electric vehicle adoption. Different subsets of the recipient sample are used to model these variables. The household vehicle count models are estimated with the whole recipient dataset (N = 1,245). On the other hand, the hybrid/electric vehicle adoption model uses survey responses from those who drive and belong to a household with at least one vehicle (N = 1,206), while the VMD model additionally excludes cases with missing or invalid VMD measures (N = 988). A respondent is classified as the adopter of a hybrid/electric vehicle if reporting a non-plug-in hybrid, plug-in hybrid, or electric vehicle in the household (4.6%) (see Appendix A for the distributions of these variables).

Multinomial logit (MNL), log-linear regression, and binary logit (BL) models are employed to model household vehicle count, VMD, and hybrid/electric vehicle adoption, respectively. To specify no-attitude, predicted-attitude, and marker-variable models for each travel behavior variable, we set three basic principles. First, we test including variables with a potential causal influence on the dependent variable. When both directions of causality could conceptually be present, variables are tested only if the direction being modeled is assumed to be the primary relationship. Second, the inclusion of variables is determined based on the interpretability of coefficient estimates (i.e., sign and magnitude) and their statistical significance (with a threshold p-value of 0.05). Third, we start from no-attitude models and explore adding predicted attitudes and MVs, while checking changes in the statistical significance of the impacts of variables tested in the no-attitude models (regardless of whether they are retained in the final no-attitude models). The purpose of the third principle is to mimic the process that transportation demand modelers who decided to add marker statements to a survey for the first time would go through. With these principles, various specifications are tested to obtain final models based on parsimony and interpretability. The summary statistics of variables used in the final models of each travel behavior variable are presented in Appendix A[7].

No-attitude and predicted-attitude models

In this section, we compare no-attitude and predicted-attitude models, starting from the *household vehicle* MNL models (Table 2). According to the no-attitude model with household characteristics, the count of members 16 and over[8], count of members 65 and over, highest education level achieved by members, annual income, and rurality (of residential location) are positively associated with vehicle counts. On the other hand, the number of members with medical conditions (which make it difficult to travel outside) has a negative impact on vehicle counts. Among the five attitudes in the predicted-attitude model, having positive attitudes toward walking, bicycling, and public transit and being tech-savvy negatively influence vehicle counts, whereas being

[7] For each subsample, 1.2–1.3% of cases have missing values for annual household income. These cases are assigned to the reference category when the variable is included in the models.

[8] For this variable, the estimated coefficients for owning one vehicle are negative (see Table 2), which does not conform to general expectations. However, based on the coefficients for the other alternatives, we can say that increasing the count of members 16 and over increases the propensity to own two or more vehicles more than the propensity to own zero or one vehicle.

family/friends-oriented, preferring suburban environments, and favoring owning and driving cars exhibit positive impacts.

With respect to how the impacts of variables in the no-attitude model changed after including the five attitudes, they, in general, remain statistically significant (at the 5% level) and keep the same direction of impact[9] (see Table 2). One interesting observation is that the statistical significance of the urbanicity variable coefficients slightly decreased. The bivariate correlations that the five predicted attitudes have with living in a suburb or small town (from -0.07 to 0.10), or rural area (from -0.08 to 0.17), are small (not shown in Table 2)[10]. However, given that the five attitudes are expected to be linked with residential location choices while explaining household vehicle counts as well, it makes sense that the neighborhood type variables have lower statistical significance in the predicted-attitude model.

The lift in model fit is non-trivial. Even after being penalized by the increased number of estimated coefficients, the market-share (MS)-based McFadden's adjusted ρ^2 increased by 5.2% (i.e., by 0.014 from 0.270 to 0.284) (Table 2)[11]. In this case, non-attitude variables are most useful in explaining vehicle ownership decisions while attitudes provide only moderate additional explanatory power. This aligns with general expectations, because owning a vehicle (and deciding how many to own) is more likely to be by necessity rather than by choice in the context of Georgia (Mokhtarian, 2024).

Table 3 presents the *log-linear VMD* models. In both no-attitude and predicted-attitude models, VMD is lower for people who are female and older, have health issues, and live with household members who drive. Meanwhile, higher VMD is expected from people who are employed, highly educated, and wealthy, have an occupation usually requiring on-site work and visiting multiple locations, live far from the workplace, live in suburban or more rural areas, and belong to a household with sufficient or surplus vehicles. The *materialistic* (related to preferences for luxury things) and *family/friends-oriented* attitudes turn out to positively affect VMD.

One noteworthy point is that the VMD ("YEARMILE") variable in NHTS is likely to involve substantial response errors, as it asks respondents to report VMD over the past 12 months from all

[9] The binary variable for suburban or small town urbanicity is preserved in the final predicted-attitude model, given that its coefficient for the two-vehicle alternative has a p-value very close to 0.05 (0.0502).

[10] The (predicted) *pro-suburban* variable is most strongly associated with the urbanicity variables (with correlations of 0.10 and 0.17).

[11] In view of the fact that the market shares are quite unbalanced (ranging from 2.6% for 0 vehicles to 38.0% for 2 vehicles), making it more difficult to improve upon the MS (alternative-specific-constant (ASC)-only) model, we wanted to ascertain how much of the log-likelihood (LL) improvement could be attributed to the true explanatory variables, as opposed to the ASCs. To do this, we obtained the log-likelihood for no-ASC models (i.e., the final models except for excluding the ASCs) (see Table 2). As an interim calculation, such a model gives explanatory power precedence to the true variables rather than to the ASCs, in contrast to the common practice of starting from the MS (ASC-only) model benchmark (see Mokhtarian, 2016, p. 62). Comparing the LL improvement (relative to the equally-likely (EL) model) of the no-ASC model to that of the final model shows that 74% (= [(–1224.773) – (– 1725.936)] / [(–1045.686) – (–1725.936)], for the no-attitude model) and 78% (= [(–1168.389) – (–1725.936)] / [(– 1010.026) – (–1725.936)], for the predicted-attitude model) of the improvement obtained by the full model can be attributed to the non-ASC variables. The EL-based adjusted McFadden's ρ^2 (of the no-ASC model) increased by 8.7% from 0.275 (= 1 – [(–1224.773 – 27) / (–1725.936)]) for the model without attitudes to 0.299 (= 1 – [(– 1168.389 – 42) / (–1725.936)]) for the model with predicted attitudes (Table 2), indicating a substantial model improvement from including the predicted attitudes.

motorized vehicles (including miles driven in work vehicles, rental cars, and any other vehicles not owned by the household)[12]. Furthermore, VMD "during the past 12 months" is more likely (than "usual" annual VMD) to be subject to irregular factors such as an infrequent long-distance trip (or the idiosyncratic absence of a normally recurring one), which makes it even more difficult to explain the reported VMD using the variables in the dataset. We conjecture that modeling "typical weekly VMD" (not available in NHTS), for instance, might result in more variables (including attitudes) being statistically significant and (potentially) improved model fit. Nevertheless, our adjusted R^2s of 0.23 – 0.24 are not unreasonable (and arguably rather good) for disaggregate models of vehicle-miles driven.

On the other hand, the lift from including attitudes is somewhat limited for the VMD models. The adjusted R^2 measure increased by 3.9% (by 0.009, from 0.228 to 0.237). It is possible that other *observable* variables (even if we did not *observe* them) could expand the amount of variance explained by the model. It is also possible, however, that most of the variance in annual VMD is due to idiosyncratic, literally unobservable, factors – including the reporting error described above.

The *hybrid/electric vehicle adoption* models[13] are distinct from previous models in a few aspects (see Table 4). First, the dependent variable is hard to explain with the individual and household characteristics measured by the NHTS. Only household annual income and highest education level achieved by household members have significant (positive) impacts in the no-attitude model. This is presumably because the share of adopters is small (4.6% = 56 of 1,206) and the adoption of hybrid/electric vehicles is more likely to be associated with attitudes and preferences than with other SED characteristics (compared to our previous two dependent variables, vehicle count and VMD).

Second, two non-attitude variables become statistically significant after including two attitudes (*pro-environmental* with a positive impact and *pro-car-owning* with a negative impact[14]), which did not occur when specifying previous models. Specifically, the p-values associated with "location: Atlanta region" and "ln(annual household VMD + 1)" decreased from 0.06 and 0.13 (when added to the no-attitude model) to 0.04 and 0.03 (in the predicted-attitude model). The positive impact of living in the Atlanta region is conjectured to be related to the lack of electric vehicle charging infrastructure elsewhere, and the relatively greater popularity of trucks (for which there were few electric vehicle options when the 2017 NHTS was administered) in the more rural remainder of the state (Tolbert, 2021). High (annual) household VMD turns out to positively affect the household-level adoption of hybrid/electric vehicles, which is expected considering that a greater demand for driving makes the lower fuel costs of hybrid/electric vehicles more attractive[15].

[12] Please see the question for the "YEARMILE" variable (NHTS_Retrieval_Instrument_20180228.pdf).

[13] The "highest education: bachelor's degree" variable is retained in the predicted-attitude model despite its coefficient having a marginal p-value (0.055), given that not many variables have statistical significance in hybrid/electric vehicle adoption models. For the same reason, this variable is retained in the marker-variable model as well (p-value = 0.051).

[14] It is probable that people with a negative attitude toward driving/owning cars are less attached to conventional vehicles for multiple reasons, including their concerns about emissions and reluctance to spend money on fuel, which may lead to adopting hybrid/electric vehicles. This argument is supported by a negative (but rather modest) correlation between (predicted) *pro-environmental* and *pro-car-owning* attitudes (-0.14).

[15] Using VMD as an explanatory variable for hybrid/electric vehicle adoption can be debated, because the cost of transportation (e.g. gasoline price, fuel efficiency) has been used to model VMD or vehicle miles traveled (VMT) in

The fact that these two variables become statistically significant only after including attitudes in the model suggests that controlling for attitudes is helping to alleviate some omitted variable biases and/or suppression effects occurring in the no-attitude model.[16]

Third, attitudes bring a substantial model lift. The MS-based McFadden's adjusted ρ^2 increased by 64.9% (i.e., by 0.050, from 0.077 to 0.127)[17] (see Table 4). One may say that this large (relative) increase is somewhat exaggerated because (given the highly unbalanced market shares) the remaining information to be explained beyond the MS model is small and the no-attitude model has a low MS-based McFadden's ρ^2. In such cases, however (see footnote 11 and Mokhtarian, 2016), a more appropriate assessment is to compare the share of the total improvement in model log-likelihood (LL) that can be attributed to the true variables, when they (rather than the alternative specific constants (ASCs)) are given explanatory power precedence. Figure 3 illustrates the relative contributions to the final model log-likelihood values of various constrained versions of the two final models. For the no-attitude model, the total improvement in model LL from the equally-likely base is (–206.142 + 835.935) = 629.793 points, while for the predicted-attitude model it is (–190.775 + 835.935) = 645.160. From that perspective alone, the improvement from including attitudes appears quite small. However, the share of that total improvement that is associated with *true* (non-ASC) explanatory variables is 51.7% (= (– 510.148 + 835.935) / 629.793) for the no-attitude model, and a whopping 97.6% (= (– 206.364 + 835.935) / 645.160) for the predicted-attitude model. Thus, the four variables added to the predicted-attitude model are in fact accounting for nearly half ((– 206.364 + 510.148) / (– 206.364 + 835.935) = 48.3%)) of the explanatory power of the non-ASC variables in the model (or, put another way, nearly doubling the explanatory power of the non-ASC variables: (835.935 – 206.364) / (835.935 – 510.148) = 1.93). In conceptual terms, they are helping to explain *why* the market shares are so unbalanced.

---

many previous studies (e.g., Munyon et al., 2018; Wang & Chen, 2014). Similarly, the lower fuel costs of hybrid vehicles have been demonstrated to create a rebound effect, with slightly increased VMT (Sun et al., 2019), i.e. hybrid adoption → VMT rather than the converse relationship of the current model. However, there is also precedent for using VMD to model hybrid/electric vehicle adoption (Diamond, 2009; Wee et al., 2018).

[16] Interestingly, pairwise correlations of the two newly-entering variables with the two (predicted) attitude variables are not high. Log(annual household VMD + 1) has a -0.06 correlation with the pro-environmental attitude and 0.04 with pro-car-owning, while the corresponding correlations for living in the Atlanta region are -0.03 and -0.06. Perhaps a more complex multi-way relationship is accounting for the observed result.

[17] Focusing on the contribution of attitudes, adding *pro-environmental* and *pro-car-owning* (but not the two additional non-attitude variables) to the no-attitude model results in an MS-based McFadden's adjusted ρ^2 of 0.118 (increasing by 53.2% from that of the no-attitude model) (not shown in Table 4). This indicates that the two attitudes offer a much larger contribution than the two additional non-attitude variables, while the contribution of the latter variables is still non-negligible.

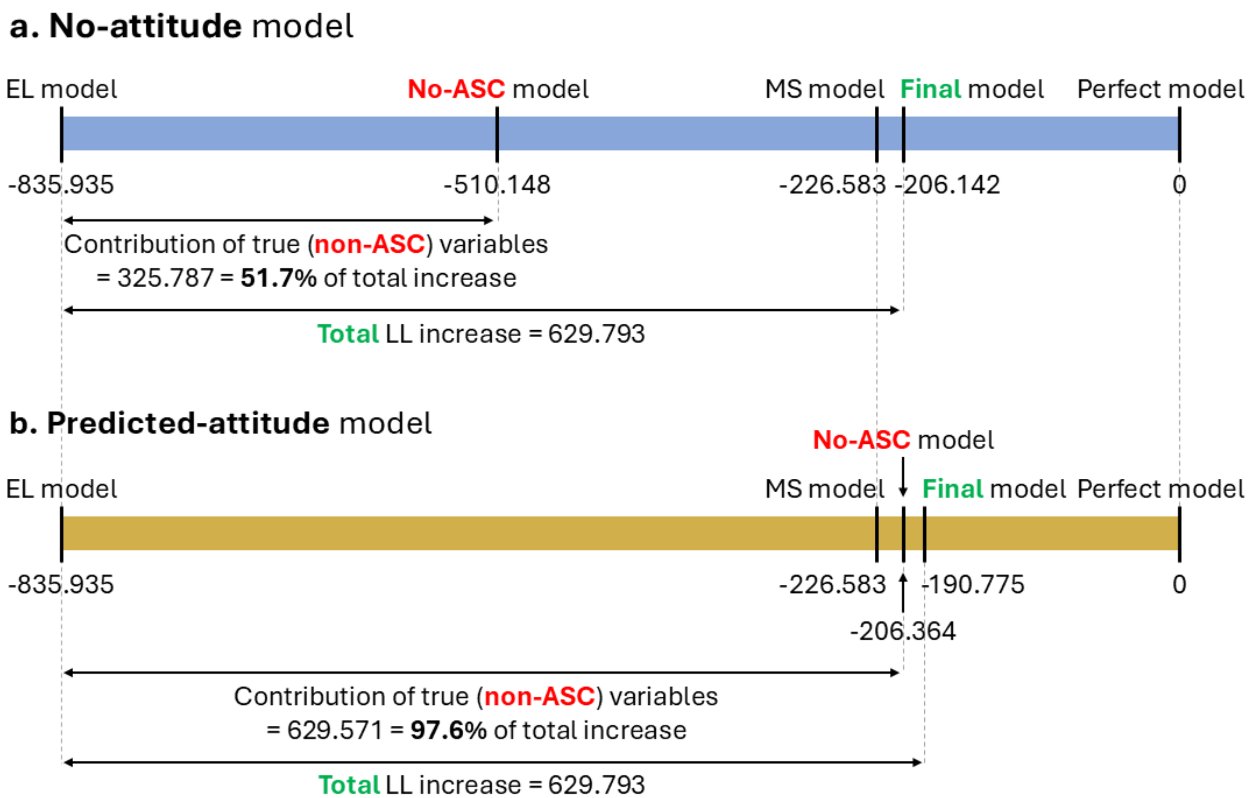


Figure 3 Comparison of log-likelihoods of hybrid/electric vehicle adoption models

Marker-variable models

Although ENR is a rather simple ML model and provides additional depth to the attitude variables because all MVs (instead of just one) are used to impute each attitude, situations may arise where it is necessary to further simplify the process of measuring and using attitudes. To simulate such situations, marker-variable models are estimated by directly incorporating the attitudinal MVs as explanatory variables. These models are specified separately from the predicted-attitude models, following the principles outlined in Section 3.2.1. The main focus is on analyzing how the model specifications and goodness-of-fit measures differ from those of the predicted-attitude models: we might expect MVs alone not to perform as well as the predicted attitude scores which exploit the information in all MVs collectively, and thus the question is, how much worse are they, and what are the consequences for the behavioral interpretation of the models? To ensure that each attitude and its MV point in the same direction, the signs are reversed for MVs having a negative pattern loading with the associated attitudes (i.e., for such variables, -2 and 2 will correspond to "strongly *agree*" and "strongly *disagree*", respectively).

As it happens, the marker-variable and predicted-attitude models end up having identical model specifications for two of the three travel behavior variables: vehicle ownership and hybrid/electric vehicle adoption. The sacrifice in model fit is rather minimal. For vehicle ownership, the marker-variable model has an MS-based McFadden's adjusted ρ^2 of 0.280 (3.7% higher than the 0.270 of the no-attitude model), which is quite close to the 0.284 of the predicted-attitude model (Table 2). On the other hand, for hybrid/electric vehicle adoption, the log-likelihood values are identical to the first decimal place, resulting in the same MS-based McFadden's adjusted ρ^2 for the two models (0.127; Table 4).

These findings point to a potentially excellent performance of singular MVs as proxies for observed attitudes in travel behavior modeling, which can be expected when they are strongly correlated with the observed attitudes they represent. In the present case, those correlations are often comparable to those of the ENR-predicted attitudes with observed attitudes (Section 3.1). For example, for the two attitudes in the hybrid/electric vehicle adoption models (*pro-environmental* and *pro-car-owning*), the *MVs'* correlations with their corresponding observed attitudes (0.955 and 0.869) differ only slightly from the correlations of the *predicted attitudes* with their observed counterparts (0.969 and 0.878) (lower by 0.014 and 0.009 in the recipient dataset; see Figure 2). This leads to (practically) the same model fit statistics for marker-variable and predicted-attitude models. However, the gaps in correlations are relatively wider for two of the five attitudes in the vehicle ownership models (*family/friends-oriented*: 0.036; *pro-suburban*: 0.075 in the recipient dataset; see Figure 2). Consequently, the marker-variable model for vehicle ownership has slightly worse model fit statistics than the predicted-attitude model, despite identical model specifications.

On the other hand, for log-linear VMD, marker-variable and predicted-attitude models differ in their specifications (Table 3)[18]. The marker-variable model does not retain the *family/friends-oriented* attitude but includes *modern urbanite* instead. While specifying the predicted-attitude model, we observed that the predicted *modern urbanite* score exhibited very large p-values,[19] whereas its single MV was statistically significant in many cases (including in the final marker-variable model in Table 3). As noted in Section 3.1, the *modern urbanite* factor is dominated by three somewhat diverse variables with similar loadings, and its MV captures a specific aspect related to land-use preferences. This puts the MV at a disadvantage in terms of internal evaluation, but helps reveal that the preference for mixed and high-density land-use (measured by the MV) is more strongly associated with VMD than the associated factor score is, whose computation dilutes that preference by combining it with other attitudinal items. This observation suggests that using a MV (instead of the predicted factor score) is worth considering when the attitude is less cohesive and the MV is suspected to be more directly associated with the dependent variable than the factor score itself is.

[18] The vehicle sufficiency variable is included in the marker-variable model, despite its coefficient's p-value being slightly above 0.05 (0.059), because the variable is conceptually closely linked to VMD and the p-value only marginally exceeds the threshold.

[19] If predicted *modern urbanite* is added to the current predicted-attitude model in Table 3, its p-value is 0.157 (not shown in Table 3). Even worse, adding predicted *modern urbanite* and removing predicted *family/friends-oriented* results in a p-value of 0.427 (not shown in Table 3).

Table 2 Vehicle ownership MNL models (in R 4.3.3 with *mlogit* package)

Variable		Estimated coefficient								
		No-attitude model			Predicted-attitude model			Marker-variable model		
		Base: 0 vehicles (2.6%)			Base: 0 vehicles (2.6%)			Base: 0 vehicles (2.6%)		
		1 veh (32.4%)	2 vehs (38.0%)	3+ vehs (27.0%)	1 veh (32.4%)	2 vehs (38.0%)	3+ vehs (27.0%)	1 veh (32.4%)	2 vehs (38.0%)	3+ vehs (27.0%)
	Alternative-specific constants (ASCs)	1.138 .	-3.708 ***	-6.981 ***	1.630 *	-3.204 ***	-6.425 ***	0.287	-4.591 ***	-7.804 ***
Attitudes	Pro-non-car-alternatives	-	-	-	-0.623 **	-0.681 **	-0.717 **	-0.503 *	-0.545 *	-0.593 **
	Tech-savvy	-	-	-	-0.344	-0.398	-0.527 *	-0.313	-0.373 .	-0.496 *
	Family/friends-oriented	-	-	-	0.490 *	0.418 *	0.412 *	0.518 *	0.456 *	0.441 .
	Pro-suburban	-	-	-	0.103	0.299	0.615 **	0.055	0.241	0.496 *
	Pro-car-owning	-	-	-	0.562 **	0.628 **	0.620 **	0.605 **	0.683 **	0.640 **
Household traits	Count: members 16 years old and older	-0.934 *	1.564 ***	2.547 ***	-1.031 *	1.496 ***	2.484 ***	-1.110 *	1.404 **	2.381 ***
	Count: members 65 years old and older	1.151 **	1.273 **	1.150 **	1.190 *	1.271 **	1.071 *	1.268 **	1.358 **	1.177 *
	Count: members with medical conditions	-0.610	-1.578 ***	-1.831 ***	-0.805 .	-1.819 ***	-2.082 ***	-0.746 .	-1.744 ***	-1.996 ***
	Highest education: some college	1.183 *	1.209 *	1.732 **	1.108 *	1.097 .	1.667 **	1.213 *	1.191 *	1.762 **
	Highest education: bachelor's degree or higher	1.597 **	1.757 **	1.704 **	1.976 **	2.178 **	2.203 **	2.072 **	2.262 ***	2.288 **
	Annual income: $25-50K	2.523 ***	2.714 ***	3.224 ***	2.712 ***	2.935 ***	3.413 ***	2.681 ***	2.905 ***	3.441 ***
	Annual income: $50K or more	3.013 **	4.172 ***	4.908 ***	3.548 **	4.790 ***	5.536 ***	3.477 **	4.696 ***	5.481 ***
	Urbanicity: suburban or small town	0.427	1.011 *	1.081 *	0.522	1.038 .	0.964 .	0.552	1.095 *	1.070 .
	Urbanicity: rural	1.618 *	2.359 **	3.162 ***	1.189	1.826 *	2.410 **	1.440 *	2.112 **	2.780 ***

		No-attitude model	Predicted-attitude model	Marker-variable model
Model log-likelihood		-1045.686	-1010.026	-1015.783
McFadden's ρ^2	MS base	0.289	0.313	0.309
	EL base	0.394	0.415	0.411
McFadden's adjusted ρ^2	MS base	0.270	0.284	0.280
	EL base	0.377	0.389	0.385
No-ASC model log-likelihood		-1224.773	-1168.389	-1162.072
Initial log-likelihood	Market-share (MS) model		-1469.702	
	Equally-likely (EL) model		-1725.936	

Notes: Significant at the 10% level (.), 5% level (*), 1% level (**), and 0.1% level (***)

Table 3 Log-linear VMD models (estimated in R 4.3.3)

Variable		Estimated coefficient: No-attitude model	Predicted-attitude model	Marker-variable model
Constant		8.561 ***	8.636 ***	8.689 ***
Attitudes	Modern urbanite	-	-	-0.054 *
	Materialistic	-	0.107 **	0.095 **
	Family/friends-oriented	-	0.061 *	-
Individual traits	Female	-0.339 ***	-0.339 ***	-0.320 ***
	Age: 45-64	-0.246 **	-0.234 *	-0.258 **
	Age: 65+	-0.296 **	-0.277 **	-0.300 **
	Worker	0.184 *	0.188 *	0.181 *
	Occupation: manufacturing, construction, maintenance, or farming	0.407 *	0.416 *	0.405 *
	Distance to work: 20 miles or more	0.562 ***	0.553 ***	0.561 ***
	Education: some college or higher	0.253 **	0.234 *	0.235 *
	Have medical conditions	-0.541 ***	-0.566 ***	-0.570 ***
	Health: fair or poor	-0.338 **	-0.351 **	-0.350 **
Household traits	Count: drivers	-0.108 *	-0.116 *	-0.111 *
	Annual income: $25-50K	0.356 ***	0.343 **	0.348 **
	Annual income: $50-100K	0.499 ***	0.450 ***	0.468 ***
	Annual income: $100K or more	0.623 ***	0.556 ***	0.591 ***
	Urbanicity: suburban, small town, or rural	0.206 **	0.198 **	0.180 *
	Vehicle sufficiency: sufficient or surplus	0.221 *	0.217 *	0.205 .
R^2		0.240	0.250	0.251
Adjusted R^2		0.228	0.237	0.238

Notes: Significant at the 10% level (.), 5% level (*), 1% level (**), and 0.1% level (***)

Table 4 Hybrid/electric vehicle adoption BL models (in R 4.3.3 with *mlogit* package)

Variable		Estimated coefficient: No-attitude model	Predicted-attitude model	Marker-variable model
Alternative-specific constants (ASCs)		-4.623 ***	-9.233 ***	-8.487 ***
Attitudes	Pro-environmental	-	0.500 ***	0.530 ***
	Pro-car-owning	-	-0.368 **	-0.430 **
Household traits	Highest education: bachelor's degree	1.252 *	1.038 .	1.053 .
	Highest education: graduate degree	1.867 ***	1.569 **	1.615 **
	Annual income: $150K or more	0.791 **	0.657 *	0.673 *
	Location: Atlanta region	-	0.592 *	0.617 *
	ln(annual household VMD + 1)	-	0.450 *	0.433 *
Model log-likelihood		-206.142	-190.775	-190.796
McFadden's ρ^2	MS base	0.090	0.158	0.158
	EL base	0.753	0.772	0.772
McFadden's adjusted ρ^2	MS base	0.077	0.127	0.127
	EL base	0.749	0.762	0.762
No-ASC model log-likelihood		-510.148	-206.364	-203.582
Initial log-likelihood	Market-share (MS) model		-226.583	
	Equally-likely (EL) model		-835.935	

Notes: Significant at the 10% level (.), 5% level (*), 1% level (**), and 0.1% level (***)

1.1.1 *3.2.4. Segmentation analysis*

To better understand how attitudes may benefit travel behavior models, we analyzed model fit statistics across various (household-level) income and education groups. We focused on the vehicle ownership and hybrid/electric vehicle adoption models, because these models showed relatively greater improvement when attitudes were included. The central idea is to compare (probability-weighted) percent correctly classified values from no-attitude and predicted-attitude models across alternatives and groups. For a given model, the "percent correctly classified" (PCC) is calculated as follows:

$$PCC_i = \frac{\sum_{n=1}^{N} P_n(i) y_{in}}{\sum_{n=1}^{N} y_{in}} \times 100(\%) \qquad \text{(Equation 2)}$$

where PCC_i = percent correctly classified for alternative i, N = sample size, $P_n(i)$ = the predicted probability of individual $\underline{n}$ selecting alternative i, and y_{in} = the binary choice variable of individual n (1 if i is chosen, or 0 otherwise). The denominator of the fraction is the number of people observed to choose alternative i, and the numerator is the expected number of people predicted by the model to choose i, among those observed to choose i. Therefore, PCC_i indicates the average probability of choosing alternative i (based on the model predictions) among those who are observed to choose alternative i (in the sample).

Focusing first on the unsegmented samples (portrayed in the rightmost panels of Figures 4 and 5), the PCC values indicate that attitudes help better predict alternatives (owning zero vehicles and adopting hybrid/electric vehicles) that are less-often chosen and, at the same time, hard to predict using household and individual traits other than attitudes. In the no-attitude model of vehicle ownership (represented by the orange bars in the identical rightmost panels of Figure 4), PCC is much lower for the zero-vehicle alternative (23.2%, compared to 45.0 – 61.2% for the other alternatives). This is closely linked to its low share in the sample (2.6%; see the sample shares in Appendix B), and potentially to the limited explanatory power that the non-attitude variables in the model can offer to the prediction of having zero vehicles. However, the attitudinal information added to the model brings a substantial growth in the PCC for zero vehicles (35.0%, an 11.8 percentage-point increase) while offering much smaller increases for other alternatives (0.6 – 1.8 percentage points).

Similarly, in the no-attitude model of hybrid/electric vehicle adoption (represented by the orange bars in the identical rightmost panels of Figure 5), PCC is much lower for the adoption alternative (8.0%, compared to 95.5% for non-adoption), probably for the same reasons why PCC is the lowest for zero vehicles in the vehicle ownership model. Although the increase in PCC is much larger for adoption (4.6 percentage points, a 58% improvement over the PCC of the no-attitude model) than for non-adoption (0.2 percentage points), it would be proper to say that the inclusion of attitudes helps the model better distinguish between adoption and non-adoption (rather than improving the prediction performance of adoption only), given that those are the only two alternatives. Overall, these observations suggest that a key benefit of including attitudes in travel behavior models is that such models can better explain infrequent choices, for which attitudes play a more prominent role.

Turning now to the comparison across income and education segments, we see that groups that are *more* likely (than others) to select the less-often chosen and hard-to-predict alternatives gain greater benefits from the inclusion of attitudes. For example, the low-income group contains almost all of the respondents with zero household vehicles (30 out of 32) while the high-income group contains none of them (see Appendix B), which aligns with the positive coefficients associated with household income in Table 2. As shown in the top row of Figure 4, the improvement in PCC for zero vehicles is most prominent in the low-income group. This is logical given that people in the low-income group are likely to have limited purchasing power, which potentially gives attitudes a large role in vehicle ownership decisions after controlling for financial status. The results for the hybrid/electric vehicle adoption model spotlight the other end of the income spectrum, but illustrate the same point (that including attitudes most benefits predictions for the group that is more likely than others to select infrequently-chosen, harder-to-predict alternatives). In this case, hybrid/electric vehicle adoption is mostly concentrated in the high-income group (51 of 56; Appendix C), which is the group that also benefits the most from the inclusion of attitudes (see the top of Figure 5). The logic is similar: once controlling for financial ability, attitudes are likely to be particularly important to explaining why some people adopt hybrid/electric vehicles and others do not.

On the other hand, we also observe that attitudes help explain why some people make choices that seem unlikely considering their SED characteristics. Household income and education are strongly positively associated (see Appendix D) and also positively influence vehicle count and hybrid/electric vehicle adoption (see Tables 2 and 4). This makes the bottom panels of Figures 4 and 5 look quite similar to the top panels in terms of the changes with the inclusion of predicted attitudes (i.e., the blue bars). However, deviating from the pattern, the bottom row of Figure 4 shows that the increase in PCC for zero vehicles is even larger in the high-education group than in the low-education group.

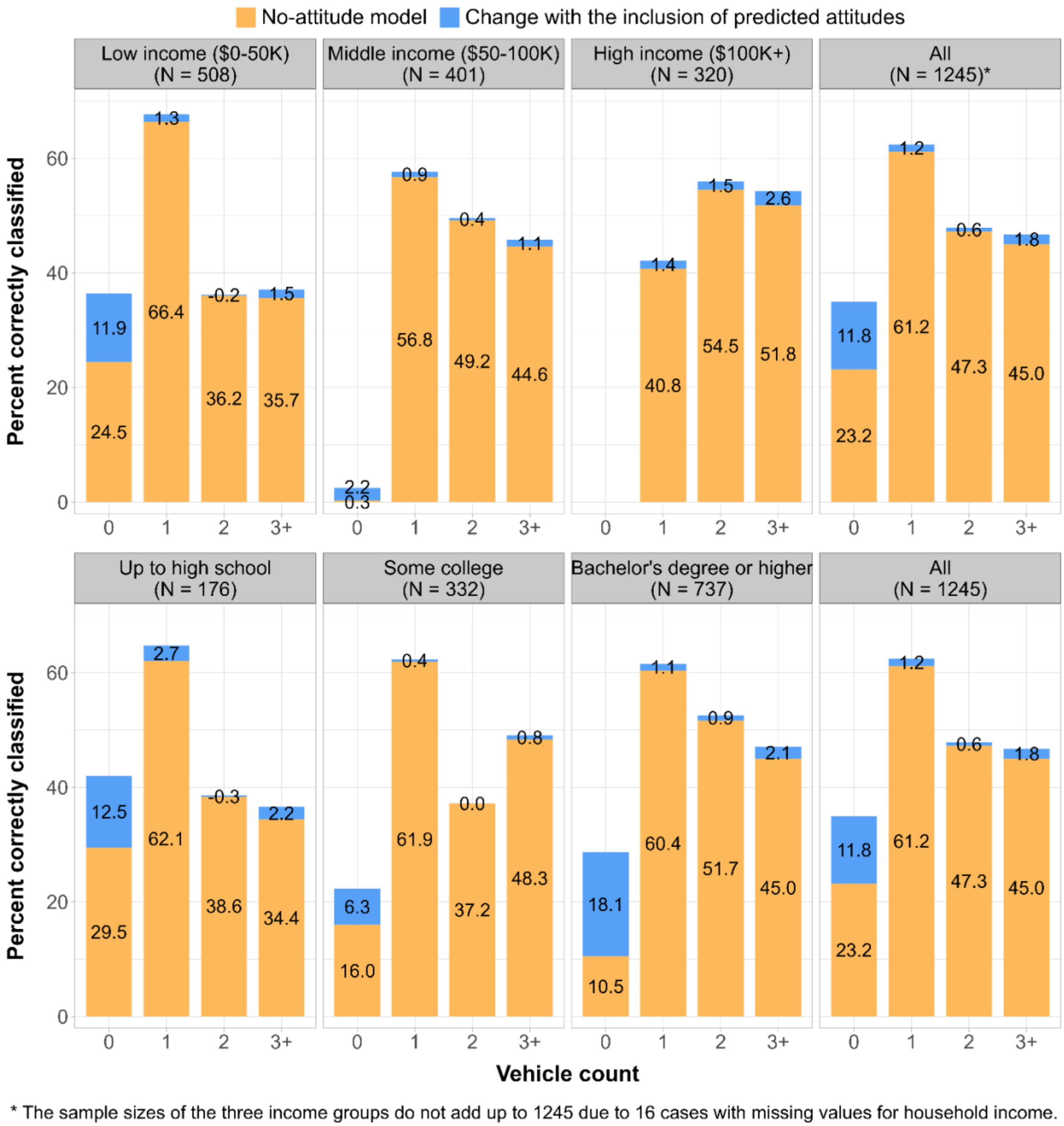
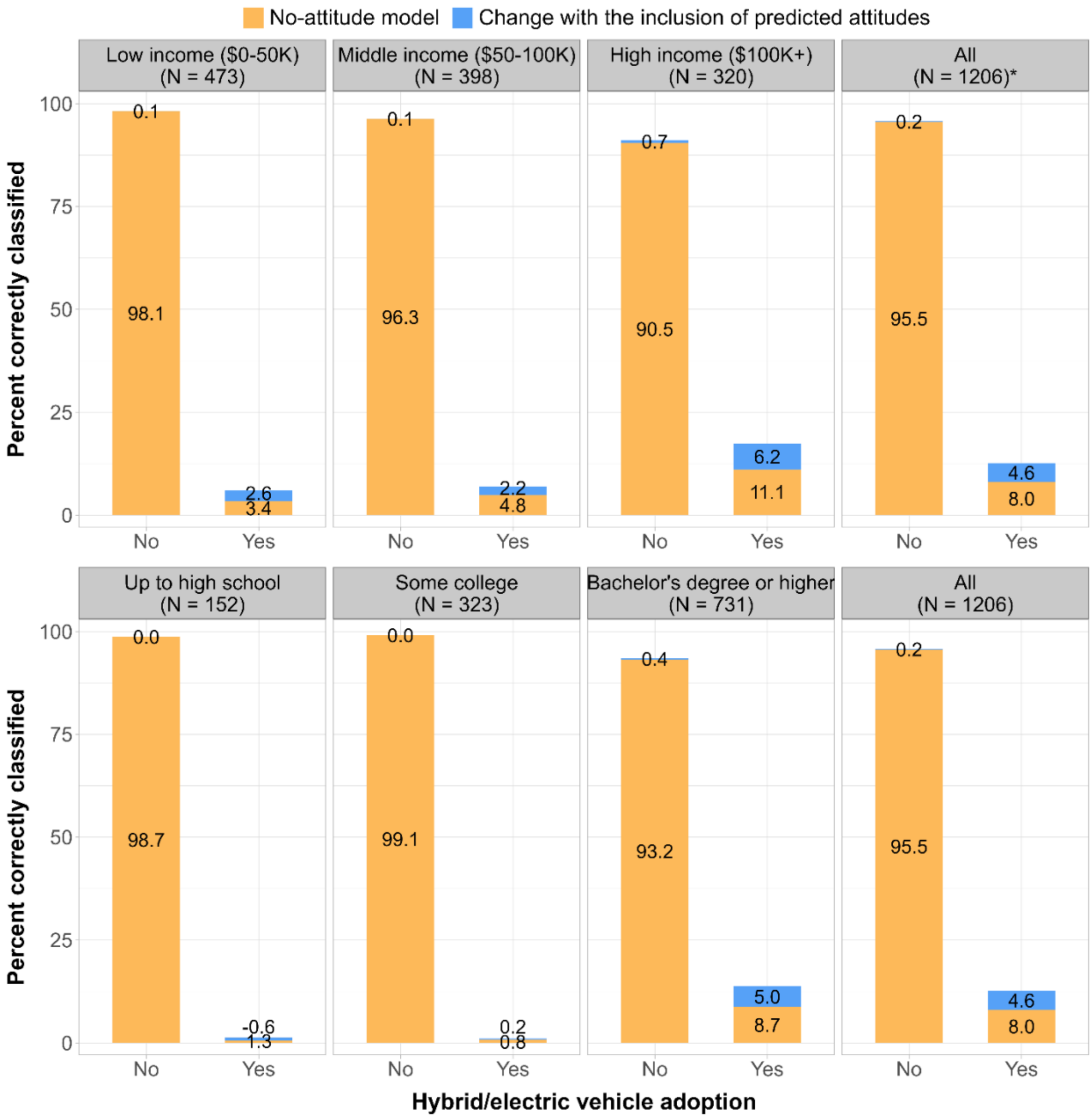


Figure 4 Vehicle ownership models: percent correctly classified values by household-level income and education



* The sample sizes of the three income groups do not add up to 1206 due to 15 cases with missing values for household income.

Figure 5 Hybrid/electric vehicle adoption models: percent correctly classified values by household-level income and education

DISCUSSION AND CONCLUSIONS

This study contributes to the growing body of evidence demonstrating the potential of using a small number of attitudinal marker variables (MVs) to improve practice-oriented travel demand models. This approach means that the burden *on the analyst* of designing, deploying, and factor-analyzing attitudes, as well as the burden *on the respondent* of reacting to a typically large number of attitudinal statements in a survey, can be substantially reduced without compromising many of the gains from having a much larger set of attitudinal variables.

Specifically, we exploited the rare opportunity of possessing an overlapping sample from two spatially congruent and almost contemporaneous surveys: the Georgia DOT Emerging Technologies (GDOT ET) survey (N = 3,288), containing 38 attitudinal statements (shown in Table 1); and the 2017 Georgia add-on to the National Household Travel Survey (N = 8,611), containing a variety of travel behavior variables. The responses of the 1,245 individuals who completed both surveys served as the recipient dataset for the analysis (see Figure 1), simulating a future situation in which strategically-chosen attitudinal variables are measured in a large-scale household travel survey. Using only the *non-overlap* sample of the GDOT ET survey as the donor dataset, we extracted 15 factors from the 38 attitudinal variables, created scores on each of those factors, and identified a single marker statement for each factor (generally the highest-loading item on that factor). Since those MVs were also available in the recipient sample, they constituted the common variables used to train elastic net regression (ENR) models to predict the factor scores from the 15 MVs alone. Those models were then applied to the recipient sample to produce predicted scores for that dataset.

t is worth pointing out that the overlap sample played two important roles in this study. First, as just mentioned, it furnished the common variables needed for the attitude prediction approach to work. In the future, if those variables are *separately* included in a household travel survey, an overlap sample will not be needed for that purpose. Second, however, it allowed us to check how well the ENR models predicted attitude scores for the recipient sample. That was the case because, since all 38 of the attitudinal items were also available in that dataset, we could apply the factor score coefficients from the factor analysis on the donor sample to the corresponding variables in the recipient sample, to compute "observed" scores for that sample. In cases where a household travel survey does not also include the larger set of attitudinal items, this check could not be performed. Accordingly, it would be desirable to conduct further tests with other overlapping survey samples as the occasion arises (or, with split-sample approaches such as that used by Soria and Mokhtarian, 2024), but the results for the present study are extremely encouraging.

To elaborate, the internal evaluation confirmed that it is possible to predict factor scores obtained from 38 attitudinal variables with relatively high fidelity (correlations in the donor dataset: 0.703 – 0.975; correlations in the recipient dataset: 0.687 – 0.975) using only 15 MVs as inputs for ENR models. In addition, the minimal performance differences of ENR models between the two datasets validate the robustness of ENR against overfitting. As expected, predicted attitudes have higher correlations with their corresponding observed attitudes than the MVs alone do. However, in general, the gaps are rather narrow, indicating excellent performance of the marker variables alone as surrogate measures of their respective associated observed attitudes. One clear exception is *modern urbanite*, for which the correlation differences are the largest (0.200 for the donor sample;

0.175 for the recipient sample) and, at the same time, the prediction of its factor scores is worst among all attitudes (correlations for donor sample: 0.703; recipient sample: 0.687). These observations are interconnected to (a) the small pattern loading of the associated marker variable and (b) the three statements most associated with the factor representing somewhat diverse aspects with pattern loadings similar in magnitude. In such instances, using all 15 MVs to predict factor scores can offer a substantial improvement over relying on one MV.

The external evaluation, performed on the recipient sample, assessed the performance of (1) the predicted attitudinal factor scores and (2) the MVs themselves in models of three different travel behavior variables (obtained from the NHTS): household vehicle count (multinomial logit), annual vehicle-miles driven (log-linear regression), and hybrid/electric vehicle adoption (binary logit). The results showed that various attitudes were significant in each of the three models, in conceptually expected ways. While the number of included attitudes is largest for vehicle ownership, improvements in model fit were most prominent for hybrid/ electric vehicle adoption. Adding attitudes did not materially harm the statistical significance of the non-attitude variables that had been initially included. To the contrary, in the model of hybrid/electric vehicle adoption, incorporating attitudes identified two additional non-attitude variables that would have been overlooked as insignificant otherwise.

Marker-variable models, specified separately from the predicted-attitude models, exhibited goodness-of-fit statistics very close to those of the predicted-attitude models, demonstrating the remarkable effectiveness of using MVs themselves in travel behavior models. Interestingly, the vehicle ownership and hybrid/electric vehicle adoption models ended up having the same attitudes for the marker-variable models and predicted-attitude models, but the (log) VMD model did not. For VMD, *modern urbanite* is included only in the marker-variable model, suggesting that a MV may perform *better* than the associated (predicted) attitudinal factor score, when the MV is more strongly related to the dependent variable of interest than is the attitude score (which is a composite of all MVs in the case of the predicted score).

Lastly, for the two discrete choice models (household vehicle count and hybrid/electric vehicle adoption), an analysis of the (probability-weighted) alternative-specific percent correctly classified identified that including (the ENR-predicted) attitudes *especially* promotes the improved prediction of alternatives that are less commonly chosen and/or relatively hard to predict with *non-attitude* variables (namely, having zero household vehicles and adoption of hybrid/electric vehicles). A segmentation analysis on income and education revealed that this improvement mainly stems from better predicting the behavior of groups that are more likely (than others) to select such alternatives. At the same time, we found that attitudes can contribute to describing behavior that contradicts what would be expected based on other variables. To our knowledge, these results are distinctive to this study, and worthy of further research to ascertain whether this pattern is typical. If so, it would certainly seem to further support the value – to transportation planning practice – of incorporating the contribution of attitudes to otherwise conventional travel demand models.

Although, as mentioned, the two surveys analyzed in this study occurred very close together in time (as well as space), MVs were measured later than other variables in the recipient dataset. Specifically, the GDOT ET survey data were collected from October 2017 to April 2018 (Kim et al., 2019), while the NHTS Georgia add-on data were collected from April 2016 to April 2017. Therefore, the attitudes used as explanatory variables in the travel behavior models were actually measured at least a few months later (by the GDOT ET survey). Although attitudes do not usually change drastically within a short period (unless life-cycle changes or other big events occur), using attitudes measured *after* the other variables in the models were (including after the behaviors that the attitudes are presumed to *cause*) may have introduced a temporal mismatch that degraded the performance of predicted-attitude and marker-variable models. A better alignment of the temporal sequence could improve the model lift somewhat.

In addition to providing empirical findings and analytical frameworks that help evaluate the benefits of measuring attitudinal MVs and incorporating predicted attitudes (or MVs themselves) into transportation demand models, this study suggests multiple routes for future research as well:

First, several other travel behavior variables (e.g., public transit usage, ridehailing usage, work-from-home adoption) can be modeled incorporating attitudes, to ascertain which attitudes matter, how much model lift we can achieve, and how model interpretations change.

Second, it would be interesting to explore how the results from models with and without attitudes differ in various scenario tests with hypothetical (but plausible) changes in *non-attitudinal* variables. Also, although the *prediction* of future attitudes is a challenge that remains to be tackled (see, e.g., Mokhtarian, 2024; Choi and Mokhtarian, 2024), it is certainly possible, and would be enlightening, to perform *scenario* tests involving hypothetical changes in attitudes. Such changes are likely to bring meaningful behavioral and policy implications, considering that in some cases, built environment (BE) characteristics and various policies may be more likely to influence attitudes than to change certain SED characteristics.

Third, whereas the present study included attitudes only directly as ordinary explanatory variables, it would be valuable to explore their role in the class membership component of latent class (regression or discrete choice) models, signifying the likely influence of attitudes on the importance (i.e. the coefficient magnitudes) of other explanatory variables (e.g., Section 5.2.2 of Shaw, 2021; Kim and Mokhtarian, 2023; Swait, 1994).

Fourth, measuring attitudes (at least with MVs) in household surveys may ultimately support exploration of the multi-way relationships among attitudes, SED, BE, and travel behavior variables using approaches such as structural equation modeling (e.g., De Vos et al., 2021; Kroesen et al., 2017; Wang & Lin, 2019), in a practice-oriented travel demand forecasting context. Such efforts can contribute to disentangling the interactions among of attitudes, SED, BE, and travel behavior, and, thus, to making better policy decisions.

Lastly, it would be beneficial to study the stability of attitudinal constructs and the transferability of attitude imputation functions, because in general, donor and recipient samples may have larger temporal and/or spatial gaps than those in the present study.

REFERENCES

Becker, H., Loder, A., Schmid, B., & Axhausen, K. W. (2017). Modeling car-sharing membership as a mobility tool: A multivariate Probit approach with latent variables. *Travel Behaviour and Society*, *8*, 26–36. https://doi.org/10.1016/j.tbs.2017.04.006

Burisch, M. (1997). Test length and validity revisited. *European Journal of Personality*, *11*(4), 303–315.

Cain, K. L., Gavand, K. A., Conway, T. L., Geremia, C. M., Millstein, R. A., Frank, L. D., Saelens, B. E., Adams, M. A., Glanz, K., King, A. C., & Sallis, J. F. (2017). Developing and validating an abbreviated version of the Microscale Audit for Pedestrian Streetscapes (MAPS-Abbreviated). *Journal of Transport & Health*, *5*, 84–96. https://doi.org/10.1016/j.jth.2017.05.004

Cao, X., Mokhtarian, P. L., & Handy, S. L. (2007). Cross-Sectional and Quasi-Panel Explorations of the Connection between the Built Environment and Auto Ownership. *Environment and Planning A: Economy and Space*, *39*(4), 830–847. https://doi.org/10.1068/a37437

Cerin, E., Conway, T. L., Saelens, B. E., Frank, L. D., & Sallis, J. F. (2009). Cross-validation of the factorial structure of the Neighborhood Environment Walkability Scale (NEWS) and its abbreviated form (NEWS-A). *International Journal of Behavioral Nutrition and Physical Activity*, *6*(1), 32. https://doi.org/10.1186/1479-5868-6-32

Cerin, E., Saelens, B. E., Sallis, J. F., & Frank, L. D. (2006). Neighborhood Environment Walkability Scale: Validity and development of a short form. *Medicine and Science in Sports and Exercise*, *38*(9), 1682–1691. https://doi.org/10.1249/01.mss.0000227639.83607.4d

Choi, S.-E.(K.), & Mokhtarian, P. L. (2024). How temporally stable are attitudes? It depends. Presentation to the ABM3 Symposium on Representation of Evolutionary Travel Behavior, Raitenhaslach, Germany, December 11–13, 2024. Available at https://www.mos.ed.tum.de/fileadmin/w00ccp/tb/abm2024_slides/232_How_temporally_stable_are_attitudes_PatriciaMokhtarian.pdf, or from the authors.

De Vos, J., Cheng, L., & Witlox, F. (2021). Do changes in the residential location lead to changes in travel attitudes? A structural equation modeling approach. *Transportation*, *48*(4), 2011–2034. https://doi.org/10.1007/s11116-020-10119-7

Diamond, D. (2009). The impact of government incentives for hybrid-electric vehicles: Evidence from US states. *Energy Policy*, *37*(3), 972–983. https://doi.org/10.1016/j.enpol.2008.09.094

Domarchi, C., Tudela, A., & González, A. (2008). Effect of attitudes, habit and affective appraisal on mode choice: An application to university workers. *Transportation*, *35*(5), 585–599. https://doi.org/10.1007/s11116-008-9168-6

Goldberg, D. P., & Hillier, V. F. (1979). A scaled version of the General Health Questionnaire. *Psychological Medicine*, *9*(1), 139–145. https://doi.org/10.1017/S0033291700021644

Gosling, S. D., Rentfrow, P. J., & Swann, W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality*, *37*(6), 504–528. https://doi.org/10.1016/S0092-6566(03)00046-1

Haddad, A. J., Mondal, A., & Bhat, C. R. (2023). Eat-in or eat-out? A joint model to analyze the new landscape of dinner meal preferences. *Transportation Research Part C: Emerging Technologies*, *147*, 104016. https://doi.org/10.1016/j.trc.2023.104016

Hatcher, L. (1994). *A Step-by-Step Approach to Using the SAS System for Factor Analysis and Structural Equation Modeling* (1st ed.). SAS Publishing.
Hendrick, C., Hendrick, S. S., & Dicke, A. (1998). The Love Attitudes Scale: Short Form. *Journal of Social and Personal Relationships*, *15*(2), 147–159. https://doi.org/10.1177/0265407598152001
Hoerl, A. E., & Kennard, R. W. (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*. https://www.tandfonline.com/doi/abs/10.1080/00401706.1970.10488634
Kash, G., Mokhtarian, P. L., & Circella, G. (2021). *Analysis of the Georgia Add-On to the 2016–2017 National Household Travel Survey* (FHWA-GA-21-1824). https://rosap.ntl.bts.gov/view/dot/57499
Kaufman, E. A., Xia, M., Fosco, G., Yaptangco, M., Skidmore, C. R., & Crowell, S. E. (2016). The Difficulties in Emotion Regulation Scale Short Form (DERS-SF): Validation and Replication in Adolescent and Adult Samples. *Journal of Psychopathology and Behavioral Assessment*, *38*(3), 443–455. https://doi.org/10.1007/s10862-015-9529-3
Kim, S. H., & Mokhtarian, P. L. (2023). Finite mixture (or latent class) modeling in transportation: Trends, usage, potential, and future directions. *Transportation Research Part B*, *172*, 134-173. https://doi.org/10.1016/j.trb.2023.03.001
Kim, S. H., Mokhtarian, P. L., & Circella, G. (2019). *The Impact of Emerging Technologies and Trends on Travel Demand in Georgia* (FHWA-GA-19-1631). https://rosap.ntl.bts.gov/view/dot/56095
Kroesen, M., Handy, S., & Chorus, C. (2017). Do attitudes cause behavior or vice versa? An alternative conceptualization of the attitude-behavior relationship in travel behavior modeling. *Transportation Research Part A: Policy and Practice*, *101*, 190–202. https://doi.org/10.1016/j.tra.2017.05.013
Malokin, A., Mokhtarian, P. L., & Circella, G. (2019). *A Transfer Learning-Based Framework for Enriching National Household Travel Survey Data with Attitudinal Variables*. http://hdl.handle.net/1853/66813
Mokhtarian, P. L. (2016). Discrete choice models' ρ^2: A reintroduction to an old friend. *Journal of Choice Modelling*, *21*, 60–65. https://doi.org/10.1016/j.jocm.2016.02.001
Mokhtarian, P. L. (2024). Pursuing the impossible (?) dream: Incorporating attitudes into practice-ready travel demand forecasting models. *Transportation Research Part A: Policy and Practice*, *190*, 104254. https://doi.org/10.1016/j.tra.2024.104254
Morrow, G. D., Clark, E. M., & Brock, K. F. (1995). Individual and Partner Love Styles: Implications for the Quality of Romantic Involvements. *Journal of Social and Personal Relationships*, *12*(3), 363–387. https://doi.org/10.1177/0265407595123003
Munyon, V. V., Bowen, W. M., & Holcombe, J. (2018). Vehicle fuel economy and vehicle miles traveled: An empirical investigation of Jevon's Paradox. *Energy Research & Social Science*, *38*, 19–27. https://doi.org/10.1016/j.erss.2018.01.007
Popuri, Y., Proussaloglou, K., Ayvalik, C., Koppelman, F., & Lee, A. (2011). Importance of traveler attitudes in the choice of public transportation to work: Findings from the Regional Transportation Authority Attitudinal Survey. *Transportation*, *38*(4), 643–661. https://doi.org/10.1007/s11116-011-9336-y
Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality*, *41*(1), 203–212. https://doi.org/10.1016/j.jrp.2006.02.001

Richins, M. L. (2004). The Material Values Scale: Measurement Properties and Development of a Short Form. *Journal of Consumer Research*, *31*(1), 209–219. https://doi.org/10.1086/383436
Robins, R. W., Hendin, H. M., & Trzesniewski, K. H. (2001). Measuring Global Self-Esteem: Construct Validation of a Single-Item Measure and the Rosenberg Self-Esteem Scale. *Personality and Social Psychology Bulletin*, *27*(2), 151–161. https://doi.org/10.1177/0146167201272002
Rummel, R. (1970). *Applied Factor Analysis*. Evanston, IL: Northwestern University.
Sallis, J. F., Cain, K. L., Conway, T. L., Gavand, K. A., Millstein, R. A., Geremia, C. M., Frank, L. D., Saelens, B. E., Glanz, K., & King, A. C. (2015). Is Your Neighborhood Designed to Support Physical Activity? A Brief Streetscape Audit Tool. *Preventing Chronic Disease*, *12*, E141. https://doi.org/10.5888/pcd12.150098
Sandvik, E., Diener, E., & Seidlitz, L. (1993). Subjective Well-Being: The Convergence and Stability of Self-Report and Non-Self-Report Measures. *Journal of Personality*, *61*(3), 317–342. https://doi.org/10.1111/j.1467-6494.1993.tb00283.x
Shaw, F. A. (2021). *Methods for enriching transportation survey datasets: With sample applications using psychometric variables*. http://hdl.handle.net/1853/64640
Silveira, S. L., & Motl, R. W. (2020). Abbreviated Neighborhood Environment Walkability scale in persons with multiple sclerosis: Initial validation of score inferences. *Journal of Transport & Health*, *19*, 100952. https://doi.org/10.1016/j.jth.2020.100952
Soria, J. (2023). Methods for marker variable selection and internal validation: Abbreviated attitudes for travel demand modeling. *Paper Presented at the Bridging Transportation Researchers Conference (BTR 5), Western Sessions, Day 2 - Track 2. August*. Available from the author. https://www.youtube.com/watch?v=lQTDRTgwlHA&list=PL9Qv7hTVQuAQoBBCCAFTBm1E2qmjGF253&index=20
Soria, J., & Mokhtarian, P. L. (2024). Using marker statements to impute attitudes: Evaluating their efficacy in vehicle ownership models. Under review; available from the authors.
Sun, S., Delgado, M. S., & Khanna, N. (2019). Hybrid vehicles, social signals and household driving: Implications for miles traveled and gasoline consumption. *Energy Economics*, *84*, 104519. https://doi.org/10.1016/j.eneco.2019.104519
Swait, J. (1994). A structural equation model of latent segmentation and product choice for cross-sectional revealed preference choice data. *Journal of Retailing and Consumer Services, 1*(2), 77–89.
Tambs, K., & Moum, T. (1993). How well can a few questionnaire items indicate anxiety and depression? *Acta Psychiatrica Scandinavica*, *87*(5), 364–367. https://doi.org/10.1111/j.1600-0447.1993.tb03388.x
Thompson, B., & Borrello, G. M. (1987). Concurrent Validity of a Love Relationships Scale. *Educational and Psychological Measurement*, *47*(4), 985–995. https://doi.org/10.1177/0013164487474014
Tibshirani, R. (1996). Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x
Tolbert, J. (2021, October 22). *Beyond Cities: Breaking Through Barriers to Rural Electric Vehicle Adoption | Article | EESI*. https://www.eesi.org/articles/view/beyond-cities-breaking-through-barriers-to-rural-electric-vehicle-adoption

Wang, D., & Lin, T. (2019). Built environment, travel behavior, and residential self-selection: A study based on panel data from Beijing, China. *Transportation*, *46*(1), 51–74. https://doi.org/10.1007/s11116-017-9783-1

Wang, T., & Chen, C. (2014). Impact of fuel price on vehicle miles traveled (VMT): Do the poor respond in the same way as the rich? *Transportation*, *41*(1), 91–105. https://doi.org/10.1007/s11116-013-9478-1

Ware, J. E., Kosinski, M., & Keller, S. D. (1996). A 12-Item Short-Form Health Survey: Construction of Scales and Preliminary Tests of Reliability and Validity. *Medical Care*, *34*(3), 220.

Wee, S., Coffman, M., & La Croix, S. (2018). Do electric vehicle incentives matter? Evidence from the 50 U.S. states. *Research Policy*, *47*(9), 1601–1610. https://doi.org/10.1016/j.respol.2018.05.003

Wu, G., Yamamoto, T., & Kitamura, R. (1999). Vehicle Ownership Model That Incorporates the Causal Structure Underlying Attitudes Toward Vehicle Ownership. *Transportation Research Record*, *1676*(1), 61–67. https://doi.org/10.3141/1676-08

Zou, H., & Hastie, T. (2005). Regularization and Variable Selection Via the Elastic Net. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *67*(2), 301–320. https://doi.org/10.1111/j.1467-9868.2005.00503.x

Appendix A Descriptive statistics of dependent and explanatory variables

Variable			Summary statistics [a]		
			Vehicle ownership (N = 1,245)	ln(yearly VMD) (N = 988)	Hybrid/electric vehicle adoption (N = 1,206)
Travel behavior variables			0 : 2.6% 1 : 32.4% 2 : 38.0% 3+: 27.0%	9.08 (1.09) [b]	No: 95.4% Yes: 4.6%
Predicted attitudes	Pro-non-car-alternatives		-0.05 (1.06)	-	-
	Tech-savvy		-0.14 (1.01)	-	-
	Materialistic		-	-0.07 (0.91)	-
	Pro-environmental		-	-	0.03 (1.06)
	Family/friends-oriented		0.00 (1.10)	0.02 (1.08)	-
	Pro-suburban		0.00 (1.12)	-	-
	Pro-car-owning		0.01 (1.01)	-	0.04 (0.97)
Marker variables (MVs)	Pro-non-car-alternatives		-0.10 (1.13)	-	-
	Tech-savvy [c]		0.01 (1.25)	-	-
	Modern urbanite		-	0.21 (1.24)	-
	Materialistic [c]		-	-0.37 (1.00)	-
	Pro-environmental [c]		-	-	-0.07 (1.03)
	Family/friends-oriented		0.82 (0.97)	-	-
	Pro-suburban		0.03 (1.12)	-	-
	Pro-car-owning		1.34 (0.87)	-	1.36 (0.84)
Individual traits	Female		-	49.3%	-
	Age (years)	18-44	-	15.3%	-
		45-64	-	40.3%	-
		65+	-	44.4%	-
	Worker		-	48.8%	-
	Occupation: manufacturing, construction, maintenance, or farming [d]		-	3.5%	-
	Distance to work: 20 miles or more [d]		-	9.3%	-
	Education	Up to high school	-	13.8%	-
		Some college	-	28.5%	-
		Bachelor's degree	-	29.3%	-
		Graduate degree	-	28.4%	-
	Have medical conditions [e]		-	7.3%	-
	Health: fair or poor [f]		-	8.2%	-
Household traits	Count: members 16 years old and older [g]	1	36.5%	-	-
		2	51.7%	-	-
		3+	11.7%	-	-
	Count: members 65 years old and older [g]	0	51.9%	-	-
		1	28.0%	-	-
		2+	20.1%	-	-
	Count: members with medical conditions [e, g]	0	83.8%	-	-
		1	14.2%	-	-
		2+	2.0%	-	-
	Count: drivers [g]	1	-	38.4%	-
		2	-	53.9%	-
		3+	-	7.7%	-
	Highest education	Up to high school	14.1%	-	12.6%

Variable		Summary statistics [a]		
		Vehicle ownership (N = 1,245)	ln(yearly VMD) (N = 988)	Hybrid/electric vehicle adoption (N = 1,206)
(of any member)	Some college	26.7%	-	26.8%
	Bachelor's degree	25.5%	-	26.1%
	Graduate degree	33.7%	-	34.5%
Annual income	Less than $25K	17.0%	12.8%	15.0%
	$25-50K	23.8%	23.1%	24.2%
	$50-100K	32.2%	34.2%	33.0%
	$100-150K	16.5%	18.2%	17.0%
	$150K or more	9.2%	10.4%	9.5%
	Missing	1.3%	1.3%	1.2%
Urbanicity of residential location	Urban / second city	20.0%	20.1%	-
	Suburban	20.7%	20.9%	-
	Small town	33.3%	32.4%	-
	Rural	25.9%	26.6%	-
Location: Atlanta region [h]		-	-	31.0%
Vehicle sufficiency: sufficient or surplus [i]		-	91.2%	-
ln(annual household VMD + 1) [j]		-	-	9.62 (1.03)

Notes:

a. Numbers shown in the table are shares, means, or (in parentheses) standard deviations as appropriate.
b. The VMD variable itself ranges from 2 to 120,000 (with a mean and standard deviation of 12,680 and 9,976).
c. For convenience, the sign is reversed given its negative pattern loading with the associated attitude. Therefore, -2 and 2 correspond to "strongly agree" and "strongly disagree", respectively. For other marker variables (MVs), -2 and 2 correspond to "strongly disagree" and "strongly agree".
d. These variables apply to workers. Therefore, the share of respondents with an occupation in manufacturing, construction, maintenance, or farming being 3.5% indicates that 7.2% (= 3.5%/48.8%) of workers have such occupations.
e. Having a condition or handicap that makes it difficult to travel outside the home.
f. Among five options (in descending order): excellent, very good, good, fair, and poor.
g. For these variables, the actual counts are used in the models, rather than the categorical variables shown in the table.
h. Following the definition from Kim et al. (2019), comprising 18 counties.
i. Having sufficient vehicles indicates that the number of vehicles equals the number of household members 16 years old or older, while having surplus vehicles indicates that the number of vehicles is larger than that.
j. The household VMD variable itself ranges from 0 to 207,641 (with a mean and standard deviation of 21,205 and 16,699).

Appendix B Distributions of vehicle count by household income and education level

		N	Count				Share (%)			
			0	1	2	3+	0	1	2	3+
	Sample	1245	32	404	473	336	2.6	32.4	38.0	27.0
Annual household income	$0-50K	508	30	262	136	80	5.9	51.6	26.8	15.7
	$50-100K	401	1	105	173	122	0.2	26.2	43.1	30.4
	$100K+	320	0	32	158	130	0.0	10.0	49.4	40.6
	Missing	16	1	5	6	4	6.3	31.3	37.5	25.0
Highest education (of any member)	Up to high school	176	19	78	49	30	10.8	44.3	27.8	17.0
	Some college	332	8	128	100	96	2.4	38.6	30.1	28.9
	Bachelor's degree or higher	737	5	198	324	210	0.7	26.9	44.0	28.5

Note: Statistics are from the whole *recipient* sample (N = 1,245), on which vehicle ownership models are estimated.

Appendix C Hybrid/electric vehicle adoption by household income and education level

		N	Count		Share (%)	
			No	Yes	No	Yes
	Sample	1206	1150	56	95.4	4.6
Annual household income	$0-50K	473	464	9	98.1	1.9
	$50-100K	398	383	15	96.2	3.8
	$100K+	320	289	31	90.3	9.7
	Missing	15	14	1	93.3	6.7
Highest education (of any member)	Up to high school	152	150	2	98.7	1.3
	Some college	323	320	3	99.1	0.9
	Bachelor's degree or higher	731	680	51	93.0	7.0

Note: Statistics are from respondents (in the *recipient* sample) who drive and belong to a household with at least one vehicle (N = 1,206), on which hybrid/electric vehicle adoption models are estimated.

Appendix D Crosstabulation of household income and education

		Annual household income				Total
		$0-50K	$50-100K	$100K+	Missing	
Highest education (of any member)	Up to high school	147	22	5	2	174
	Some college	186	116	28	2	330
	Bachelor's degree or higher	175	263	287	12	725
	Total	508	401	320	16	1245

Note: Statistics are from the whole *recipient* sample (N = 1,245).