Nonlinear effects of changes in built environment and life events on mode choice: A longitudinal analysis

Senkai Xie[1], Feixiong Liao[2]

[1]Urban planning and Transportation group, Eindhoven University of Technology, Eindhoven, the Netherlands, s.xie@tue.nl
[2]Urban planning and Transportation group, Eindhoven University of Technology, Eindhoven, the Netherlands, f.liao@tue.nl

Keywords: changes in built environment; life events; evolving mode choice; nonlinear effect; light gradient boosting machine (LightGBM)

Extended abstract

1. Introduction

The rapid increase in the adoption and use of private cars over the past decades has brought about unprecedented convenience in mobility in our societies. The increased reliance on private cars has also led to a myriad of challenges, including congestion, air pollution, and energy consumption. To address the associated transport externalities, many countries have applied transport policies to control the use of carbon-fueled cars and promote green transport modes. Scholars in the transport field have scrutinized the relationships between the built environment and travel behavior to derive effective transport policies. It is concluded that residents would reduce car use and walk more if they were closer to destinations and had more or less equivalent mode choice alternatives (Ewing and Cervero, 2010; Stevens, 2017). It is commonly recognized that understanding the effects of changes in the built environment and life events on mode choice is essential for promoting a shift from carbon-fueled cars to sustainable transport modes (Aditjandra et al., 2012; de Haas et al., 2018). However, due to the lack of longitudinal data, most studies have neglected the effects of changes in the built environment and life events.

The literature suggests that life events play an important role in daily mode choice (Janke and Handy, 2019; Wang et al., 2020; Kalter et al., 2021) because life events provide a window of opportunity to reconsider their mobility alternatives. However, only a few studies addressed the impacts of life events on the changes in mode choice (Wang et al., 2020). Furthermore, it remains unclear about the short- and long-term effects of life events on evolving mode choice. Recent studies also show that built environment attributes have nonlinear effects on travel behavior (Ding et al., 2018a; Wagner et al., 2022). Ignoring the nonlinear association may result in a misunderstanding of the effect of built environment characteristics and provide false interpretations for planning and policy decisions. However, as stated above, the majority of existing studies applied a cross-sectional research design that cannot pinpoint the inner relationships accounting for changes in mode choice.

Therefore, this study aims to assess the influences of changes in the built environment and life events on evolving travel patterns with a focus on transport mode choice. To highlight the dynamic and nonlinear effects on modal shift, we apply the light gradient boosting machine (LightGBM) method to the Netherlands Mobility Panel (MPN) data. LightGBM is a machine learning method that fits well with nonlinear relationships and can handle discrete dependent variables. We attempt to answer the following research questions: (1) What are the evolving transport mode choice? (2) Do changes in the built environment and life events cause modal shift and in which way if any? (3) What are the nonlinear effects of socio-demographics, built environment, and life events on evolving mode choice? This study contributes threefold to the literature. First, this study uncovers the evolving mode choice from a longitudinal perspective considering the Netherlands as a study area, which enriches the study of dynamic mode choice in high-density developed countries. Second, the analysis casts light on how changes in built environment attributes and life events affect the evolution of mode choice. Third, it provides a comprehensive assessment of the relative importance of different variables and nonlinear relationships. To that end, the remainder of this extended abstract is structured as follows. Section 2 presents the data and methods used in this study. Section 3 presents the results. Section 4 summarizes the key findings and implications.

2. Data and methods

2.1 Data

The panel data used in this study is extracted from the Netherlands Mobility Panel (MPN), which covers various subjects related to individual and household attributes, life events, and three-day travel diaries in the Netherlands. The MPN data is an annual dataset administered by the Netherlands Institute for Transport Policy Analysis (KiM) since 2013 (Hoogendoorn-Lanser et al., 2015). Approximately 2,500 households participate in the MPN survey each year.

To show the evolving patterns of mode choice, five years of the MPN data (from 2015 to 2019) are used. After data cleaning, the sample contains 1194 representative individuals. The trip frequency of the four transport modes (car, public transport, bicycle, and walking) is measured from the three-day travel diaries and used as input variables for latent class clustering. For ease of comparison, most of the variables use the values of changes in 2019 with reference to 2015 (see Table 1). For socio-demographic and built environment attributes, the variables capture the changes over the five years, which are to a large extent related to life events. For a few selected life events such as changes in residence, employment, education, and family composition, we are particularly interested in the effects of length of changes.

Table 1. Variable definition and statistics.

Variables	Definition	Min.	Max.	Mean	St. Dev.
Socio-demographics					
Gender	1: male 2: female	1.00	2.00	1.54	0.50
Age	Age in 2019 1 = 18-30 years old 2 = 31-60 years old 3 = 61 years old and older	1.00	3.00	2.19	0.65
Education level	Highest education level in 2019 1 = low 2 = medium 3 = high	1.00	5.00	4.24	1.13
Employment status	The employment situation in 2015 and 2019 0: unemployed in both years 1: employed in one year 2: employed in both years	0.00	2.00	1.16	0.92
Working hour *	The changes in levels of working hours in a recent week from 2015 to 2019	-7.00	6.00	0.03	1.82
Household income *	The changes in levels of gross income in the household from 2015 to 2019	-6.00	5.00	0.04	0.76
Marital status	The marital status from 2015 to 2019 -1: divorced 0: single 1: get married 2: married.	-1.00	2.00	1.33	0.96
Household size *	The changes in number of people in the household from 2015 to 2019	-3.00	2.00	-0.01	0.50
Number of young children *	The changes in number of young children in the household from 2015 to 2019	-3.00	2.00	-0.02	0.45
Transport card ownership *	The transport card ownership in 2015 and 2019 0: didn't own in both years 1: owned in one year 2: owned in both years	0.00	2.00	1.34	0.50
Car ownership *	The changes in number of cars by the household from 2015 to 2019	-3.00	2.00	-0.02	0.58
Built environment attributes					
Urban density *	The changes in levels of urban density from 2015 to 2019	-3.00	2.00	-0.10	0.45
Distance to the nearest city center (km) *	The changes in distance from home to the nearest city center from 2015 to 2019	-33.29	26.48	-0.08	2.34
Distance to the nearest train station (km) *	The changes in distance from home to the nearest train station from 2015 to 2019	-16.44	14.08	0.10	1.41

Table 1. (continued)

Variables	Definition	Min.	Max.	Mean	St. Dev.
Distance to the nearest IC junction (km) *	The changes in distance from home to the nearest IC junction from 2015 to 2019	-23.65	24.97	-0.12	2.14
Distance to the nearest bus stop (km) *	The changes in distance from home to the nearest bus stop from 2015 to 2019	-1.25	3.97	0.01	0.22
Distance to the nearest highway (km) *	The changes of distance from home to the nearest highway from 2015 to 2019	-11.76	12.76	-0.03	1.02
Selected life events					
Relocation	0: no moved 1: moved in 2019 2: moved in 2018 3: moved in 2017 4: moved in 2016 5: moved in 2015	0.00	5.00	0.47	1.26
New job	0: no new job 1: started in 2019 2: started in 2018 3: started in 2017 4: started in 2016 5: started in 2015	0.00	5.00	0.86	1.60
Stopping work	0: no stopped work 1: stopped work in 2019 2: stopped work in 2018 3: stopped work in 2017 4: stopped work in 2016 5: stopped work in 2015	0.00	5.00	0.44	1.22
New education program	0: no new education program 1: started in 2019 2: started in 2018 3: started in 2017 4: started in 2016 5: started in 2015	0.00	5.00	0.35	1.14
Childbirth	0: no childbirth 1: childbirth in 2019 2: childbirth in 2018 3: childbirth in 2017 4: childbirth in 2016 5: childbirth in 2015	0.00	5.00	0.32	1.12

* Value of changes from 2015 to 2019

2.2 Methods

To understand the dynamic effects of socio-demographics and changes in built environment and life events, the conceptual framework for studying the evolving mode choice among four common alternatives is depicted in Fig. 1. We apply a machine learning approach to develop a gradient-boosted decision trees (GBDT) algorithm for multiple choice applications under the LightGBM framework to explore the nonlinear relationships (Ke et al., 2017). GBDT first integrates several individual decision trees internally and then accumulates all the results of these decision trees (Friedman, 2001).

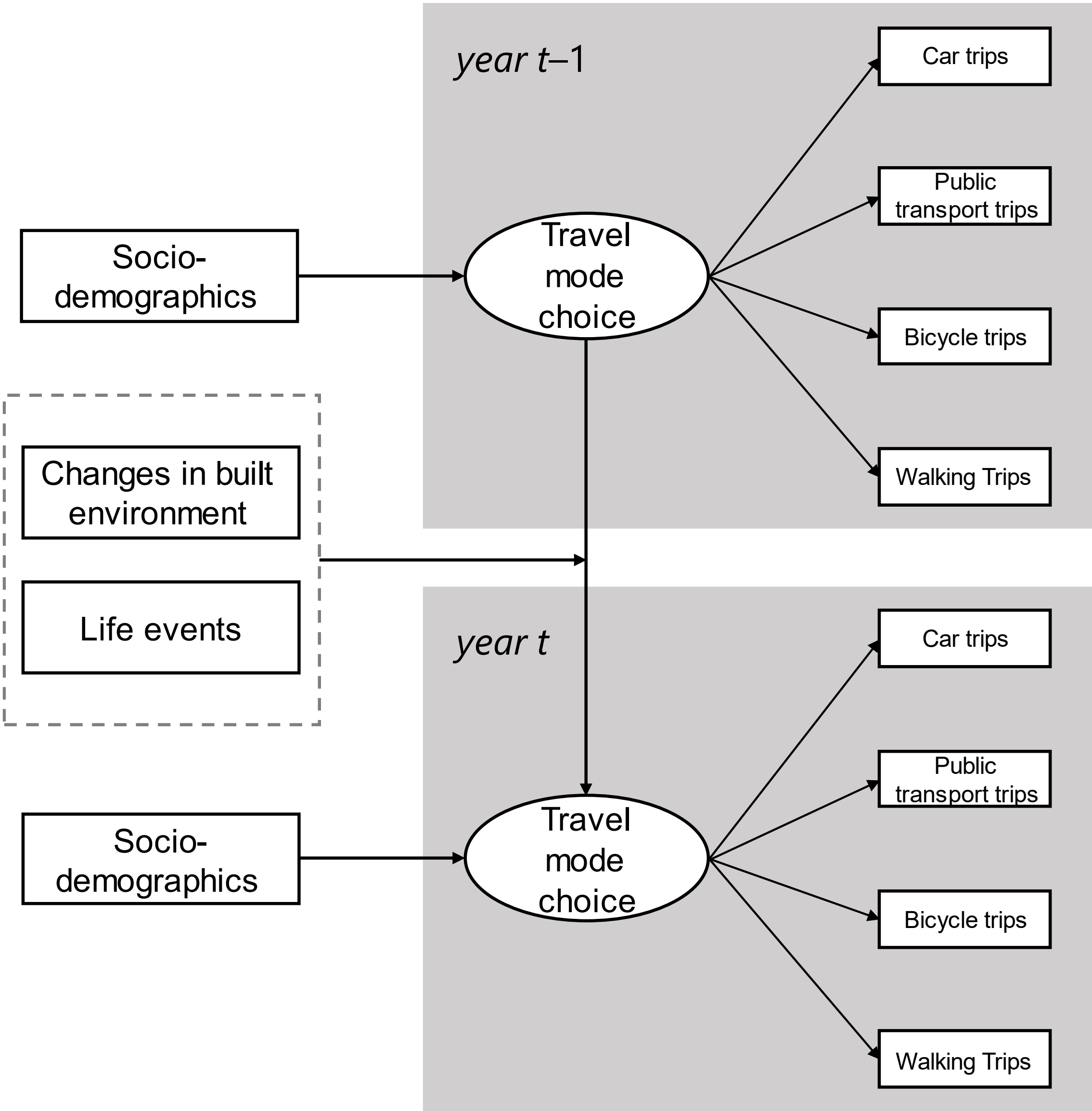


Fig. 1. Conceptual model.

Mathematically, the GBDT concerns dataset $\{x_i, y_i\}_{i=1}^{N}$ with N data points, where x is the set of independent variables (e.g., built environment attributes), and $f(x)$ approximates the dependent variable y (e.g., evolving mode choice) of the prediction function. The function is based on an additive model and formed by combining several decision trees.

$$f(x) = \sum_{m=1}^{M} f_m(x) = \sum_{m=1}^{M} \sigma_m h(x_i; a_m) \quad (1)$$

where M is the number of trees to be fitted, σ_m represents the weight of decision tree m and is used to determine the ability of an individual decision tree to predict the result. a_m represents the mean values of split locations and the terminal node for each splitting variable in the individual decision tree. $h(x_i; a_m)$ is the individual decision tree.

Compared to the traditional regression model, LightGBM can fit both linear and nonlinear relationships. It can handle continuous and discrete variables flexibly and efficiently, and it is also insensitive to multicollinearity. We employ the LightGBM package in Python to develop the LightGBM model and use 5-fold cross-validation to obtain robust model results. The choice of learning rate is 0.01 and the maximum depth of the decision tree is 10. After 133 iterations, the model obtains the optimal results. In this case, the pseudo-R2 of the model is 0.59, which indicates a relatively strong model fit (Wu et al., 2019).

3. Results

3.1 Evolving mode choice

The travel frequency of the four modes (car, public transport, bicycle, and walking) on weekdays from 2015 to 2019 (5 waves) is used to define the mode choice using the method of latent class clustering. Models with different numbers of classes (from 1 to 10) are used to determine the appropriate number of classes. The model identifies four classes by the criteria that the percentage increase in the LL value is relatively small along with a descending BIC value and the minimum class size is over 10% (Magidson and Vermunt, 2004; Haustein and Kroesen, 2022). Table 3 presents the profiles of the four classes, of which car was the dominant chosen mode. Based on the use frequency, clusters 1–4 are labeled as “Car”, “Bicycle”, “Public Transport”, and “Walking”, respectively.

Table 2. Model fit of the latent class models with different classes.

	Classes	LL	BIC	Npar	% increase in LL compared to 1-class model
Model1	1	-35504.4	71043.5	4	13.86%
Model2	2	-30559.4	61197.04	9	22.30%
Model3	3	-28638.6	57398.87	14	25.72%
Model4	4	-28058.2	56281.62	19	26.80%
Model5	5	-27874.4	55957.57	24	27.51%
Model6	6	-27697.5	55647.15	29	28.07%
Model7	7	-27562.9	55421.38	34	28.44%
Model8	8	-27471.3	55281.63	39	28.70%
Model9	9	-27398.7	55179.99	44	28.71%
Model10	10	-27386.3	55198.59	49	13.86%

LL: Log-Likelihood

BIC(LL): Bayesian Information Criterion (based on LL)

Npar: Number of parameters

After clustering the mode choice, the evolving mode choice is derived by the following steps. First, we construct the evolving sequences of mode choice of individuals over over the five years. Second, we classify the sequence as either stable or changing. Mode choice that remained the same for four or more years are considered stable, while residents whose mode choice changed in the last two years compared to the first two years are considered changing. Third, we aggregate bicycle, public transport, and walking into green modes, as including too many modes would make the results too complex for interpretation. Thus, four types of evolving mode choice are identified for analysis after classification as shown in Table 4, of which modal shift did not take place for the majority. Taking *car to green modes* for example, for a person whose mode choice was car in the first three years and green modes in the later two years, the evolving travel pattern is C-C-C-G-G. This sequence is classified as evolving from car to green modes.

Table 3. Trip rates of latent class clustering.

	Cluster 1: Car use (C)	Cluster 2: Bicycle (B)	Cluster 3: Public Transport (PT)	Cluster 4: Walking (W)
Car trip frequency (per day)	2.50	0.63	0.88	1.21
Public transport trip frequency (per day)	0.00	0.01	4.64	0.02
Bicycle trip frequency (per day)	0.17	2.46	0.58	0.50
Walking trip frequency (per day)	0.21	0.36	0.63	2.31
Class size	46.51%	29.09%	11.65%	12.75%

Table 4. Evolving travel pattern classification.

	stable use of car	**stable use of green modes**	**car to green modes**	**green modes to car**
Patterns size	41.54%	40.54%	7.87%	10.05%
Example of sequences	C-C-C-C-C	G-G-G-G-G	C-C-C-G-G	G-G-C-C-C

Note: C: Car; G: Green modes, including bicycle, public transport, and walking.

3.2 Relative importance

Table 4 presents the relative importance and rankings of all variables in predicting the evolving mode choice. Socio-demographic variables collectively contribute 64.64%, while built environment attributes and the lengths of changes in selected life events contribute 16.12% and 19.24%, respectively. Distance to the nearest bus stop has the largest contribution in predicting evolving mode choice of built environment attributes, with a relative importance of 5.95%. Distance to the nearest highway and distance to the nearest train station rank 2nd and 3rd among built environment attributes. Among the selected life events, starting a new job plays an important role in predicting evolving mode choice, with a contribution of 8.99%, ranking 3rd among all the variables.

Table 5. Relative importance of three types of variables.

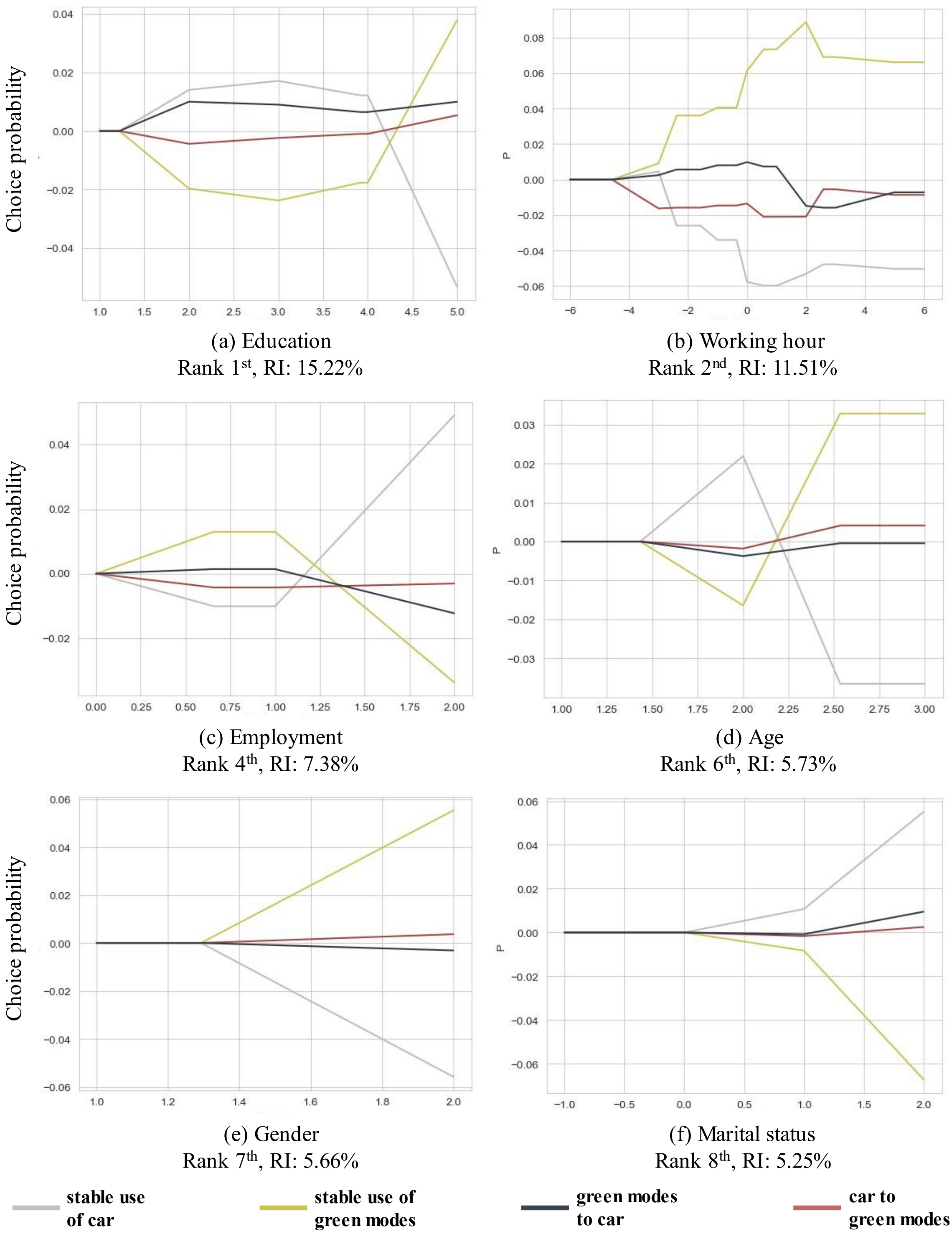
Variables	**Relative importance**	**Rank**
Socio-demographics (64.64%)		
Education	15.22%	1
Working hour	11.51%	2
Employment	7.38%	4
Age	5.73%	6
Gender	5.66%	7
Marital status	5.25%	8
Car ownership	5.18%	9
Transport card ownership	4.56%	10
Income	1.86%	18
Household size	1.76%	19
Number of young children	0.52%	22
Built environment attributes (16.12%)		
Distance to the nearest bus stop (km)	5.95%	5
Distance to the nearest highway (km)	3.30%	11
Distance to the nearest train station (km)	2.75%	13
Distance to the nearest IC junction (km)	2.11%	17
Urban density	1.43%	20
Distance to the nearest city center (km)	0.58%	21
Selected life events (19.24%)		
New job	8.99%	3
Stopping work	2.76%	12
New education program	2.58%	14
Childbirth	2.47%	15
Moving home	2.45%	16

3.3 Nonlinear associations with evolving mode choice

The results of the partial dependence plots (PDPs) show that many variables have nonlinear effects on evolving mode choice. Fig. 2 illustrates the relationships between socio-demographic variables and evolving mode choice. Using education as an example, the x-axis represents the level of education, and the y-axis represents the predicted probability of becoming one of the evolving mode choices. Highly-educated people are more likely to be in stable use of green modes. Changes in working hours are positively associated with the stable use of green modes. Employment is also positively correlated with stable use of car. Results suggest that middle-aged, low or medium-educated men with stable employment are most likely to drive for a long term, which is in line with Ton et al. (2019). Although the transition probability is low in PDPs, the nonlinear effects are significant as assessed in PDPs (Kim, 2021).

Fig. 3 illustrates the nonlinear effects of changes in the built environment on evolving mode choice. Among the built environment variables, the changes in the distance to the nearest bus stop have the highest relative importance in predicting the evolution of mode choice. The threshold of distance to the nearest bus stop is around 450 m. Within the threshold, the transition from car to green modes has a positive association with the changes in distance to the nearest bus stop. As the distance exceeds 450 m, there is no discernible impact on the evolution of mode choice. In general, being close to services increases the probability of adopting green modes. This is consistent with the results of Cao et al. (2019) who found that moving inward would reduce car ownership and use through a longitudinal analysis. However, changes in distance to the nearest bus stop are positively associated with the use of green modes within 450 m. The larger the distance, the higher the likelihood of switching to green modes. A plausible explanation is that the removal of bus stops stimulates multimodality for daily activities, that is, increasing trips executed with green modes.

Fig. 4 shows that life events have a significant impact on the evolution of mode choice. The relationship between starting a new job and mode choice is found to be nonlinear, exhibiting a U-shaped pattern. Within the first two years of getting a new job, people are more likely to use green modes. However, a higher inclination towards car use is observed after the second year. This U-shaped relationship illustrates how the short-term effects of starting a new job are associated with green travel while promoting car use in the long run. Similarly, stopping work and mode choice have a U-shape relationship. The difference is that after two years of stopping work, people are more likely to use green modes. Thresholds for a new education program appear in the 4th year as seen in Fig. 4(c). Furthermore, there is a positive correlation between childbirth and the modal shift from green modes to car, indicating that parents are more willing to drive after childbirth. On the other hand, childbirth also shows a positive relationship with the switch from car to green modes. In either way, childbirth means an opportunity window for modal shifts (de Haas et al., 2018).



(a) Education
Rank 1st, RI: 15.22%

(b) Working hour
Rank 2nd, RI: 11.51%

(c) Employment
Rank 4th, RI: 7.38%

(d) Age
Rank 6th, RI: 5.73%

(e) Gender
Rank 7th, RI: 5.66%

(f) Marital status
Rank 8th, RI: 5.25%

Fig. 2. Nonlinear relationships between socio-demographic variables and evolving mode choice.

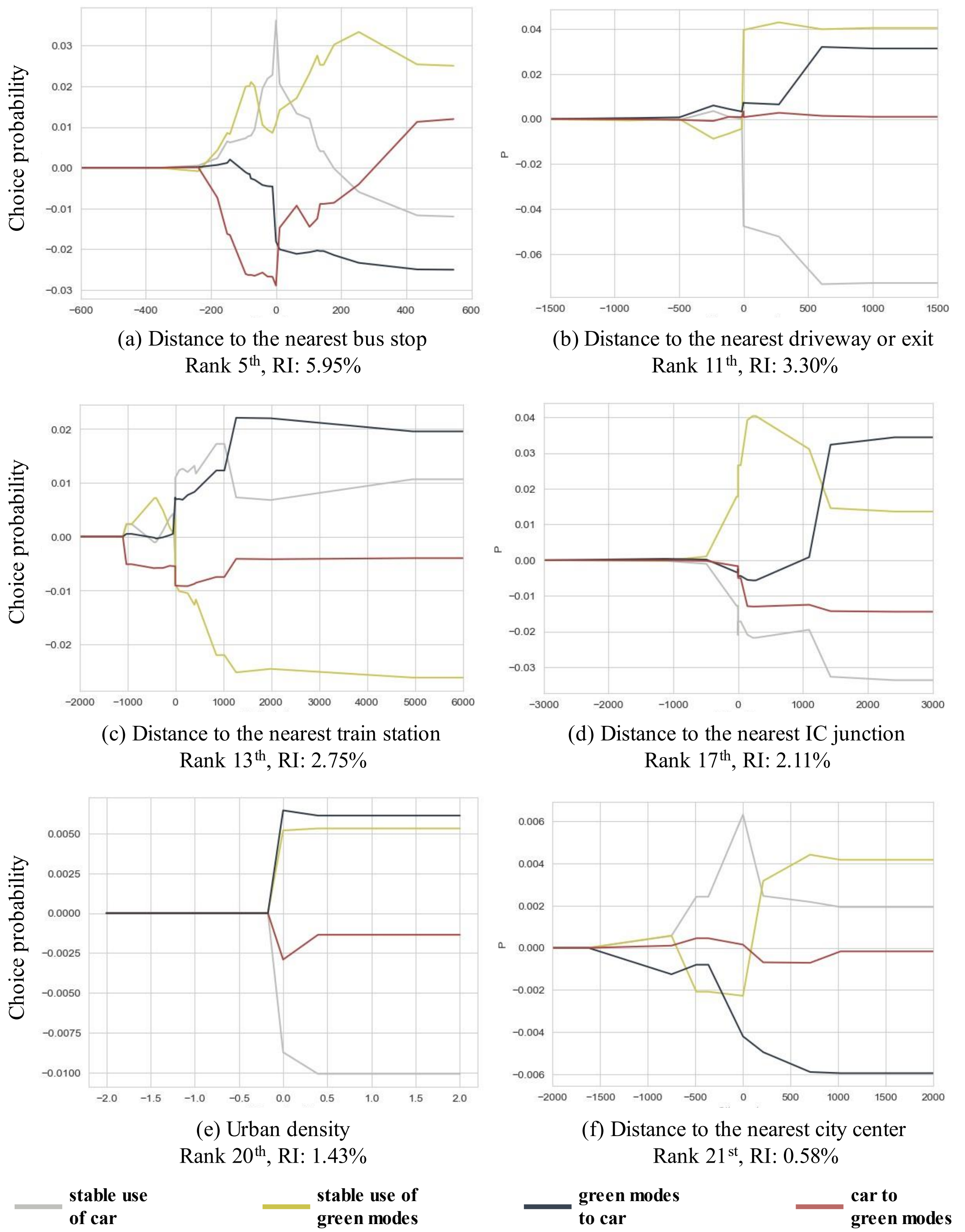


Fig. 3. Nonlinear relationships between built environment attributes and evolving mode choice.

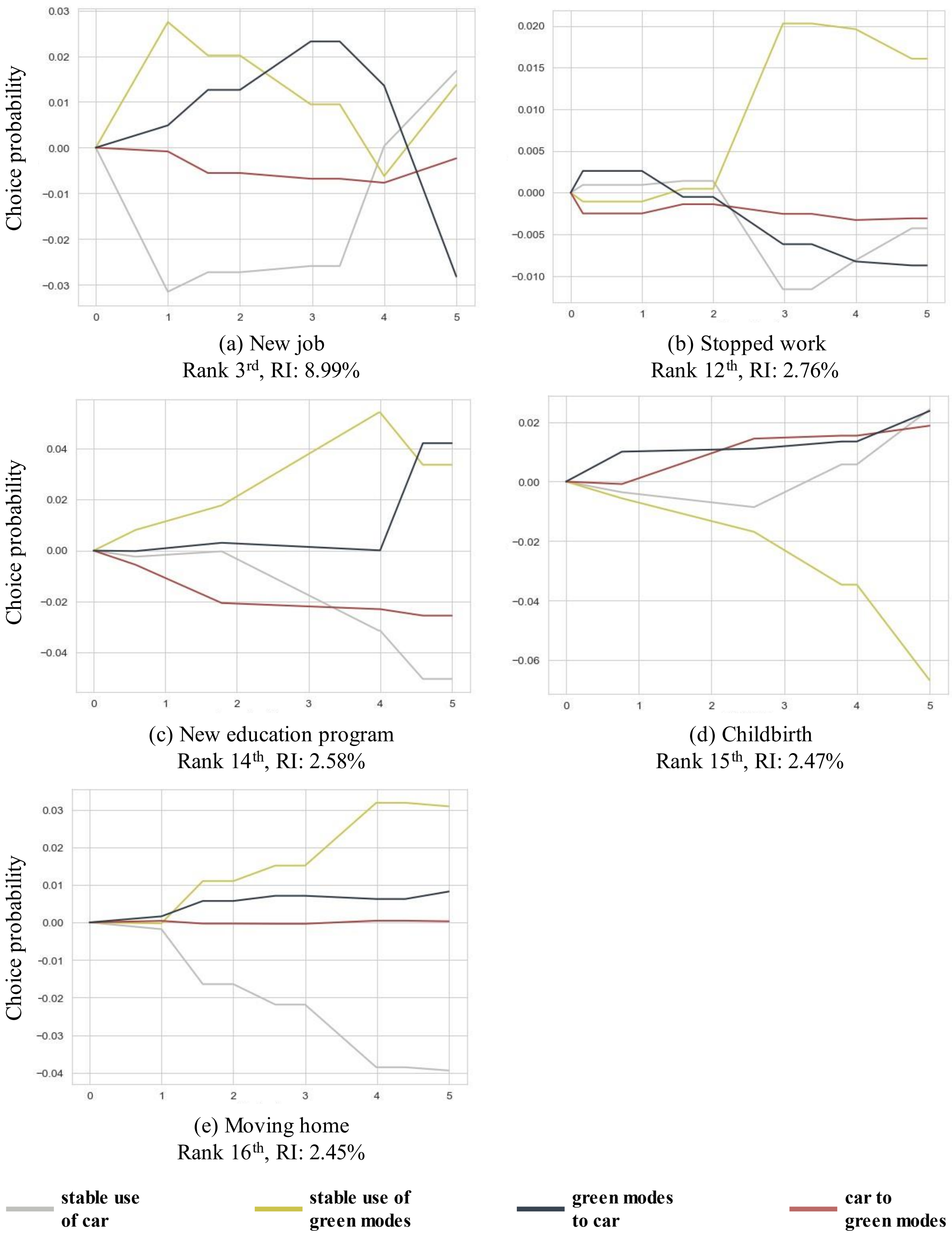


Fig. 4. Nonlinear relationships between life events and evolving mode choice.

4. Discussion

This study applies the LightGBM approach to the mobility panel data from the Netherlands to analyze the relationships between socio-demographics, built environment characteristics, and life events. The results of latent class clustering show that there are four classes of mode choice and four types of evolving mode choice. The study shows that approximately 82% of individuals have a stable mode choice over the five years. The changes in built environment attributes have great predictive power and exhibit nonlinear relationships with the evolving mode choice. Individuals moving closer to service facilities are correlated with green mobility. Furthermore, life events have both short- and long-term effects on mode choice with a U-shape impact, implying that short-term effects are divergent from long-term effects.

The results indicate that the increased distance to transport facilities would discourage car use. Utilizing the derived thresholds, policymakers and planners can deploy the built environment attributes to facilitate the use of green modes efficiently. Furthermore, our study uncovers that life events, especially starting a new job and stopping work have important impacts on evolving mode choice. Policymakers who want to use life events as an opportunity to influence travel mode choices must have a clear understanding of the underlying short- and long-term effects. For example, policymakers or operators may promote emerging green modes to substitute carbon-fueled vehicles by offering new parents free experiences of electric cars and car-sharing, thus influencing them to use these low-carbon modes.

Nevertheless, a few aspects deserve further exploration. First, to further investigate the autoregressive and cross-lagged effects, future research should explore the causal relationships by using other machine learning and econometric methods. Second, more built environment variables are needed to form the '5Ds' measurement framework of the built environment (Ewing and Cervero, 2010). Third, since the Netherlands has well-developed public transport and cycling infrastructure, the findings may not apply to other countries, and therefore empirical studies in other spatial contexts are needed to validate and complement our study.

References

Aditjandra, P. T., Cao, X. J., & Mulley, C. (2012). Understanding neighbourhood design impact on travel behaviour: An application of structural equations model to a British metropolitan data. *Transportation research part A: policy and practice*, 46(1), 22-32.

Cao, X. J., Næss, P., & Wolday, F. (2019). Examining the effects of the built environment on auto ownership in two Norwegian urban regions. *Transportation Research Part D: Transport and Environment*, 67, 464-474.

De Haas, M. C., Scheepers, C. E., Harms, L. W. J., & Kroesen, M. (2018). Travel pattern transitions: applying latent transition analysis within the mobility biographies framework. *Transportation Research Part A: Policy and Practice*, 107, 140-151.

Ding, C., Cao, X. J., & Næss, P. (2018). Applying gradient boosting decision trees to examine non-linear effects of the built environment on driving distance in Oslo. *Transportation Research Part A: Policy and Practice*, 110, 107-117.

Ewing, R., & Cervero, R. (2010). Travel and the built environment: A meta-analysis. *Journal of the American planning association*, 76(3), 265-294.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, 1189-1232.

Haustein, S., & Kroesen, M. (2022). Shifting to more sustainable mobility styles: A latent transition approach. *Journal of Transport Geography*, 103, 103394.

Hoogendoorn-Lanser, S., Schaap, N. T., & OldeKalter, M. J. (2015). The Netherlands Mobility Panel: An innovative design approach for web-based longitudinal travel data collection. *Transportation Research Procedia*, 11, 311-329.

Janke, J., & Handy, S. (2019). How life course events trigger changes in bicycling attitudes and behavior: Insights into causality. *Travel behaviour and society*, 16, 31-41.

Kalter, M. J. O., Puello, L. L. P., & Geurs, K. T. (2021). Exploring the relationship between life events, mode preferences and mode use of young adults: A 3-year cross-lagged panel analysis in the Netherlands. *Travel behaviour and society*, 24, 195-204.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... & Liu, T. Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, 30.

Kim, E. J. (2021). Analysis of travel mode choice in Seoul using an interpretable machine learning approach. *Journal of Advanced Transportation*, 2021, 1-13.

Magidson, J., & Vermunt, J. K. (2004). Latent class models. *The Sage handbook of quantitative methodology for the social sciences*, 175-198.

Stevens, M. R. (2017). Does compact development make people drive less?. *Journal of the American Planning Association*, 83(1), 7-18.

Ton, D., Duives, D. C., Cats, O., Hoogendoorn-Lanser, S., & Hoogendoorn, S. P. (2019). Cycling or walking? Determinants of mode choice in the Netherlands. *Transportation research part A: policy and practice*, 123, 7-23.

Wagner, F., Milojevic-Dupont, N., Franken, L., Zekar, A., Thies, B., Koch, N., & Creutzig, F. (2022). Using explainable machine learning to understand how urban form shapes sustainable mobility. *Transportation Research Part D: Transport and Environment*, 111, 103442.

Wang, X., Shao, C., Yin, C., & Guan, L. (2020). Built environment, life events and commuting mode shift: Focus on gender differences. *Transportation research part D: transport and environment*, 88, 102598.

Wu, X., Tao, T., Cao, J., Fan, Y., & Ramaswami, A. (2019). Examining threshold effects of built environment elements on travel-related carbon-dioxide emissions. *Transportation Research Part D: Transport and Environment*, 75, 1-12.