Transport Policy 133 (2023) 45–53

Contents lists available at ScienceDirect

Transport Policy

journal homepage: www.elsevier.com/locate/tranpol

Modeling vehicle-miles of travel accounting for latent heterogeneity

Fatemeh Nazari [a,*], Abolfazl (Kouros) Mohammadian [b]

[a] Department of Civil Engineering, University of Texas Rio Grande Valley, USA
[b] Department of Civil, Materials, and Environmental Engineering, University of Illinois at Chicago, USA

ARTICLE INFO

Keywords:
Vehicle-miles of travel
Taste heterogeneity
Latent class regression
Fuel cost
Telework

ABSTRACT

Vehicle use is associated with negative externalities such as traffic congestion, air pollution, and greenhouse gas emissions. Particularly in the U.S. as a car-oriented country, vehicle use — in terms of vehicle-miles of travel (VMT) — has been on the rise and is projected to increase in the future. To curb the VMT growth and mitigate the associated externalities, policy makers can design informed strategies based on VMT predicted by vehicle use models. However, traditional vehicle use models capture merely the *observed heterogeneity* across vehicle decision making units (e.g., individuals) and ignore the *latent or taste heterogeneity* sourced in individuals' attitudes and lifestyle preferences, which may cause biased and inconsistent results that mislead implications for policy makers. To address this research gap, the present study introduces a latent class regression model, where a probabilistic multinomial logit component *endogenously* classifies a sample of vehicle use observations so as to be homogeneous within and heterogeneous across the classes with respect to VMT. At the same time, a finite set of linear regression equations in the number of the latent classes yields class-specific VMT. The model is estimated on a sample dataset from the State of California identifying three latent classes, verifying the hypothesis of positing vehicle use on both observed and unobserved heterogeneity. The estimation results are analyzed to infer implications of potential policies aiming at reducing VMT through increasing fuel cost and switching to telework, and to evaluate the efficiency of resource allocation to policies by targeting different classes with distinctive characteristics.

1. Introduction

Car-oriented societies experience negative externalities, such as traffic congestion, air pollution, greenhouse gas (GHG) emissions, and dependence on oil import. For instance, 82% of daily trips in the U.S. are accomplished by cars, half of which are single-occupant, according to the most recent travel trends concluded from the national household travel survey (McGuckin and Fucci, 2017). Over time, vehicle use — in terms of vehicle-miles of travel (VMT) — has shown an increasing trend in the U.S., with a 20% increase over the past decade (US FHWA, 2022). Moreover, VMT is anticipated to further increase with the advent of emerging transportation technologies, for instance, by adding deadhead miles of unoccupied autonomous vehicles (Noruzoliaee and Zou, 2022). To curb the VMT growth and mitigate the associated negative externalities, *policy makers* can design informed strategies based on the results of the VMT prediction models. For instance, the State of California envisions a 40% GHG emissions reduction by 2030 and continued to an 80% reduction by 2050 assuming 5% and 15% reductions in VMT, respectively (California Air Resources Board, 2021). To meet that goal, VMT reduction strategies are suggested which are related to land use, such as residential density and land-use mix diversity, as well as to transportation systems, such as public transit improvements, road pricing, and programs aimed at changing people's travel choices (Salon et al., 2012; Byars et al., 2017).

However, these policy strategies need to be effectively designed by tailoring them for the population segments with distinctive characteristic profiles and lifestyle preferences. To do so, there is a research need for an in-depth understanding of the process of vehicle use decision made by households (or individuals). In addition to the policy relevance of such research effort, the expected findings are further intriguing to *travel behavior analysts* due to the key role of vehicle use in explaining the individuals' travel behavior. In particular, the relevant literature on the

* Corresponding author.
E-mail addresses: fatemeh.nazari@utrgv.edu (F. Nazari), kouros@uic.edu (A.(K. Mohammadian).

https://doi.org/10.1016/j.tranpol.2023.01.005
Received 22 December 2021; Received in revised form 2 January 2023; Accepted 11 January 2023
Available online 14 January 2023
0967-070X/

vehicle ownership problem[1] signifies the role of the vehicle use decision in travel behavior-related decisions in both short- and long-term such as adoption of autonomous vehicles (Nazari et al., 2018b).

1.1. Vehicle use modeling

One approach to modeling vehicle use is through investigating the problem at the aggregate level. Rentziou et al. (2012) estimate a random parameters model to predict VMT of various road functional classes in the U.S., which are then used for assessing the influence of policies governing fuel tax and population density on future energy consumption and GHG emissions. A recent study verifies a hypothesis that vehicle use can be explained mainly by economic factors such as gasoline price (Bastian et al., 2016).

Alternatively, vehicle use can be modeled at the disaggregate level,[2] which is the focus of the present study, as a function of demographic attributes of decision makers, built environment factors, and vehicle attributes. Mannering (1983) investigates vehicle use of two-vehicle households in the U.S. by simultaneously estimating two linear equations. The findings signify the influences of household income and vehicle fuel efficiency on the allocation of vehicle use to the household vehicles, which are subsequently utilized for analyzing the relevant policy implications. Focused on multi-vehicle households in the U.S., Greene and Hu (1985) estimate linear equations and conclude that a 25% increase in gasoline price leads to a 5% decline in vehicle use. Guo (2013) explores VMT by estimating two linear equations on households with and without garage parking space to suggest policies on residential parking in New York. The results reveal the larger impact of street parking than access to garage on trips made by vehicles, and less vehicle use for households without access to off-street parking. Singh et al. (2018) present a joint modeling framework of the household residential location choice and VMT in New York, which is considerably affected by socio-demographic attributes and built environment factors. In another joint modeling system of a quadratic generalized multilevel structural equation, Zhang et al. (2021) evaluate the impacts of land use densification policies on vehicle use in Beijing, China. The impact of such policies on vehicle use reduction in low-density neighborhoods are found to be significant but indirect through the mediation of other travel decisions.

1.2. Latent or taste heterogeneity

While providing valuable insights, most of the existing studies on vehicle use can capture merely the *observed heterogeneity* across the decision-making units, e.g., individuals, by associating the model outcome to their observed attributes such as income. However, the decision process might be further influenced by, for instance, the individuals' attitudes and lifestyle preferences unknown to the analyst, which are referred to as *the latent (unobserved)* or *taste heterogeneity*. Ignoring taste heterogeneity in the estimation process may cause bias and in turn lead to inconsistent model estimation results and misleading implications for analysts and policy makers (Greene, 2000). The earlier attempts to address this issue exogenously segment a sample dataset into heterogeneous classes based only on the observed attributes and then estimate class-specific vehicle use models. Yet, this "deterministic" segmentation cannot fully account for latent heterogeneity, thereby calling for a "probabilistic" model.

A promising solution is employing latent class or finite mixture models to probabilistically segment a sample into a finite number of mutually exclusive classes to have attributes and preferences homogenous within and heterogeneous across the classes (see Heckman and Singer (1986) for a related theoretical discussion). The latent class model sometimes may be superior to its counterparts[3] in addressing the heterogeneity issue due to yielding more intuitive market segmentation (Greene and Hensher, 2003).

The contribution of the present study is to close the above-discussed gap by presenting the first empirical attempt to relate the vehicle use decision to both observed and unobserved/taste heterogeneities through estimating a latent class model. The study outcome, which is the continuous variable describing VMT, requires the model specification to be a latent class regression (LCR) (McCullagh and Nelder, 1983; DeSarbo et al., 1989; Wedel and Kistemaker, 1989; Wedel and DeSarbo, 1994). There exists limited applications of the LCR model in, for instance, sociology (Yamaguchi, 2000) and logistics theory (Garver et al., 2008). The empirical analysis of the proposed model identifies three heterogeneous classes with respect to vehicle use verifying the hypothesis that both observed and latent heterogeneities shape vehicle use. Moreover, the study findings can be leveraged for an effective design of policies tailored for the identified classes instead of a one-shot policy for the entire sample. The policies are used to evaluate the goal of the State of California for the 5% reductions in VMT through increasing fuel cost and telework.

The next section presents the methodology of the LCR model. The empirical estimation on a dataset described in section 3 is analyzed in section 4. The paper concludes in section 5.

2. Methodology

Fig. 1 shows the conceptual framework of the LCR model entailing a multinomial logit (MNL) model which *endogenously* segments the sample of vehicles into a finite number of latent classes so as to be homogeneous within and heterogeneous across the classes with respect to vehicle use (see Ben-Akiva et al. (2002) for a discussion on the endogeneity bias). At the same time, a finite set of linear regression equations in the number of the latent classes yields the vehicle use for each class. The simultaneous estimation of the MNL model and the class-specific linear equations avoids any *measurement error* possibly caused by the sequential estimation (Greene, 2000).

The LCR model is built on the linear regression equation $y_i = \boldsymbol{x}_i'\boldsymbol{\alpha} +$

[1] The vehicle ownership problem has various aspects, which are comprehensively discussed in a review study by Anowar et al. (2014), such as: (i) vehicle ownership level (Zhang et al., 2017); (ii) choice of vehicle attributes such as fuel type (Nazari et al., 2018a, 2019); (iii) vehicle transaction decision (Gilbert, 1992); and (iv) vehicle utilization or use (Zhang et al., 2021), which is the focus of the present study.

[2] Due to the possible endogeneity of vehicle use to the other aspects of the vehicle ownership decision, a stream of disaggregate studies analyze a hypothesis on this endogeneity by jointly estimating vehicle use and, for instance, vehicle ownership level (Liu and Cirillo, 2016). This is out of the scope of the present study.

[3] An alternative is the random parameters model, which allows parameters to have a continuous random distribution over the sample (McFadden and Train, 2000). This model has the flexibility advantage over the latent class model, since the latter approximates the underlying continuous random distribution of parameters with a discrete one. The latent class model has the semiparametric specification superiority to the fully parametric random parameters model developed earlier, which is limited to have "a priori" assumption about the mixture distribution of the parameters, thus cannot find the source of heterogeneity. More recent versions of the random parameters model, however, can effectively identify these sources. Examples are random parameters model with heterogeneity in means and variances (Waseem et al., 2019), correlated random parameters ordered probit model (Fountas et al., 2018), grouped random parameters bivariate probit model (Sarwar et al., 2017), and correlated random parameters model with heterogeneity in means (Ahmed et al., 2021). Regardless, both latent class and random parameters models *implicitly* treat taste heterogeneity, while an approach to *explicitly* address the issue incorporates the underlying observed psychometric indicators into choice models (Ben-Akiva et al., 2002). Interested readers in the recent methodological advancements are referred to Washington et al. (2020).

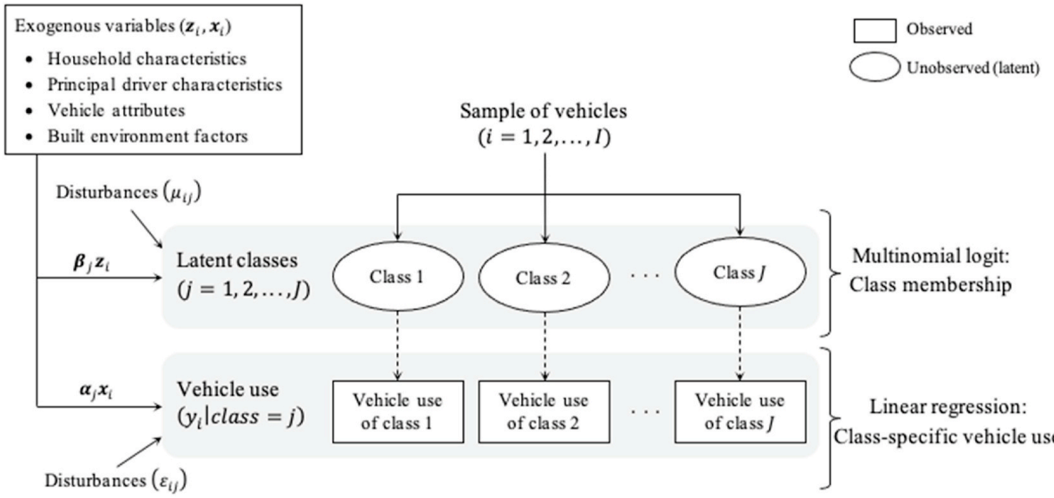


Fig. 1. Conceptual framework: Latent class regression model of vehicle use.

ε_i, $\varepsilon_i \sim N[0,\ \sigma_i^2]$, wherein y_i denotes vehicle use for vehicle $i\,(i = 1, 2, 3, \ldots, I)$ in terms of the logarithm of annual VMT, which is explained by a linear function of the set of the exogenous variables $(\boldsymbol{x_i})$ multiplied by the set of the associated parameters $(\boldsymbol{\alpha})$. The logarithmic specification follows the former vehicle use models (Kim and Kim, 2004; Spissu et al., 2009; Eluru et al., 2010; Singh et al., 2018).

Estimation of the above equation results in only one set of parameters $\boldsymbol{\alpha}$ for all sample vehicles. However, vehicle use may vary across heterogeneous classes of vehicles with various features which are held by households and driven by individuals with distinctive characteristics. In other words, there might be a taste heterogeneity in vehicle use that, if ignored, the estimation results might be unreliable and inconsistent (Greene, 2000). To tackle, the equation is extended to the LCR (or finite mixture) model assuming a discrete random distribution for the parameters (McCullagh and Nelder, 1983; DeSarbo et al., 1989; Wedel and Kistemaker, 1989; Wedel and DeSarbo, 1994). The LCR endogenously segments I vehicles into J classes and simultaneously regresses a linear regression of VMT for each class.

The probability of VMT for vehicle i belonging to class j is expressed as equation (1). The first term is the probability of VMT conditional on placing in class j (equation (2)). The second term, which is the probability of vehicle i belonging to class j ($j = 1, 2, \ldots, J$), is an MNL (equation (3)). The utility equation of the MNL is written as $U_{ij} = z_i^{'}\boldsymbol{\beta}_j + \mu_{ij}$, wherein z_i is the set of the exogenous variables, $\boldsymbol{\beta}_j$ is the set of the corresponding parameters, and μ_{ij} is the error component with the extreme value distribution.

$$Prob\ (y_i, class = j) = Prob\ (y_i|class = j) \bullet Prob\ (class = j) \tag{1}$$

$$Prob\ (y_i|class = j) = N\left[\boldsymbol{x}_i^{'}\boldsymbol{\alpha}_j, \sigma_{ij}^2\right] = \frac{1}{\sigma_{ij}\sqrt{2\pi}} \bullet \exp\left(-\frac{\left(y_i - \boldsymbol{x}_i^{'}\boldsymbol{\alpha}_j\right)^2}{2\sigma_{ij}^2}\right) \quad i = 1, 2, \ldots, I;\ j = 1, 2, \ldots, J \tag{2}$$

$$Prob\ (class = j) = \frac{\exp\left(z_i^{'}\boldsymbol{\beta}_j\right)}{\sum_{j=1}^{J} \exp\left(z_i^{'}\boldsymbol{\beta}_j\right)} \quad j = 1, 2, \ldots, J \tag{3}$$

The log-likelihood function of the LCR model is written as equation (4), which is solved using maximum likelihood to estimate the parameters.

$$\mathrm{Ln}\ L = \sum_{i=1}^{I} \ln \sum_{j=1}^{J} Prob\ (y_i, class = j) = \sum_{i=1}^{I} \ln \sum_{j=1}^{J} Prob\ (y_i|class = j) \bullet Prob\ (class = j) \tag{4}$$

3. Data

The LCR model of vehicle use is empirically estimated on California Vehicle Survey (National Renewable Energy Laboratory, 2019). We use a sample dataset on 7387 residential light duty vehicles which are held by 3963 households and driven by 6444 individual principal drivers (i.e., the household members driving the vehicles more than the other members).

3.1. Vehicle attributes

The statistical distribution of the sample vehicles is presented in

Table 1
Statistical distribution of vehicle attributes.

Variable description	Category	# observations	Share (%)
Annual vehicle-miles of travel or VMT	Mean = 9527.921, SD* = 7971.461	–	–
Logarithm of annual VMT⁺	Mean = 8.879, SD = 0.796	–	–
Vehicle age (years)	Mean = 2.937, SD = 4.599	–	–
Vehicle fuel efficiency (miles per gallon)	Mean = 31.127, SD = 21.533		
Fuel cost per mile (dollars)	Mean = 0.183, SD = 0.072	–	–
Fuel cost per mile (dollars) for households with:			
low income (<$50K)	Mean = 0.027, SD = 0.073	–	–
medium income ($50K ≤ < $150K)	Mean = 0.087, SD = 0.104	–	–
high income (≥$150K)	Mean = 0.052, SD = 0.089	–	–
Vehicle ownership type at acquisition time	Purchased new	3876	52.47
	Purchased used or previously owned	2446	33.11
	Leased (new or used)	861	11.66
	Other	204	2.76

Sample size = 7387.
*SD: standard deviation.
⁺ Model outcome.

Table 1. *The logarithm of annual VMT* specifies the model outcome with the average and standard deviation values observed respectively at 8.879 and 0.796 per year. To capture the economic aspects of vehicle use, namely operating cost, a factor on *fuel cost per mile* (in dollars) is calculated for each vehicle, which is equal to the average fuel price (in dollars) in the residence county of the household who holds the vehicle divided by the vehicle fuel efficiency. This factor is computed for three household groups with three annual income levels. The lowest and the highest average fuel costs are respectively observed for households with low- and medium-income levels.

3.2. Household-related factors

The sample households are defined by socio-economic characteristics and the built environment factors. The statistical distribution of the relevant factors is presented in Table 2. For instance, 82.44% of the households have access to parking space at their residence.

3.3. Individual-related factors

The individuals, who are the principal drivers of the sample vehicles, are characterized by their socio-economic attributes (Table 3).

4. Estimation results

4.1. The number of classes

The number of classes for a latent class model is exogenously determined through trying various models to select the best one based on both model fitness and interpretability. The former is measured by the log-likelihood value at the estimated parameters ($LL(\beta)$), the Akaike information criterion (AIC), which is equal to $(2K - 2LL(\beta))/N$, and the Bayesian information criterion (BIC), which is calculated as $(-2LL(\beta) + K\ln(N))/N$. The indices K and N respectively denote the number of model parameters and the number of observations. The superior model has smaller values of $-LL(\beta)$, AIC, and BIC, conditional on providing more intuitive estimation results with better interpretability (Greene, 2000). The statistics of five models with similar specifications except for the number of classes is summarized in Table 4, indicating that the 3-class model provides better behaviorally interpretable results despite marginal improvements in $LL(\beta)$, AIC, and BIC for higher number of classes.

Table 2
Statistical distribution of households-related factors.

Variable description	Category	# observations	Share (%)
Socio-demographic characteristics			
Household structure		–	–
# kids (age <5 years)	Mean = 0.098, SD* = 0.377	–	–
# teenagers (5 years ≤ age <12 years)	Mean = 0.126, SD = 0.424	–	–
# young children (12 years ≤ age <16 years)	Mean = 0.089, SD = 0.332	–	–
# adults (age ≥16 years)	Mean = 1.992, SD = 0.818	–	–
Decision on future household vehicle(s)	Solely by one member	1678	42.34
	Primarily by one member	1027	25.91
	Shared among members	1258	31.75
Built environment factors			
Access to parking space at residence	Yes	3267	82.44
	No	696	17.56
Residential location	Urban area	3220	81.25
	Suburban area	594	14.99
	Rural area	149	3.76

Sample size = 3963.
*SD: standard deviation.

Table 3
Statistical distribution of individual principal drivers' socio-economic characteristics.

Variable description	Category	# observations	Share (%)
Gender	Male	3175	49.28
	Female	3200	49.65
	Other	69	1.07
Employment status	Full-time	2856	44.32
	Part-time	721	11.19
	Self-employed	439	6.81
	Not paid	2428	37.68
Job location type	One location	2938	45.59
	Regularly varied	665	10.32
	Telework	413	6.41
	Not paid	2428	37.68
Educational attainment	High school graduate or less	701	10.88
	Associate degree	1749	27.14
	College graduate	1988	30.85
	Post-graduate	2006	31.13
Ethnicity	Hispanic, Latino, or Spanish origin	636	9.87
	Non-Hispanic	5534	85.88
	Prefer not to answer	274	4.25
Race	White	4382	68.00
	Asian	1090	16.92
	African American	202	3.13
	American Indian	114	1.77
	Other	656	10.18

Sample size = 6444.

4.2. Interpretation of results

4.2.1. Class-specific vehicle use

The estimation results of the vehicle use specific for the three latent classes and the pooled sample of vehicles are presented in Table 5. The set of exogenous variables is restricted to be the same across the equations during the estimation process. Almost all parameters are found to be statistically significant at a 95% confidence interval, indicating their statistical reliability. Comparing the estimated parameters reveals that the taste variation in VMT across the three classes is mainly rooted in distinctive attributes of vehicles belonging to each class and held by households and driven by individuals with different characteristics, as shown in Table 6 and discussed in section 4.2.2.

The *constant* parameters vary across the classes which are explained by unknown factors omitted from the model. The associated absolute values in a descending order belong to classes 2, 1, and 3, reflecting the corresponding order of vehicle use if all exogenous factors entering the model are identical.

Among the vehicle attributes influencing vehicle use, *vehicle fuel cost per mile* (in dollars), which partly reflects commodity price of travel, negatively affects VMT of the members of the second and the third classes. This is in line with the findings of the previous studies in the U.S. (Mannering, 1983; Mannering and Winston, 1985; Kim and Kim, 2004; Hang et al., 2016; Liu and Cirillo, 2016; Bastian et al., 2016). However, the magnitudes of the corresponding parameters vary across the three income levels for the two classes, signifying the two-fold heterogeneity. Within class 2, the largest impact of fuel cost is on the vehicle use of those with medium-income level, followed by high- and low-income levels. In contrast, the largest and the smallest impacts of fuel cost on vehicle use of members of class 3 are respectively on the low- and the medium-income households. The results further suggest that the vehicle use response of the members of the third class to fuel cost almost follows the same pattern found for the pooled sample.

Vehicle age affects the vehicle use of only the members of the second

Table 4
Summary statistics for latent class regression models with different number of classes.

	Model without classification	Models with latent classification			
Number of classes	1	2	**3**	4	5
Number of parameters (K)	16	40	**64**	88	112
$LL(\beta)$	−8431.645	−7946.685	**−7853.916**	−7221.023	−7032.506
AIC	2.287	2.162	**2.144**	1.979	1.934
BIC	2.302	2.200	**2.204**	2.061	2.039

Note: Boldfaced column indicates the model with the best number of classes.

Table 5
Estimation results of the latent class regression model: Class-specific vehicle use.

Exogenous variable	Latent class regression						1-class regression	
	Class 1		Class 2		Class 3			
	coef.	t-stat	coef.	t-stat	coef.	t-stat	coef.	t-stat
Constant	9.265***	517.43	9.674***	145.80	8.568***	137.42	8.975***	300.50
Vehicle attributes								
Fuel cost per mile (dollars) for households with:								
low income (<$50K)	−0.040	−0.41	−0.755**	−2.13	−1.287***	−5.57	−1.161***	−7.64
medium income ($50K ≤ < $150K)	−0.064	−0.90	−1.107***	−4.48	−0.654***	−3.39	−0.795***	−6.73
high income (≥$150K)	0.056	0.74	−1.000***	−3.90	−0.902***	−4.05	−0.901***	−6.95
Logarithm of vehicle age (years)	−0.004	−0.40	−0.863***	−22.67	0.021	0.55	−0.205***	−11.64
Vehicle ownership type at acquisition time								
Purchased used or previously owned	−0.019	−1.22	0.282***	5.33	0.036	0.82	0.094***	3.78
Leased (new or used)	0.050***	3.20	−0.052	−0.74	0.186***	2.66	0.113***	3.88
Built environment factors								
Access to parking space at residence								
Yes	0.007	0.55	0.007	0.14	0.095**	2.16	0.055**	2.39
Residential location								
Rural area	0.073***	2.85	0.371***	3.88	−0.140*	−1.91	0.028	0.60
Principal driver characteristics								
Employment status								
Full-time employed	0.008	0.76	0.372***	8.86	0.353***	10.47	0.311***	16.31
Self-employed	−0.008	−0.31	0.178**	2.10	0.258***	3.47	0.191***	4.64
Job location type								
Regularly varied	0.027*	1.66	0.165***	2.71	0.235***	4.74	0.198***	6.65
Telework	−0.033	−1.22	−0.227***	−2.69	−0.072	−0.99	−0.097**	−2.34
Ethnicity								
Hispanic, Latino, or Spanish origin	0.032**	2.02	0.134*	1.94	0.085*	1.72	0.097***	3.24
Race								
Asian	−0.030**	−2.15	−0.102*	−1.92	0.054	1.18	0.009	0.38

Note: ***, **, and * respectively indicate parameters statistically significant at the 99%, 95%, and 90% confidence intervals.

class, meaning that they drive older vehicles less than newer ones. A similar behavior is found for the pooled sample, yet the associated impact is smaller, as suggested by the smaller absolute value of the corresponding parameter than that of the second class. The negative influence of this factor on vehicle use is consistent with the findings of the past studies in the U.S. (Greene and Hu, 1985; Hang et al., 2016). Besides, the logarithmic form of this factor implies the associated diminishing impact as vehicles become older.

Vehicle ownership type appears in the model as two dummy variables, which are found to be a possible source of heterogeneity across the three classes. The first variable equals 1 for the vehicles that are either purchased as used vehicles or previously owned and 0 otherwise. This variable appears in the equation of the members of class 2 with a positive sign, meaning that this ownership type increases VMT. This factor influences the pooled sample in the same direction, yet with a smaller magnitude. The second variable takes the value 1 if a vehicle is leased. This variable enters the VMT equations of the first and the third classes with a larger parameter for the third class, revealing the associated positive effects, especially on the third class. Besides, this factor positively affects VMT of the pooled sample with a magnitude larger and smaller than those of the first and the third classes, respectively.

As for the built environment factors influencing vehicle use, the first factor appears in the model as a dummy variable taking the value 1 for those with *access to parking space at their residence*. This factor, which reflects the availability or affordability of parking supply, is found to positively affect only the members of class 3. The same impact is also found for the pooled sample, though with a smaller magnitude. This finding is previously reported by a study on households residing in the New York City (Guo, 2013).

The residential location of the households appears as a dummy variable taking the value 1 for those who reside in rural areas. The results signify the influential role of this factor in the equations of the three classes yet with different signs, implying the presence of taste variation. Living in rural areas increases VMT of the persons who belong to the first and the second classes, especially the second class. This might be due to the fact that rural areas usually have lower residential and population density and thus, have lower accessibility to and availability of public transit due to the lack of sufficient infrastructure support for public transit. This is in line with the previous studies which report less vehicle use for urban areas (Mannering, 1983; Greene and Hu, 1985; Schimek, 1996; Stevens, 2017; Singh et al., 2018; Zhang et al., 2021). Conversely, residing in rural areas decreases VMT of the members of class 3, whose smaller VMT, which seems contradictory with the common sense, might be due to the gradual change in their behavior over time by adjusting their travel needs for rural residence through less frequent activity involvement. In contrast, VMT of the pooled sample is found not to be

Table 6
Estimation results of the latent class regression model: Multinomial logit of class membership.

Exogenous variable	Class 1		Class 2		Class 3	
	coef.	t-stat	coef.	t-stat	coef.	t-stat
Constant	−1.933***	−9.57	−0.600**	−2.08		
Household characteristics						
Household structure						
# kids (age <5 years)	0.203*	1.88	−0.391*	−1.73	0.000	–
# teenagers (5 years ≤ age <12 years)	0.243**	2.36	0.050	0.32	0.000	–
# adults (age ≥16 years)	0.203***	3.52	−0.096	−1.05	0.000	–
Decision on future household vehicle(s)						
Solely by one household member	−0.058	−0.49	−0.509***	−3.05	0.000	–
Principal driver characteristics						
Gender						
Male	−0.284***	−2.66	−0.324**	−2.36	0.000	–
Educational attainment						
College graduate	0.314**	2.45	0.534***	3.07	0.000	–
Post-graduate	0.367***	2.73	0.787***	4.67	0.000	–
Predicted class membership (%)	14.710		26.866		58.425	

Note: ***, **, and * respectively indicate parameters statistically significant at the 99%, 95%, and 90% confidence intervals.

influenced by this factor.

Switching the focus to the principal drivers' characteristics, *the employment status* appears in the model in the form of two dummy variables; one taking the value 1 for full-time employment and the other equals 1 for self-employment. Being a full-time employee or self-employed is expected to increase vehicle use presumably due to work trips (Giuliano and Dargay, 2006; Zhang et al., 2021). The estimation results verify more VMT for both groups, especially the full-time employees who are the members of classes 2 and 3. The magnitude of the relevant parameters is almost similar for both classes, which is slightly larger than that of the pooled sample. Conversely, self-employed persons belonging to the third class drive more than those who are places in the second class, as indicated by the corresponding larger parameter.

Job location type of principal drivers enters the model as two dummy variables. One variable equals 1 for individuals with varied job locations and the other variable takes the value 1 for individuals who telework. As expected, the former increases VMT for all three classes and the pooled sample, whereas the latter decreases VMT, but only for class 2 and the pooled sample. Also, there is a taste variation with respect to both factors across the three classes. The varied job location increases VMT of classes 3, 2, and 1 in a descending manner. Also, smaller VMT is likely tied with telework if the person belongs to class 2, whose behavior almost mimics the travel pattern of the pooled sample.

Ethnicity, defined as a dummy variable, which takes the value 1 for Hispanic or Latino individuals, is found to have positive effect on VMT of classes 2, 3, and 1 in a descending order, signifying the ethnicity-related heterogeneity. Besides, the magnitudes of this factor on all three classes are different from that of the pooled sample.

One *race* type is found to be statistically significant in the model as a dummy variable taking 1 for Asians. The appearance of this variable in the equations of classes 1 and 2 with the negative signs implies that Asians who belong to these two classes, especially class 2, likely drive less than others. Similarly, previous studies report the negative tendency of Asians to vehicle use in the State of California (Brownstone and Golob, 2009; Spissu et al., 2009). Besides, the race factor is found to be insignificance in the VMT of the pooled sample.

4.2.2. *Multinomial logit of class membership*

The LCR model also entails the MNL component to endogenously distinguish the three vehicle classes based on the characteristics of the households holding the vehicles and their principal drivers (Table 6). The estimation results reveal that almost all parameters are statistically reliable attributed to their statistical significance at a 95% confidence interval. The set of the exogenous variables are restricted to be the same across the equations during the estimation process.

The *constant* parameters reveal that if all conditions are equal, the probability of belonging to classes 1 or 2 is less than that of class 3, and the lowest probability is for class 1.

The three classes are distinguished in part by *household structure* which is defined by the numbers of household kids, teenagers, and adults. It is found that vehicles held by households with more kids are most (respectively, least) likely placed in class 1 (respectively, class 2). The numbers of teenagers and adults in a household appear only in the equation of the first class with a positive sign, meaning that the presence of more teenagers and adults in a household leads to placing its vehicles in the first class. The class membership is also influenced by *the role of household members in future vehicle decision(s)*, which is significant as a dummy variable taking the value 1 for households whose vehicle decisions are made solely by one member. Specifically, the vehicle(s) of the households with a positive response to this factor are less likely categorized in the second class.

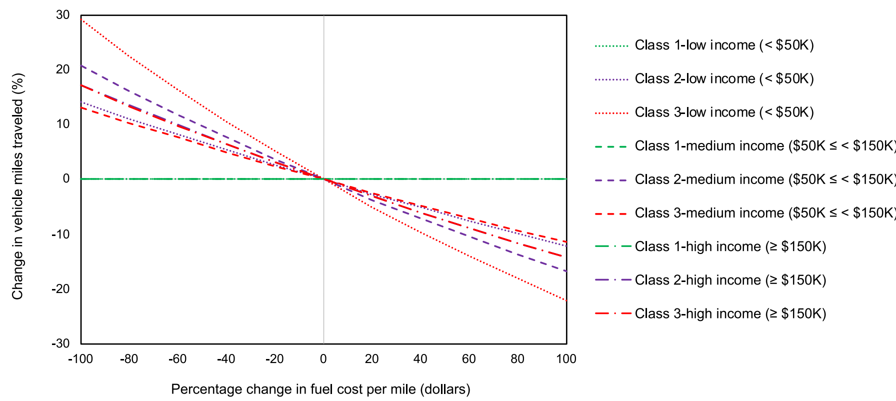
Gender of principal driver, defined as a dummy variable equaling 1 for males, appears in the equations of classes 1 and 2. Male drivers less likely drive vehicles categorized in these two classes, especially class 2. *Education* is found to be influential in the equations of classes 1 and 2 as two dummy variables; one is equal to 1 for the college graduates with a 4-year degree and the other takes the value 1 for the post-graduate individuals. Both factors positively influence the membership in these two classes, with a larger impact on the second class, indicating that those who are highly educated most likely belong to these two classes, especially class 2.

4.3. *Sensitivity analysis and policy implications*

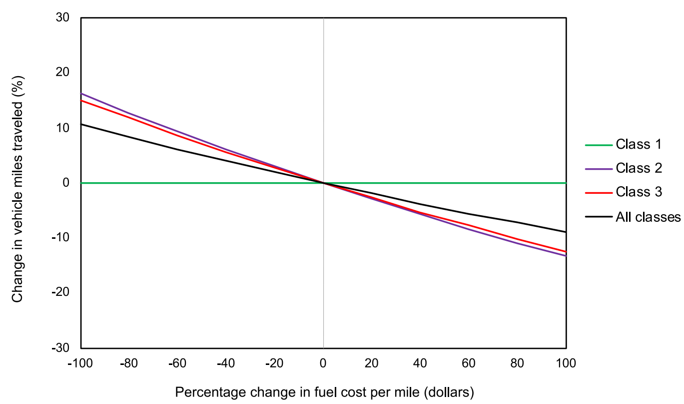
The estimated LCR model can be further analyzed by measuring the sensitivity of VMT of the three latent classes with respect to two policy-sensitive factors found significant in the model, namely, *fuel cost per mile in dollars* and *telework*. The results can be utilized for inferring policy implications, based on which policies can be effectively designed by tailoring them for the three classes instead of a one-shot policy for the entire sample.

Assuming that the impact of the fuel cost change is merely through the direct change in VMT, fuel cost is changed within the range of −100%–100% in 10% intervals to predict the VMT of each vehicle in the sample in response to the fuel cost variations. Then, the percentage changes (compared to a base case with no changes in fuel cost) in the cumulative VMT are calculated for nine groups of vehicles. These groups are the three latent classes which are further categorized by the three household annual income levels, i.e., low, medium, and high levels. The results are displayed in Fig. 2(a), which overall reveal that the order of VMT variations for the nine groups follows the magnitudes of the associated parameters (presented in Table 5).

The highest impact of the fuel cost change is found to be in the VMT of the third-class vehicles and held by low-income households, as shown by the dotted red line. Specifically, the VMT of this group decreases by 21.94% if fuel cost doubles (given that the average fuel price of the sample is 4.507 dollars per gallon, doubling fuel cost means increasing the average value to 9.014 dollars per gallon). On the other extreme, the fuel cost reduction down to almost 0 dollars leads to the 29.10% increase



(a) Three classes with three income levels



(b) Three classes and all classes

Fig. 2. The modeled VMT sensitivity to fuel cost.

in the VMT of this group. This sensitive group likely encompasses vehicles held by households with an income less than $50K, less teenagers and adults, and whose vehicle decisions are made solely by one household member, as well as vehicles driven by male principal drivers who are not highly educated. The second highest sensitive group to the fuel cost variations is vehicles of class 2 and held by medium-income households (shown by the dashed purple line). The fuel cost increase and decrease up to 100% lead to respectively 16.67% less and 20.62% more VMT of the vehicles of this group, whose principal drivers are probably not-highly-educated males and are held by households with an income between $50K and $150K, less kids, and whose vehicle decisions are not made solely by one member.

Moreover, Fig. 2(b) exhibits the percentage change in the cumulative VMT of the three classes regardless of the income level, as well as that of the sample vehicles altogether, in response to the changes in fuel cost. The lowest impact is on the vehicles placing in class 1, shown by the solid green line, which are held by households with more kids, teenagers, and adults, and are driven by highly educated male principal drivers.

The changes in VMT of all classes, which is depicted by the black solid line, suggest a 10.71% decrease and an 8.84% increase in VMT in response to the extreme changes (± 100%) in fuel cost. This can indicate the fuel cost elasticity of the whole sample as 10.71%, which is consistent with the findings of the previous research efforts in the U.S. For instance, Mannering (1983) finds the 11.3% fuel price elasticity in multivehicle households and Greene and Hu (1985) predict a 5% decline in vehicle use due to a 25% increase in gasoline price. More recently, Bastian et al. (2016) finds a gasoline price elasticity of 14% in the U.S. Besides, it can be concluded that, without narrowing down the target population to any specific group of people, the changes in fuel cost would induce at most a 10.71% change in VMT. Clearly, this is an underestimation of the change in VMT vehicles of classes 2 and 3, which can be captured through the LCR model.

Overall, it can be concluded that there exists heterogeneity across the nine groups with respect to the changes in VMT in response to change in fuel cost. This finding can be leveraged for suggesting effective policies specific to the nine groups instead of a uniform policy for the entire sample. For instance, targeting a 5% VMT reduction, which is the goal of the State of California by 2030 (California Air Resources Board, 2021), requires an average 60% increase in fuel cost (e.g., through increasing fuel tax) for the entire sample. Alternatively, this goal can be effectively met if fuel cost change is determined specific to the nine groups. In details, VMT can be reduced by 5% through increasing fuel cost, for instance, by almost 20% and 40% for the members of class 3 who have respectively low- and medium-income levels. On implementation of such group-specific policies, one suggestion is applying the required changes in fuel cost on the counties of the State by matching the demographic attributes of each county with the characteristic profiles of each of the nine groups.

The estimated model is further analyzed by focusing on a factor describing the principal drivers' job location, in particular, telework which is usually assessed as a policy strategy to directly affect VMT. For individuals who do not telework, the telework time is increased up to 100% in 10% increments. Then, the corresponding changes in their VMT is calculated, which is distinguished for the members of the three classes shown by the three colored lines in Fig. 3. As an example, a 20% increase in telework in exchange for a 20% less work time in the current job location results in 4.53% less VMT for those individuals who drive vehicles of class 2 and work either at only one or regularly varied locations.

The most sensitive class to the telework factor involves individuals of the second class, such that their VMT drops by 20.48% if they switch fully to telework. This highlights the maximum expected change in VMT

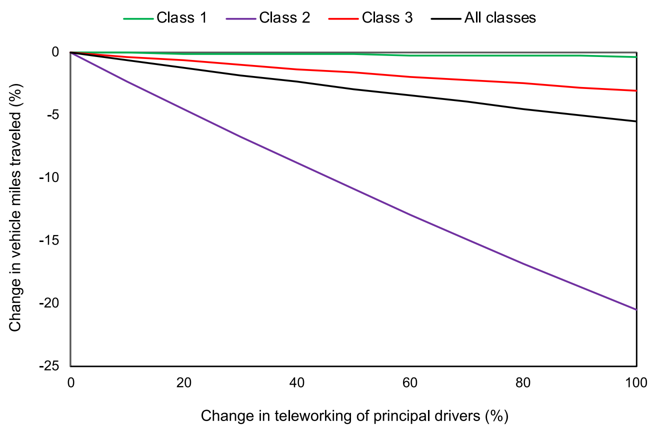


Fig. 3. The modeled VMT sensitivity to teleworking of principal drivers of the three classes and all classes.

through devising policies that encourage telework and are geared towards the second-class members, which likely encompasses vehicles held by households with less kids whose vehicle decisions are not made solely by one member, and vehicles driven by non-male highly educated individuals.

Conversely, the first-class vehicles experience the slightest changes in VMT in response to the telework factor, as evidenced by the maximum decrease in VMT of vehicles in class 1 being less than 1% in response to switching to full telework. This determines the least sensitive class to telework policies, which encompasses vehicles held by households with more kids, teenagers, and adults, and driven by male individuals who are highly educated. The corresponding changes for the third class are also small, which is up to a 3.02% reduction in VMT.

Regardless of the class membership, the cumulative changes in VMT of all vehicles in response to the telework factor are shown by the black line. It can be concluded that employing telework policies without targeting any specific group of population is likely expected to lead to up to a 5.45% reduction in VMT.

These findings can be further utilized for recommending effective policies. For instance, the target of up to 17% VMT reduction only through telework, which is concluded by California Center for Jobs and the Economy (2020) as the potential contribution of telework to the VMT reduction, can be achieved, for instance, by an almost 80% increase of telework time only for the individuals who belong to class 2 with no need to target the entire sample. Given the characteristic profiles of the class 2 members, whose VMT is more sensitive to telework compared to the other classes, policy makers can effectively identify them and apply the telework policies to achieve higher returns in investment.

5. Conclusions

The present study updates the conventional vehicle use models, which are prone to bias due to the ignorance of latent or taste heterogeneity stemming from individuals' attitudes and lifestyle preferences. To do so, a latent class regression model is presented which hypothesizes that vehicle use — in terms of vehicle-miles of travel (VMT) — is affected by both observed and latent heterogeneities. The latent heterogeneity is inferred by a probabilistic market classification of the sample through a multinomial logit and the class-specific linear regression equations. The empirical estimation of the model on a sample dataset from the State of California supports the study hypothesis. The results are further utilized to assess the implications of policies targeted at reducing VMT by increasing fuel cost and telework specific to each of the three classes.

To address the study limitations, two directions are suggested for future research. On the modeling approach, further research can comparatively analyze the estimation results of the latent class model with the random parameters model (see footnote 3 for a relevant discussion). The outcomes are expected to shed lights on the applicability of the two models for policy analysis in the context of vehicle use. On the sample dataset, this study uses one collected in 2019 which is right before the COVID-19 pandemic. A future study can focus on examining whether and how the pandemic influences the individuals' travel behavior, specifically their vehicle use, by collecting more recent datasets measuring VMT during and after the pandemic. The results can be noteworthy from the policy-making perspective since the pandemic likely shifts the individuals' travel patterns, such as VMT, mainly due to the shift in job location from conventional workplaces to homes.

Author statement

Fatemeh Nazari: Conceptualization, Methodology, Data curation, Software, Writing-Original draft preparation.

Abolfazl (Kouros) Mohammadian: Conceptualization, Supervision.

Data availability

Data will be made available on request.

Acknowledgements

This research was funded in part by the US National Science Foundation (NSF 2112650). The opinions expressed are solely those of the authors, and do not necessarily represent those of NSF.

References

Ahmed, S.S., Cohen, J., Anastasopoulos, P.C., 2021. A correlated random parameters with heterogeneity in means approach of deer-vehicle collisions and resulting injury-severities. Anal. Method. Acc. Res. 30, 100160 https://doi.org/10.1016/j.amar.2021.100160.

Anowar, S., Eluru, N., Miranda-Moreno, L.F., 2014. Alternative modeling approaches used for examining automobile ownership: a comprehensive review. Transport Rev. 34, 441–473. https://doi.org/10.1080/01441647.2014.915440.

Bastian, A., Börjesson, M., Eliasson, J., 2016. Explaining "peak car" with economic variables. Transport. Res. Pol. Pract. 88, 236–250. https://doi.org/10.1016/j.tra.2016.04.005.

Ben-Akiva, M., Walker, J., Bernardino, A.T., Gopinath, D.A., Morikawa, T., Polydoropoulou, A., 2002. Integration of choice and latent variable models. Perpetual Motion: Travel Behav. Res. Opport. Appl. Challeng. 431–470.

Brownstone, D., Golob, T.F., 2009. The impact of residential density on vehicle usage and energy consumption. J. Urban Econ. 65, 91–98. https://doi.org/10.1016/j.jue.2008.09.002.

Byars, M., Wei, Y., Handy, S., 2017. State-level Strategies for Reducing Vehicle Miles of Travel (No. UC-ITS-RR-2017-27-A). Institute for Transportation Studies, National Center for Sustainable Transportation, University of California, Davis.

California Air Resources Board, 2021. 2020 mobile source strategy. Retrieved from. https://ww2.arb.ca.gov/sites/default/files/2021-12/2020_Mobile_Source_Strategy.pdf.

California Center for Jobs and the Economy, 2020. California Workers: Modernized Telecommuting Policies to Build Equity and Reduce Costs. Retrieved from. https://centerforjobs.org/wp-content/uploads/Telecommuting-FINAL.pdf.

DeSarbo, W.S., Oliver, R.L., Rangaswamy, A., 1989. A simulated annealing methodology for clusterwise linear regression. Psychometrika 54, 707–736. https://doi.org/10.1007/BF02296405.

Eluru, N., Bhat, C.R., Pendyala, R.M., Konduri, K.C., 2010. A joint flexible econometric model system of household residential location and vehicle fleet composition/usage choices. Transportation 37, 603–626. https://doi.org/10.1007/s11116-010-9271-3.

Fountas, G., Anastasopoulos, P.C., Abdel-Aty, M., 2018. Analysis of accident injury-severities using a correlated random parameters ordered probit approach with time variant covariates. Anal. Method. Acc. Res. 18, 57–68. https://doi.org/10.1016/j.amar.2018.04.003.

Garver, M.S., Williams, Z., Taylor, G.S., 2008. Employing latent class regression analysis to examine logistics theory: an application of truck driver retention. J. Bus. Logist. 29, 233–257. https://doi.org/10.1002/j.2158-1592.2008.tb00094.x.

Gilbert, C.C., 1992. A duration model of automobile ownership. Transp. Res. Part B Methodol. 26, 97–114. https://doi.org/10.1016/0191-2615(92)90001-D.

Giuliano, G., Dargay, J., 2006. Car ownership, travel and land use: a comparison of the US and Great Britain. Transport. Res. Pol. Pract. 40, 106–124. https://doi.org/10.1016/j.tra.2005.03.002.

Greene, D.L., Hu, P.S., 1985. The influence of the price of gasoline on vehicle use in multivehicle households. Transport. Res. Rec.: J. Transport. Res. Board 988, 19–24.
Greene, W.H., 2000. Econometric Analysis. Prentice-Hall, New York, NY, USA.
Greene, W.H., Hensher, D.A., 2003. A latent class model for discrete choice analysis: contrasts with mixed logit. Transp. Res. Part B Methodol. 37, 681–698. https://doi.org/10.1016/S0191-2615(02)00046-2.
Guo, Z., 2013. Home parking convenience, household car usage, and implications to residential parking policies. Transport Pol. 29, 97–106. https://doi.org/10.1016/j.tranpol.2013.04.005.
Hang, D., McFadden, D., Train, K., Wise, K., 2016. Is vehicle depreciation a component of marginal travel cost?: a literature review and empirical analysis. J. Transport Econ. Pol. 50, 132–150.
Heckman, J.J., Singer, B., 1986. Econometric analysis of longitudinal data. Handb. Econom. 3, 1689–1763. https://doi.org/10.1016/S1573-4412(86)03009-X.
Kim, H.S., Kim, E., 2004. Effects of public transit on automobile ownership and use in households of the USA. Review of Urban & Regional Development Studies. J. Appl. Region. Sci. Confer. 16, 245–262. https://doi.org/10.1111/j.1467-940X.2005.00090.x.
Liu, Y., Cirillo, C., 2016. Small area estimation of vehicle ownership and use. Transport. Res. Transport Environ. 47, 136–148. https://doi.org/10.1016/j.trd.2016.05.002.
Mannering, F., Winston, C., 1985. A dynamic empirical analysis of household vehicle ownership and utilization. Rand J. Econ. 16 (2), 215–236. http://www.jstor.org/stable/2555411.
Mannering, F.L., 1983. An econometric analysis of vehicle use in multivehicle households. Transport. Res. Gen. 17, 183–189. https://doi.org/10.1016/0191-2607(83)90040-7.
McCullagh, P., Nelder, J.A., 1983. Generalized Linear Models. Taylor and Francis, Boca Raton, FL, USA.
McFadden, D., Train, K., 2000. Mixed MNL models for discrete response. J. Appl. Econom. 15, 447–470.
McGuckin, N., Fucci, A., 2017. Summary of Travel Trends: 2017 National Household Travel Survey. U.S. Department of Transportation (US DOT), Federal Highway Administration, Washington DC, USA. Retrieved from: https://nhts.ornl.gov/assets/2017_nhts_summary_travel_trends.pdf (No. FHWA-PL-18-019).
National Renewable Energy Laboratory, Transportation Secure Data Center, 2019. Transportation Secure Data Center: 2019 California Vehicle Survey Conducted by California Energy Commission. Retrieved from: https://www.nrel.gov/transportation/secure-transportation-data/.
Nazari, F., Mohammadian, A.K., Stephens, T., 2019. Modeling electric vehicle adoption considering a latent travel pattern construct and charging infrastructure. Transport. Res. Transport Environ. 72, 65–82. https://doi.org/10.1016/j.trd.2019.04.010.
Nazari, F., Mohammadian, A.K., Stephens, T., 2018a. Dynamic household vehicle decision modeling considering plug-In electric vehicles. Transport. Res. Rec.: J. Transport. Res. Board 1–10. https://doi.org/10.1177/0361198118796925.
Nazari, F., Noruzoliaee, M., Mohammadian, A.K., 2018b. Shared versus private mobility: modeling public interest in autonomous vehicles accounting for latent attitudes. Transport. Res. C Emerg. Technol. 97, 456–477. https://doi.org/10.1016/j.trc.2018.11.005.
Noruzoliaee, M., Zou, B., 2022. One-to-many matching and section-based formulation of autonomous ridesharing equilibrium. Transp. Res. Part B Methodol. 155, 72–100. https://doi.org/10.1016/j.trb.2021.11.002.
Rentziou, A., Gkritza, K., Souleyrette, R.R., 2012. VMT, energy consumption, and GHG emissions forecasting for passenger transportation. Transport. Res. Pol. Pract. 46, 487–500. https://doi.org/10.1016/j.tra.2011.11.009.
Salon, D., Boarnet, M.G., Handy, S., Spears, S., Tal, G., 2012. How do local actions affect VMT? A critical review of the empirical evidence. Transport. Res. Transport Environ. 17, 495–508. https://doi.org/10.1016/j.trd.2012.05.006.
Sarwar, M.T., Anastasopoulos, P.C., Golshani, N., Hulme, K.F., 2017. Grouped random parameters bivariate probit analysis of perceived and observed aggressive driving behavior: a driving simulation study. Anal. Method. Acc. Res. 13, 52–64. https://doi.org/10.1016/j.amar.2016.12.001.
Schimek, P., 1996. Household motor vehicle ownership and use: how much does residential density matter? Transportation Research Record. J. Transport. Res. Board 120–125.
Singh, A.C., Astroza, S., Garikapati, V.M., Pendyala, R.M., Bhat, C.R., Mokhtarian, P.L., 2018. Quantifying the relative contribution of factors to household vehicle miles of travel. Transport. Res. Transport Environ. 63, 23–36. https://doi.org/10.1016/j.trd.2018.04.004.
Spissu, E., Pinjari, A.R., Pendyala, R.M., Bhat, C.R., 2009. A copula-based joint multinomial discrete-continuous model of vehicle type choice and miles of travel. Transportation 36, 403–422. https://doi.org/10.1007/s11116-009-9208-x.
Stevens, M.R., 2017. Does compact development make people drive less? J. Am. Plann. Assoc. 83, 7–18. https://doi.org/10.1080/01944363.2016.1240044.
U.S. Federal Highway Administration (US FHWA), 2022. Moving 12-month Total Vehicle Miles Traveled [M12MTVUSM227NFWA]. Federal Reserve Bank of, St. Louis. Retrieved from. https://fred.stlouisfed.org/series/M12MTVUSM227NFWA.
Waseem, M., Ahmed, A., Saeed, T.U., 2019. Factors affecting motorcyclists' injury severities: an empirical assessment using random parameters logit model with heterogeneity in means and variances. Accid. Anal. Prev. 123, 12–19. https://doi.org/10.1016/j.aap.2018.10.022.
Washington, S., Karlaftis, M., Mannering, F., Anastasopoulos, P., 2020. Statistical and Econometric Methods for Transportation Data Analysis, third ed. Chapman and Hall/CRC, New York, NY, USA.
Wedel, M., DeSarbo, W.S., 1994. A review of recent developments in latent class regression models. Adv. Method. Market. Res. 352–388.
Wedel, M., Kistemaker, C., 1989. Consumer benefit segmentation using clusterwise linear regression. Int. J. Res. Market. 6, 45–59. https://doi.org/10.1016/0167-8116(89)90046-3.
Yamaguchi, K., 2000. Multinomial logit latent-class regression models: an analysis of the predictors of gender-role attitudes among Japanese women. Am. J. Sociol. 105, 1702–1740.
Zhang, W., Lu, D., Chen, Y., Liu, C., 2021. Land use densification revisited: nonlinear mediation relationships with car ownership and use. Transport. Res. Transport Environ. 98, 102985 https://doi.org/10.1016/j.trd.2021.102985.
Zhang, Z., Jin, W., Jiang, H., Xie, Q., Shen, W., Han, W., 2017. Modeling heterogeneous vehicle ownership in China: a case study based on the Chinese national survey. Transport Pol. 54, 11–20. https://doi.org/10.1016/j.tranpol.2016.10.005.