9-2018

Transferring time-series discrete choice to link-based route choice in space: Estimating vehicle type preference using recursive logit model

Fabian BASTIN
University of Montreal

Yan LIU
University of Maryland at College Park

Cinzia CIRILLO
University of Maryland at College Park

Tien MAI
Singapore Management University, atmai@smu.edu.sg

Article

TRR
JOURNAL OF THE TRANSPORTATION RESEARCH BOARD

Transportation Research Record
2018, Vol. 2672(49) 81–90

DOI: 10.1177/0361198118796731
journals.sagepub.com/home/trr
SAGE

Transferring Time-Series Discrete Choice to Link-Based Route Choice in Space: Estimating Vehicle Type Preference using Recursive Logit Model

Fabian Bastin[1], Yan Liu[2], Cinzia Cirillo[2], and Tien Mai[1]

Abstract
This paper considers a sequential discrete choice problem in a time domain, formulated and solved as a route choice problem in a space domain. Starting from a dynamic specification of time-series discrete choices, we show how it is transferrable to link-based route choices that can be formulated by a finite path choice multinomial logit model. This study establishes that modeling sequential choices over time and in space are equivalent as long as the utility of the choice sequence is additive over the decision steps, the link-specific attributes are deterministic, and the decision process is Markovian. We employ the recursive logit model proposed in the context of route choice in a network, and apply it to estimate time-series vehicle type choice based on Maryland Vehicle Stated Preference Survey data. The model has been efficiently estimated by a dynamic programming approach; the values of estimated coefficients provide important patterns on vehicle type preferences. Compared with a naive model based on sequential multinomial logit choices which are independent over time and a dynamic discrete choice model which considers agent's future expectations, the smaller root mean square error of recursive logit model indicates that it has a better performance in estimating sequential choices over time. We also compare the predictive powers and find that the proposed model outperforms the basic approach and the dynamic approach.

Discrete choice theory, introduced by McFadden (*1*), has proven to be an efficient framework to predict and explain individual choices. When individuals express a sequence of choices instead of a single one, the derivation and estimation of behavioral models become challenging. Despite an individual's sequence of choices being possible in a space domain or a time domain, in the literature of discrete choice analysis, most behavioral models capture time-series choices instead of link-based choices in space. Modeling the two types of decisions in sequence is in fact transferable as long as (a) the alternative-specific (link-specific) attributes are deterministic, (b) the perceived total utility is additive over the sequence of choices, and (c) the decision process is Markovian (*2*).

To estimate individuals' time-series choices, dynamic discrete choice models of consumer behaviors were first used in economics and social science (*3*). The earliest generation of studies includes Wolpin (*4*) on fertility and child mortality, Miller (*5*) on job matching and occupational choice, Pakes (*6*) on patent renewal, and Rust (*7*) on engine replacement. Rust (*7*) was the first to propose a dynamic logit model and to formulate the time-series discrete choices of bus engine replacement as an optimal stopping problem. His model is conceived for a single agent, a homogeneous product, and infinite time horizon. The utilities of the dynamic model are nonlinear, composed of information on current alternatives, expectations about future alternatives, and independent and identically Gumbel-distributed random components. Melnikov (*8*) extended Rust's formulation to consider binary decisions, whether to buy or postpone the purchase of a printer, based on the expected evolution of printer quality and price. His dynamic model considers heterogeneous products and homogeneous consumers, and assumes a consumer can only make one purchase over the time horizon. Lorincz (*9*) further extended the dynamic structure by considering the so-called persistent

[1]Department of Computer Science and Operations Research, University of Montreal and CIRRELT, Montréal, QC, Canada
[2]Department of Civil and Environmental Engineering, University of Maryland, College Park, MD

Corresponding Author:
Address correspondence to Cinzia Cirillo: ccirillo@umd.edu

effect, which allows consumers who already had a product to upgrade it instead of replacing it.

Dynamic discrete choice models for car ownership estimation have been developed in recent decades, including models accounting for inertia effect with lagged variables (*10*, *11*) and models capturing forward-looking behaviors with Bellman equations (*12–14*). Ben-Akiva and Abou-Zeid (*10*) proposed a dynamic formulation with the integration of Hidden Markov Chain to model sequence of discrete choices and evolution of latent variables, which was applied for driving behavior analysis. Realizing the importance of a consumer's future expectation in dynamic problems, Schiraldi (*12*) first introduced a dynamic structural approach with optimal stopping formulation to study car replacement decisions in a second-hand vehicle market. His model accounts for consumer's heterogeneity, future expectation, and price endogeneity in an infinite time horizon. However, it ignores the market evolution. To overcome the limitation, Cirillo et al. (*13*) further proposed a regenerative optimal stopping formulation with a stochastic diffusion process in order to capture not only the optimal car purchase time but also consumer's decisions on vehicle type in a dynamically changing vehicle market in a finite horizon. Taking advantage of approximate dynamic programming, the model has been estimated by the maximum likelihood technique. In the literature, a significant number of interesting applications aim at solving the empirical issues in economics, such as unobserved heterogeneity, initial conditions, state dependency, measurement error, endogeneity, and identification (*15*). However, the computational complexity of model estimation is considered to impede the development of these dynamic structures.

On the other hand, the idea of using a sequential link choice model to describe path choice becomes popular in the context of traffic assignment (*16–18*). More recently, Fosgerau et al. (*2*) developed the recursive logit (RL) model and applied it to the route choice problem by formulating each path choice as a sequence of link choices. At each node, individuals are assumed to choose the outgoing link with the maximum utility including an instantaneous cost and an expected downstream utility (i.e., value function) identified by the Bellman equation. This study builds a bridge between the sequential link-based route choice model and the finite path choice multinomial logit (MNL) model; it provides an interpretation of the route choice problem in a dynamic discrete choice formulation.

Based on the RL model, this paper proposes the transferability between the dynamic discrete choice problem and the route choice problem with an application on vehicle type preferences. To the best of our knowledge, this study is the first to establish that sequential choices in a time domain can be formulated as a route choice problem and solved as an individual-specific shortest path problem. Furthermore, the model estimation benefits from the computational advantages of the RL model. In particular, the value function can be quickly computed by solving systems of linear equations, and the overall path likelihood function has a closed form and can be efficiently maximized by a nonlinear optimization algorithm.

The remainder of this paper is organized as follows. First, the methodology is described and the dynamic discrete choice problem formulated as a route choice problem. Then, Maryland Vehicle Stated Preference Survey Data is introduced for model estimation; while the next section demonstrates an application on vehicle type choice using the RL model and performs an out-of-sample validation. The final section offers concluding remarks and avenues for future research.

Methodology

We consider the situation where an individual i, within a set $\mathrm{I} = \{1, ..., I\}$, has to make choices at time stages $t = 1, ..., T_i$, where T_i is the time horizon for the individual i, within finite choice sets A_t that can vary over time. At each stage, each alternative j presents some utility U_{ij} for individual i, and the total utility gained by an individual is supposed to be additive over the sequence of choices:

$$U_i = \sum_{t=1}^{T_i-1} U_{ij_t}, \tag{1}$$

where j_t is the alternative chosen at stage t. We assume that the individual aims to maximize his/her total utility, or in other words, to select the alternatives j_t such that U_i is maximized. Moreover, at each stage t, we assume that the individual situation can be represented by a state x_{it} and that the set of possible states X_{it} at each stage is finite. The state evolution over time is governed by a transition function, such that:

$$x_{i,t+1} = f(x_{it}, j_t) \tag{2}$$

The perceived utility is a function of the state, and therefore, the choice sequence is a solution to the problem:

$$\max_{j_t, t=1,...,T_i-1} \sum_{t=1}^{T-1} U_{ij_t}(x_{it}) \tag{3}$$

or equivalently:

$$\min_{j_t, t=1,...,T_i-1} -\sum_{t=1}^{T-1} U_{ij_t}(x_{it}) \tag{4}$$

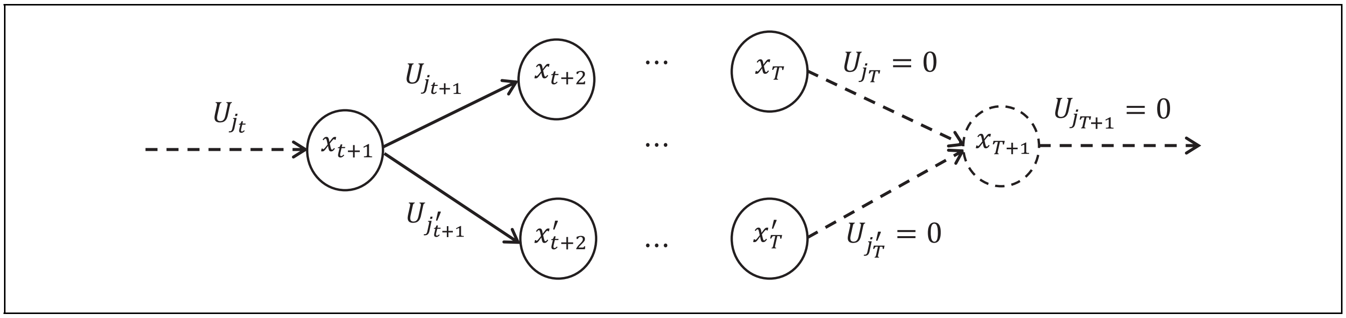

Figure 1. An example network with arbitrary root and artificial destination.

We can represent the problem with a graph where each node represents a possible state and each link corresponds to the transition from a state x_{it} to another state $x_{i,t+1}$ after the choice of some alternative j_t and has the weight $U_{ij_t}(x_{it})$. The graph can be rooted at $x_{i,t+1}$ and we add an artificial terminal node corresponding to an arbitrary state $x_{i,T+1}$ and artificial links connecting each possible state $x_{i,T}$ to $x_{i,T+1}$ with weight $U_{ij_T}(x_{iT}) := 0$. The network in Figure 1 shows a detailed illustration for the graph, with the dash line representing artificial node and links (individual index i is omitted in Figure 1 for simplification purpose).

The choice problem can be treated as a deterministic dynamic program, or equivalently as the computation of the shortest path from $x_{i,1}$ to $x_{i,T+1}$ where in addition the nodes can be numbered in topological order (*19*).

Unfortunately, the alternative utilities cannot be observed exactly, and as in classical discrete choice theory, we assume that the utility $U_{ij_t}(x_{it})$ can be decomposed as the sum of an observable utility $V_{ij_t}(x_{it})$ and a random term $\epsilon_{ij_t}(x_{it})$. The problem can then be viewed as a route choice problem with known origin and destination, and we will assume that the random terms are independent and identically distributed (i.i.d.) and Gumbel distributed. The sequence of choices can be viewed as the choice of a path in an oriented graph, where the origin is the individual state at the beginning of the choice process, and the destination is the artificial terminal node. In this way, a dynamic choice problem can be transferred to a route choice problem.

The estimation of route choices is usually viewed as a difficult problem, as the set of possible paths exponentially increases with the size of the network, so that this equivalence could seem useless in practice. However, recent developments in route choice theory have made this problem tractable. Especially, we can use the estimation techniques proposed by Fosgereau et al. (*2*) to estimate the model, e.g., the expected downstream utilities as well as their gradients can be computed by solving systems of linear equations, and the choice probability for each path is computed by the choice probabilities of the links on that path. We note that such model can be consistently estimated and quickly used for prediction without sampling of choice sets. We here rely on the MATLAB implementation from Mai (*20*) to perform the RL model estimation on time-series vehicle type choices.

Description of Maryland Vehicle Stated Preference Survey

The data used for the empirical analysis was collected from a stated preference survey, which was designed to analyze household vehicle preferences in a dynamic environment (we refer the reader to reference *21* for details in survey design). The survey was conducted under a self-interview, web-based format. Table 1 describes the survey methodology employed.

The stated choice experiment places respondents in a hypothetical nine-year future period starting from 2014. Each year includes two scenarios with a total of 18 scenarios possible. In each scenario, respondents are shown current prices of gasoline and electricity as well as characteristics of four vehicles—the current vehicle and three new vehicles: a gasoline, a hybrid electric, and a battery electric car. Respondents then choose whether to keep their current vehicle or to purchase a new one. If the respondents keep their current vehicle, they then go to the next scenario with a new set of vehicles. Otherwise, their chosen vehicle becomes their current primary vehicle and the respondents are accelerated three years into the future. After this acceleration, the respondents are returned to the scenario progression with the first scenario for the third year after purchase. Figure 2 shows the progression of vehicle type choice.

The stated preference panel data contains 3598 observations of vehicle type choices from 456 Maryland residents (households) over the hypothetical 18 scenarios. For each observation, the choice set contains four alternatives including residents' current vehicle and three new vehicles. New vehicle characteristics such as purchasing price, fuel economy, size, and recharging range are changing over time in the dynamic marketplace. The data is quite representative for Maryland residents, regarding the distributions of households'/respondents'

Table 1. Survey Design

Characteristic	Details
Time frame	May–June 2014
Target population	Maryland households
Sampling frame	Households with Internet access in the state of Maryland
Sample design	Recruitment panel
Use of interviewer	Self-administered
Mode of administration	Self-administered via Internet
Computer assistance	Computer-assisted, web-based self-interview
Reporting unit	One person aged 18 or older per household reports for the entire household
Time dimension	Cross-sectional survey with hypothetical longitudinal stated-choice experiment
Frequency	One 2-week phase of collecting responses
Levels of observations	Household, vehicle, person

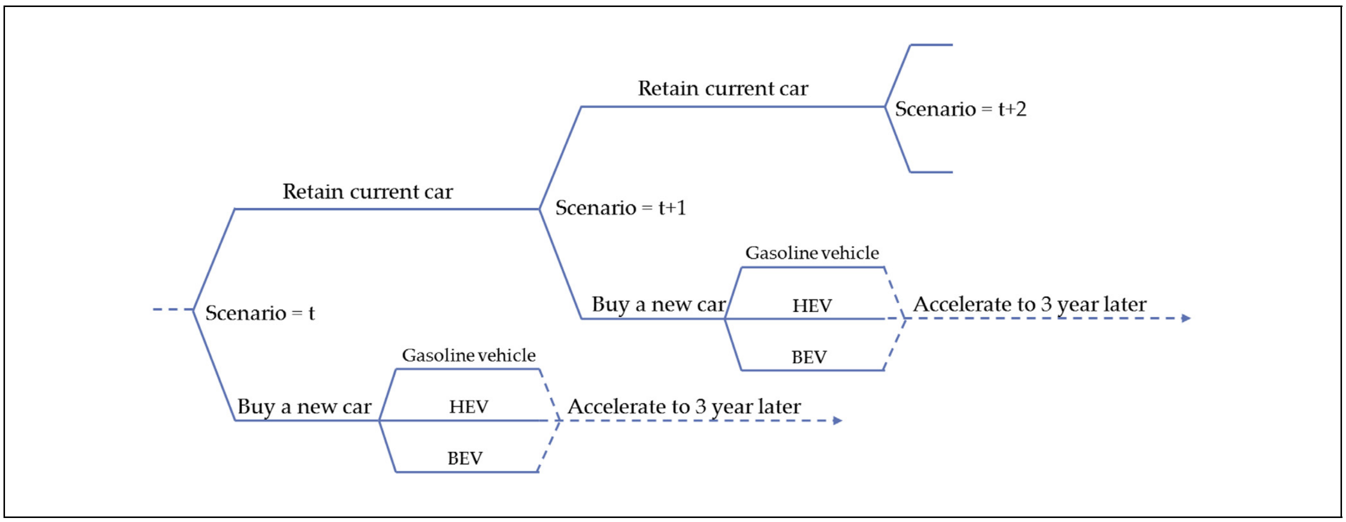

Figure 2. Decision progression of vehicle type choice.

characteristics and their current vehicle characteristics (reference *21* provides summary tables of descriptive data statistics).

Besides, the sequence of households' vehicle purchasing decisions over time provides essential evidence to capture households' preference on switching to greener vehicle types. According to the survey design, at each scenario households can be either in the market or out-of-market. If a household is in the market, he/she has an opportunity to decide either to retain his/her current car or to purchase a new car. Figure 3 shows the shares of households in the market, retaining current vehicle, buying a new gasoline vehicle (GV), a new hybrid electric vehicle (HEV), and a new battery electric vehicle (BEV) over the 18 scenarios.

Application on Vehicle Type Choice

In this section, we construct a link-based network to describe households' vehicle type choices over time and employ an RL model to predict future vehicle preferences on the time-series data in Maryland.

Link-Based Network to Represent Vehicle Type Choice

To transfer this dynamic vehicle type choice problem to a route choice problem, we focus on the sequence of households' vehicle type choices and vehicle characteristics over time. In particular, a link-based network is constructed to represent the decision progression of vehicle type choices over the 18 scenarios, with link attributes representing vehicle characteristics. Therefore, the sequence of choices made by a household over time is represented by a path in the network; with a total of 456 path observations available. The entire network contains 74 nodes and 129 links with 2971 possible paths from origin to destination. Figure 4 provides a partial network of vehicle type choices over four years (eight scenarios). If respondents retain their current vehicle at scenario (*j*-1), there will be four outgoing links from the black node

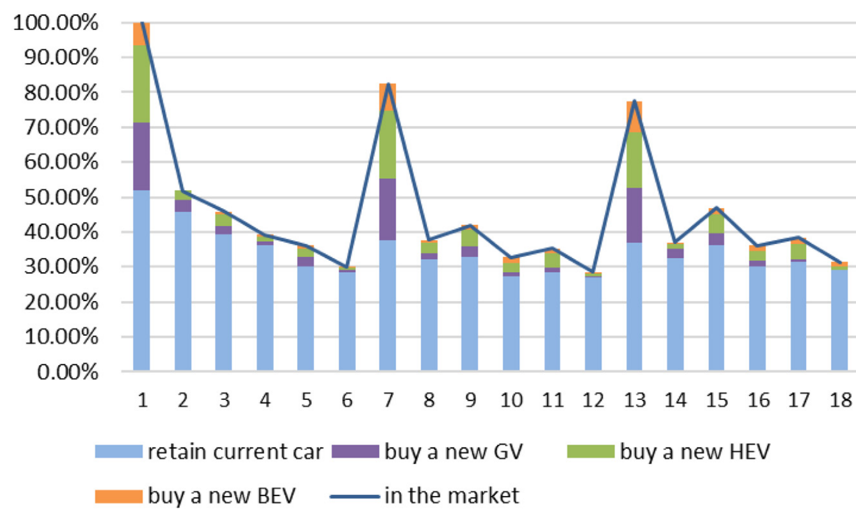

Figure 3. Shares of vehicle type choices for households in the market over time.

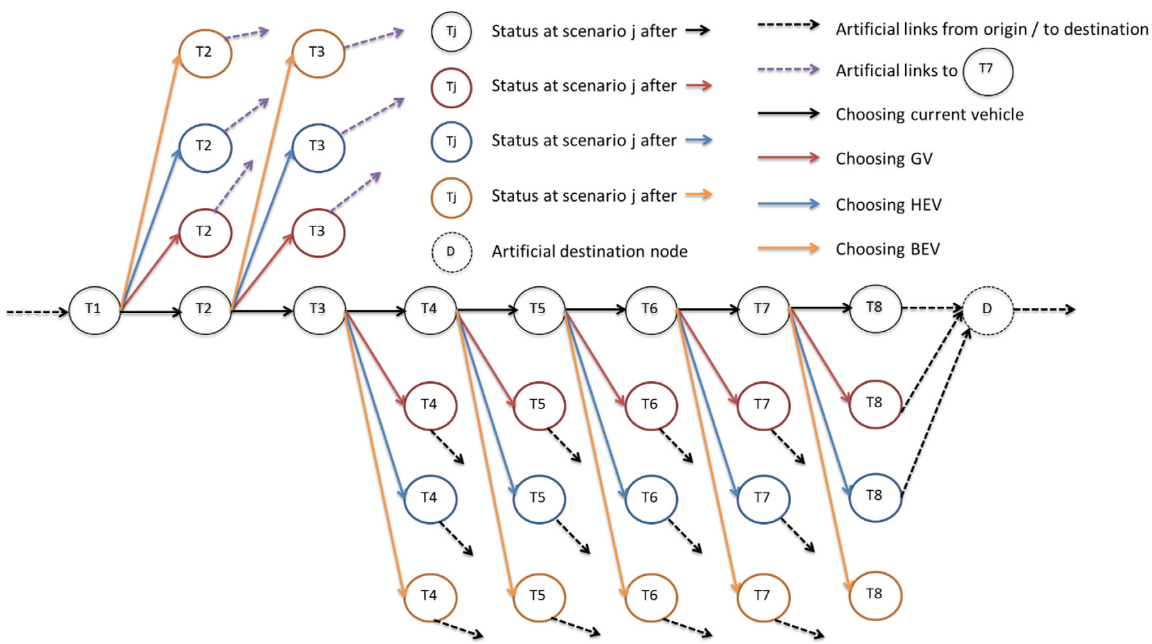

Figure 4. An example network representing vehicle type choices over eight scenarios.

at scenario (*j*-1): the four links represent keeping current vehicle (black arrow), buying a new GV (red arrow), buying a new HEV (blue arrow), and buying a new BEV (yellow arrow) at the next scenario *j*. Otherwise, respondents will jump directly to the first scenario in the third year from purchase (purple dashed arrow). This progression continues until respondents reach the artificial destination node. We assume all paths will eventually go to the unique destination. The main difference between the network considered in this paper with a typical transport network is that our network has only one origin and one destination.

Estimation Results of RL Model

We estimate the RL model on the path observations (sequence of vehicle type choices) of 456 Maryland residents. In the model specification, four outgoing links (alternatives) are considered at each scenario, and the link attributes include household social demographics and car characteristics such as number of workers within a household, respondents' age, education level and gender-related indicators, purchase price, fuel economy (i.e., miles per gallon), vehicle size, and recharging range. The instantaneous utility function (v) for each link is defined as:

$$\begin{aligned} v(g|k) = {} & \mathrm{ASC}_g + \beta_{\mathrm{VP},g}\mathrm{VP}_g + \beta_{\mathrm{FK},g}\mathrm{FK}_g \\ & + \beta_{\mathrm{FU},g}\mathrm{FU}_g + \beta_{\mathrm{VS},g}\mathrm{VS}_g + \beta_{\mathrm{GP}}\mathrm{GP} + \varepsilon_g \end{aligned} \tag{5}$$

$$\begin{aligned} v(h|k) = {} & \mathrm{ASC}_h + \beta_{\mathrm{EF},h}\mathrm{EF}_h + \beta_{\mathrm{YH},h}\mathrm{YH}_h + \beta_{\mathrm{VP},h}\mathrm{VP}_h \\ & + \beta_{\mathrm{FK},h}\mathrm{FK}_h + \beta_{\mathrm{FU},h}\mathrm{FU}_h + \beta_{\mathrm{VS},h}\mathrm{VS}_h + \varepsilon_h \\ v(e|k) = {} & \mathrm{ASC}_e + \beta_{\mathrm{EM},e}\mathrm{EM}_e + \beta_{YH,e}YH_e \\ & + \beta_{\mathrm{VP},e}\mathrm{VP}_e + \beta_{\mathrm{FK},e}\mathrm{FK}_e \\ & + \beta_{\mathrm{FU},e}\mathrm{FU}_e + \beta_{\mathrm{VS},e}\mathrm{VS}_e + \beta_{\mathrm{VR},e}\mathrm{VR}_e \\ & + \beta_{\mathrm{EP}}\mathrm{EP} + \varepsilon_e \\ v(c|k) = {} & \beta_{\mathrm{VP},c}\mathrm{VP}_c + \beta_{\mathrm{NW},c}\mathrm{NW}_c + \varepsilon_c \end{aligned}$$

where

- k is the status retaining current vehicle
- g, g, h, and c are the outgoing links from status k that represent buying a new GV, HEV, BEV, and retaining current vehicle, respectively
- ASC_i is the alternative specific constant of link i, where $i \in \{g, h, e, c\}$
- EF_i, YG_i, EM_i, NW_i represent indicator of educated female (EF), indicator of young household head (YH), indicator of educated male (EM), and number of workers (NW)
- VP_i, FK_i, FU_i, VS_i, and VR_i represent vehicle price (VP), fuel economy for the group who knows the fuel economy of current car (FK), fuel economy for the group who does not know the fuel economy of current car (FU), vehicle size (VS), and vehicle recharging range (VR) of link i, respectively
- GP and EP are gas price (GP) and electricity price (EP)
- All β are alternative-specific coefficients to be estimated
- ϵ_i is the error component of link i following a Gumbel distribution

Table 2 compares the estimation results between an MNL model, an RL model, and a dynamic discrete choice model proposed by Liu and Cirillo (*22*). It should be noted that all attributes considered are observation specific.

In order to approve the capability of estimating vehicle type choice using the RL model, we compare the estimation results of the RL model not only with the static MNL model, but also with the dynamic discrete choice model.

In the specification of MNL model, choices made by the same household are treated as independent observations and panel effects are omitted. Their choices are assumed to be myopic and no forward-looking behavior is considered. We can observe that many estimation coefficients from the MNL model are not significant, such as fuel economy of GV, size of gasoline and hybrid vehicles, recharging range of electric vehicle, and fuel prices.

On the other hand, the RL model and the dynamic model account for the dependences in sequence of choices made by the same household. Further, they assume that households have expectations about future market. There are two main differences between the RL model and the dynamic model. First, the RL model assumes that households have perfect information of future market, while the dynamic model allows households to have limited forward-looking time. Second, the four alternatives in the vehicle type choice problem are treated equally by the RL model, while keeping the current vehicle and making a purchase are treated differently by the dynamic model. We refer readers to reference (*22*) for details in the formulation of the dynamic discrete choice model.

Most estimation coefficients of the RL model and the dynamic model are significant and have reasonable signs except gasoline price and electricity price. The values of "Rho-square" and "Run time" suggest that the RL model has even better performance in estimation than the dynamic model. The capability of estimating vehicle type choices using the RL model indicates that the estimation of dynamic discrete choice in a time domain can be efficiently solved by a route choice model in a space domain.

Some important patterns can be observed from the values of estimated coefficients in the RL model. The negative value of number of workers suggests that if there are more workers in a household, they tend to buy a new car. The indicators of young people are positive, which indicates households with a young household head have higher preference on hybrid and electric cars. The absolute values of vehicle price coefficients for GV (0.364), HEV (0.476), and BEV (0.817) increase, which indicates that households are more sensitive to the sale price of hybrid and electric cars, possibly because these vehicles have newer technology not fully known and they are more expensive. The coefficients for vehicle size of both gasoline-powered and electricity-powered vehicle have positive sign as households prefer larger cars. In addition, households care more about the size of BEV (0.884) than GV (0.459) and HEV (0.408), possibly because of the smaller size of electric cars currently in the market. The

Table 2. Comparison of Estimation Results between MNL, RL, and a Dynamic Model

Attributes [units]	Current	Gasoline	Hybrid	Electric	MNL estimate (standard error)	RL estimate (standard error)	Dynamic model estimate (standard error)
ASC_gas		X			−2.910 (0.47)	−5.838 (0.67)	—
ASC_hyd			X		−2.583 (0.50)	−6.059 (0.61)	2.148 (0.47)
ASC_ele				X	−6.293 (1.14)	−10.547 (1.45)	−3.198 (0.64)
Number of vehicles	X				0.042 (0.05)[a]	—	0.185 (0.01)
Number of workers	X				−0.121 (0.04)	−0.153 (0.02)	−0.027 (−0.01)
D_young_hyd [1/0]			X		0.158 (0.11)[a]	0.550 (0.19)	0.489 (0.12)
D_educated_female_hyd [1/0]			X		0.195 (0.11)[a]	0.245 (0.18)[a]	0.434 (0.11)
D_young_ele [1/0]				X	1.038 (0.17)	1.440 (0.25)	1.651 (0.17)
D_educated_male_ele [1/0]				X	0.410 (0.19)	0.461 (0.26)[a]	0.739 (0.16)
Vehicle_price_gas [$10,000]		X			−0.345 (0.10)	−0.364 (0.11)	−0.492 (0.06)
Vehicle_price_hyd [$10,000]			X		−0.497 (0.10)	−0.476 (0.13)	−0.592 (0.09)
Vehicle_price_ele [$10,000]				X	−0.614 (0.20)	−0.817 (0.24)	−0.794 (0.16)
Vehicle_price_cur [$10,000]	X				−0.097 (0.03)	−0.156 (0.03)	—
mpg_known_gas [100 mpg]		X			1.435 (0.98)[a]	5.115 (1.27)	10.111 (0.71)
mpg_known_hyd [100 mpg]			X		1.676 (0.68)	5.365 (0.83)	8.569 (0.60)
mpg_known_ele [100 mpg]				X	2.523 (0.67)	3.860 (0.80)	4.838 (0.50)
mpg_unknown_gas [100 mpg]		X			1.057 (1.00)[a]	3.715 (1.28)	8.631 (0.66)
mpg_unknown_hyd [100 mpg]			X		0.860 (0.69)[a]	3.715 (0.87)	5.630 (0.56)
mpg_unknown_ele [100 mpg]				X	2.603 (0.67)	3.372 (0.79)	4.060 (0.48)
Vehicle_size_gas [small, medium, large]		X			0.221 (0.13)[a]	0.459 (0.16)	0.905 (0.08)
Vehicle_size_hyd [small, medium, large]			X		0.166 (0.11)[a]	0.408 (0.13)	0.706 (0.09)
Vehicle_size_ele [small, medium, large]				X	0.637 (0.22)	0.884 (0.27)	0.752 (0.16)
Recharging_range [100 miles]				X	0.684 (0.43)[a]	1.225 (0.45)	2.010 (0.36)
Gasoline_price [$1]		X			−0.015 (0.09)[a]	0.005 (0.12)[a, b]	0.547 (0.07)[b]
Electricity_price [$1]				X	−0.154 (0.18)[a]	0.078 (0.25)[a, b]	0.123 (0.058)[b]
Number of observations					3598	456	456
Number of individuals					456	456	456
Null log-likelihood LL(0)					−8201.659	−8201.659	−8201.659
Final log-likelihood LL($\hat{\beta}$)					−3532.370	−2318.937	−2779.839
Rho-square					0.569	0.717	0.661
Adjusted rho-square					0.566	0.702	0.644
Run time					1 sec	6 min 34 sec	7 min 52 sec

Note: — = the corresponding variable is not considered.
[a]The coefficient is not significant at significant level of 0.05.
[b]The sign of the coefficient is not as expected.

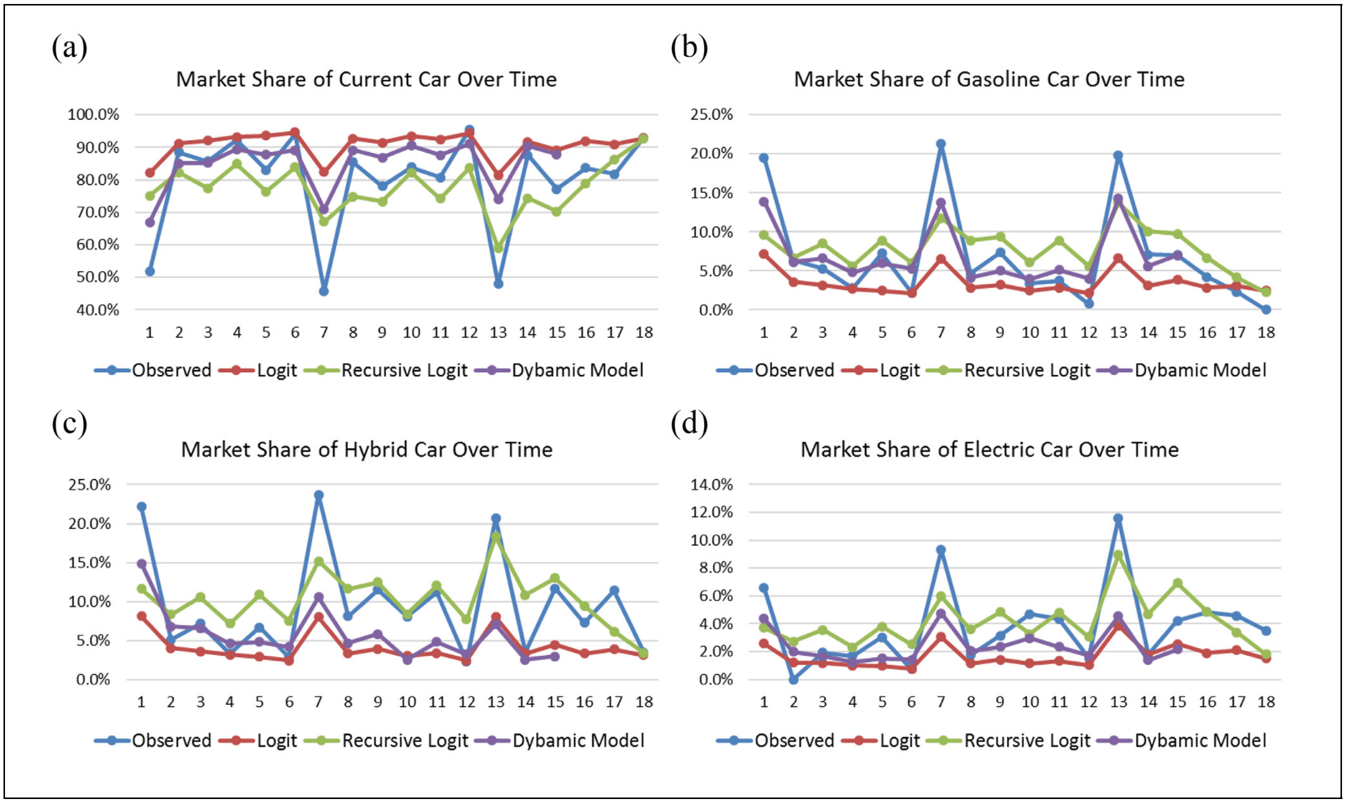

Figure 5. Comparison between observed and predicted market shares on testing dataset over time.

coefficients for recharging range of BEV are positive as expected since a larger range allows for longer trips.

To conclude, the RL model and the dynamic model give very reasonable estimation values which are consistent with general expectations. On the other hand, lower estimated values are obtained from the MNL model, suggesting that the MNL models more conservatively predict how households value fuel economy, vehicle size, and recharging range.

Out-of-Sample Model Validation

In order to validate the prediction performance of the RL model, the sample is randomly divided into two parts: a training dataset containing 80% of the sample, and a testing dataset containing the remaining 20% of the sample. We re-estimate the MNL, the RL, and the dynamic models on the training dataset and apply the estimated coefficients to the testing dataset for predicting the market share of different vehicle types. Figure 5 compares the observed and predicted market shares of the four alternatives along the 18 scenarios over the nine-year period.

From Figure 5, we observe that the MNL model predicts a more stable market share and it is incapable of capturinge fluctuations and peaks. Specifically, it predicts well only the upper bounds of the market share of retaining the current vehicle and the lower bounds of buying a new GV, HEV, and BEV. On the other hand, the RL model is able to recover the fluctuations and sudden changes in consumer demands. It has better performance in predicting the peaks and valleys of market share. The trend of market shares predicted by the dynamic model is in between that of the MNL and RL models. It is able to capture the fluctuations, peaks, and valleys of the actual market shares. However, the model can only predict market shares for 15 time periods, sacrificing the last three time periods to capture household forward-looking behavior.

To further measure the aggregate magnitude of errors in prediction over time, we compare the root mean square errors (RMSE) of market share for the MNL, the RL, and the dynamic models in Figure 6. It should be noted that the reported RMSE are the average values of five validation results. The values suggest that the MNL model has larger prediction errors, especially in reproducing the market share of the current vehicle; while in comparison the other two models perform well. Considering the average over four alternatives, the RL model produces the smallest error.

Concluding Remarks

This paper employed an RL model in the context of route choice in a network to capture sequential vehicle

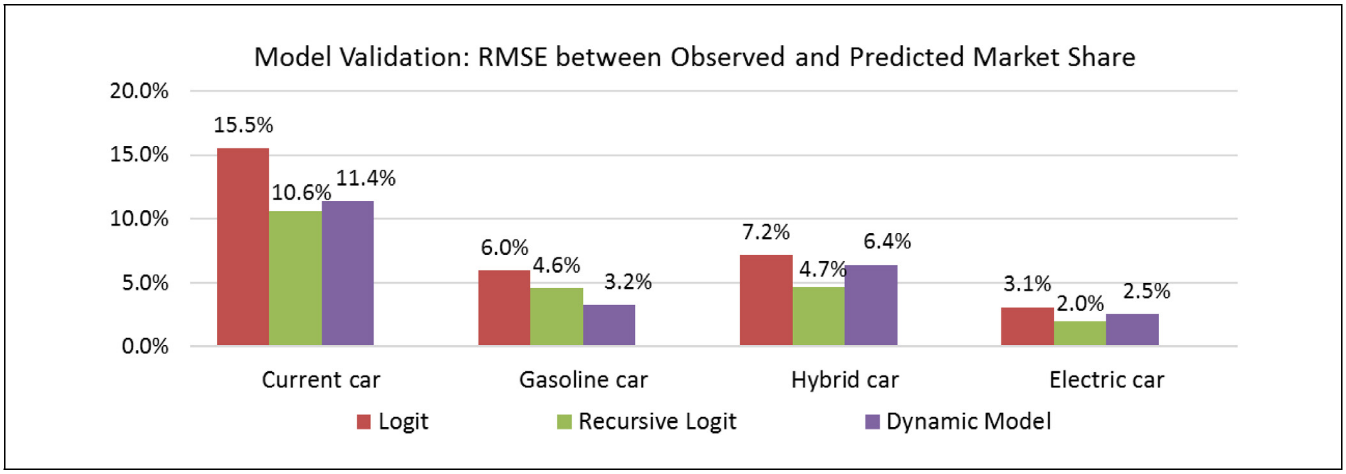

Figure 6. RMSE between observed and predicted vehicle market shares on testing data.

type choices over time and to forecast time-dependent vehicle type preferences. Starting from a dynamic specification of time-series discrete choices, we showed how it is equivalent to link-based route choices that can be formulated by a finite path choice MNL model. This study established that sequential choices over time and in space are transferrable for modeling, as long as the utility of sequential choices is additive, the link-specific attributes are deterministic, and the decision process is Markovian.

Based on the MATLAB implementation from Mai (*20*), we applied the RL model to estimate time-series vehicle type choices using Maryland Vehicle Stated Preference Survey data. The vehicle purchase decision progression of 456 Maryland residents over a nine-year period has been constructed as a link-based network; each node represents a possible state and each link corresponds to the transition between states after a decision is made. The model has been efficiently estimated by a dynamic programming approach; the values of estimated coefficients provide important patterns on vehicle type preferences. Compared with the MNL model and the dynamic model, the smaller RMSE of RL model indicates that it has a better performance in estimating sequential choices over time. The capability of estimating vehicle type choices using the RL model suggests that the estimation of dynamic discrete choice in a time domain can be efficiently solved by a link-based route choice model in a space domain.

There is a large literature on the estimation of time-series discrete choices in economics, social science, and engineering, i.e., patent renewal, welfare gains from industry innovation, machine replacement, car ownership and type choice. However, many dynamic applications of discrete choice models have mainly considered individuals' previous actions (i.e., inertia effect) and do not care much about future expectations. In addition, the computational complexity of model estimation is considered to impede the development of the dynamic structures. The methodology presented in this paper offers a novel and efficient way to model these time-series discrete choices econometrically.

The employed model however retains the well-known independence of irrelevant alternatives (IIA) property which is undesirable in a route choice setting (*23*), but also in the more general setting of sequence of decisions, as two sequences can share common states and decision. To relax the IIA property and to improve the prediction power, Mai, Forsgerau and Frejinger (*24*) extended the RL formulation to a nested RL model that allows path utilities to be correlated in a fashion similar to nested logit (*25*, *26*) and where links can have different scale parameters. Mai, Bastin and Frejinger (*27*) later applied the RL approach to estimate more complex route choice formulations, as mixed logit models, introducing a decomposition technique to alleviate the numerical costs. Such extensions should be investigated in future research. Another undesirable feature of the proposed approach is that individuals make choices with perfect information of the future. While this is necessary to develop the equivalence with a route choice model, this is clearly unrealistic. We could however combine the proposed approach to stochastic programming techniques (*13*) in a context of approximate dynamic programming (*28*) in future research.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. 1131535 and Natural Sciences and Engineering Research Council of Canada (NSERC). The authors would like to acknowledge the reviewers for generously providing their valuable comments and suggestions.

Author Contributions

The authors confirm contribution to the paper as follows: study conception: Fabian Bastin; study design: Cinzia Cirillo; data collection: Cinzia Cirillo and Yan Liu; coding and model: Tien Mai; analysis and interpretation of results: Yan Liu; draft manuscript preparation: Yan Liu; Cinzia Cirillo, Fabian Bastin. All authors reviewed the results and approved the final version of the manuscript.

References

1. McFadden, D. Conditional Logit Analysis of Qualitative Choice Behaviour. In *Frontiers in Econometrics* (Zarembka, P., ed.), Academic Press, New York, NY, 1973, pp. 105–142.
2. Fosgerau, M., E. Frejinger, and A. Karlstrom. A Link Based Network Route Choice Model with Unrestricted Choice Set. *Transportation Research Part B: Methodological*, Vol. 56, 2013, pp. 70–80.
3. Liu, Y., and C. Cirillo. A Generalized Dynamic Discrete Choice Model for Green Vehicle Adoption. *Transportation Research Procedia*, Vol. 23, 2017, pp. 868–886.
4. Wolpin, K. I. An Estimable Dynamic Stochastic Model of Fertility and Child Mortality. *Journal of Political Economy*, Vol. 92, No. 5, 1984, pp. 852–874.
5. Miller, R. A. Job Matching and Occupational Choice. *Journal of Political Economy*, Vol. 92, No. 6, 1984, pp. 1086–1120.
6. Pakes, A. *Patents as Options: Some Estimates of the Value of Holding European Patent Stocks*. NBER Working Paper No. 1340. National Bureau of Economic Research, Cambridge, MA, April 1984. http://www.nber.org/papers/w1340.pdf.
7. Rust, J. Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher. *Econometrica: Journal of the Econometric Society*, Vol. 55, No. 5, 1987, pp. 999–1033.
8. Melnikov, O. Demand for Differentiated Durable Products: The Case of the US Computer Printer Market. *Economic Inquiry*, Vol. 51, No. 2, 2013, pp. 1277–1298.
9. Lorincz, S. *Persistence Effects in a Dynamic Discrete Choice Model-Application to Low-End Computer Servers*. Report No. MT-DP-2005/10. IEHAS Discussion Papers, 2005. http://hdl.handle.net/10419/108082.
10. Ben-Akiva, M., M. Abou-Zeid, and C. Choudhury. Hybrid Choice Models: From Static to Dynamic. In *Triennial Symposium on Transportation Analysis VI*, Phuket, Thailand, March 19, 2007.
11. Nolan, A. A Dynamic Analysis of Household Car Ownership. *Transportation Research Part A: Policy and Practice*, Vol. 44, No. 6, 2010, pp. 446–455.
12. Schiraldi, P. Automobile Replacement: A Dynamic Structural Approach. *The RAND Journal of Economics*, Vol. 42, No. 2, 2011, pp. 266–291.
13. Cirillo, C., R. Xu, and F. Bastin. A Dynamic Formulation for Car Ownership Modeling. *Transportation Science*, Vol. 50, No. 1, 2015, pp. 322–335.
14. Gillingham, K., F. Iskhakov, A. Munk-Nielsen, J. Rust, and B. Schjerning. *A Dynamic Model of Vehicle Ownership, Type Choice, and Usage*. Unpublished Working Paper, 2015. http://bschjerning.com/papers/iruc.pdf.
15. Aguirregabiria, V., and P. Mira. Dynamic Discrete Choice Structural Models: A Survey. *Journal of Econometrics*, Vol. 156, No. 1, 2010, pp. 38–67.
16. Bell, M. G. Alternatives to Dial's Logit Assignment Algorithm. *Transportation Research Part B: Methodological*, Vol. 29, No. 4, 1995, pp. 287–295.
17. Akamatsu, T. Cyclic Flows, Markov Process and Stochastic Traffic Assignment. *Transportation Research Part B: Methodological*, Vol. 30, No. 5, 1996, pp. 369–386.
18. Baillon, J. B., and R. Cominetti. Markovian Traffic Equilibrium. *Mathematical Programming*, Vol. 111, No. 1–2, 2008, pp. 33–56.
19. Bertsekas, D. P. *Dynamic Programming and Optimal Control (3rd ed.)* Athena Scientific, Belmont, MA, 2005.
20. Mai, T. *Dynamic Programming Approaches for Estimating and Applying Large-Scale Discrete Choice Models*. University of Montreal, Montreal, QC, Canada, 2016.
21. Cirillo, C., Y. Liu, and M. Maness. A Time-Dependent Stated Preference Approach to Measuring Vehicle Type Preferences and Market Elasticity of Conventional and Green Vehicles. *Transportation Research Part A: Policy and Practice*, Vol. 100, 2017, pp. 294–310.
22. Liu, Y., and C. Cirillo. A Generalized Dynamic Discrete Choice Model for Green Vehicle Adoption. *Transportation Research Procedia*, Vol. 23, 2017, pp. 868–886.
23. Mai, T., E. Frejinger, and F. Bastin. A Misspecification Test for Logit-Based Route Choice Models. *Economics of Transportation*, Vol. 4, No. 4, 2015, pp. 215–226.
24. Mai, T., M. Fosgerau, and E. Frejinger. A Nested Recursive Logit Model for Route Choice Analysis. *Transportation Research Part B*, Vol. 75, 2015, pp. 100–112.
25. Ben-Akiva, M. *The Structure of Travel Demand Models*. PhD thesis. Massachusetts Institute of Technology, Cambridge, MA, 1973.
26. McFadden, D. Modeling the Choice of Residential Location. *Transportation Research Record: Journal of the Transportation Research Board*, 1978. 673: 72–77.
27. Mai, T., F. Bastin, and E. Frejinger. A Decomposition Method for Estimating Recursive Logit-Based Route Choice Models. *EURO Journal on Transportation and Logistics*, Vol. 7, No. 3, 2018, pp. 253–275.
28. Powell, W. *Approximate Dynamic Programming: Solving the Curses of Dimensionality (2nd ed.)* Wiley, Hoboken, NJ, 2011.

The Standing Committee on Transportation Demand Forecasting (ADB40) peer-reviewed this paper (18-04149).

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation and NSERC.