Dynamic discrete-continuous choice modeling approach for car use, ownership and fuel type choice based on register data

Aurélie Glerum
Transport and Mobility Laboratory,
School of Architecture, Civil and Environmental Engineering,
EPFL - Ecole Polytechnique Fédérale de Lausanne, Switzerland
Email: aurelie.glerum@epfl.ch

Emma Frejinger
Department of Computer Science and Operations Research,
Faculté des arts et des sciences,
Université de Montréal

Anders Karlström
Department of Transport Science,
School of Architecture and the Built Environment,
KTH - Royal Institute of Technology, Stockholm, Sweden

Muriel Beser Hugosson
Department of Transport Science,
School of Architecture and the Built Environment,
KTH - Royal Institute of Technology, Stockholm, Sweden

Michel Bierlaire
Transport and Mobility Laboratory,
School of Architecture, Civil and Environmental Engineering,
EPFL - Ecole Polytechnique Fédérale de Lausanne, Switzerland

1 Introduction

From a modeling perspective, car ownership and car usage models poses a number of challenges. In particular, cars are durable, differentiated goods. Static modeling approaches are typically used (e.g. Berry et al., 1995), but they cannot address the durable feature of a car. Cars entering the fleet tend to stay for a long time, and policies affecting the new purchases therefore have long term impact (e.g. in Sweden cars stay in the fleet for 14 years on average). To address the observed shortcomings of the static approach, structural dynamic models have gained increased attention in the transportation literature. In particular, dynamic discrete choice models (DDCM) are used. For instance, Cirillo and Xu (2011) model the choice of vehicle type and individuals' replacement decisions as a regenerative optimal stopping problem and apply it to a stated preferences context (Xu, 2011). In another novel contribution, Schiraldi (2011) presents a model taking the second-hand market of car into account when analyzing the influence of a scrappage subsidies in Italy on choice of vehicle type and replacement behavior. Interestingly, it allows to capture the unobserved transaction costs due to changes in the household fleet composition.

Car acquisition decisions are often associated with decisions regarding their usage. A static joint modeling approach for both how long individuals keep a car and how much they drive has been applied earlier by De Jong (1996), who used duration models and regressions techniques. More recently, Gillingham (2012) models the choice of vehicle type and monthly mileage using vehicle registration data in California.

In this research, we model households' decisions on car ownership and usage using a *dynamic discrete-continuous choice model (DDCCM).* We account for the future expectations of households regarding the utility of their vehicle(s) in the presence of transaction costs. An important and novel feature of this work consists of using register data of all vehicles and all individuals in Sweden. Moreover, this paper contributes to the specification and estimation of a dynamic model where, each year, a household takes a decision in a complex setting, i.e. the household chooses the number of cars (or not to have a car at all, as the transportation alternatives, such as public transit, also change), the ownership of company cars, the purchase of new or second-hand cars, the choice of annual mileages and fuel types.

2 Background and data

There are fairly few studies using dynamic models for car related choices and one of the reasons may be that data is difficult to obtain. In this study we use register data over the whole Swedish population that combines the population and car registers for the years 1998 to 2008. These registers are based on individuals and we have extensive socio-economic data such as net income, house and work locations, type of employment in addition to characteristics of each owned car (make, model, fuel consumption, fuel type, age, etc.) and the mileage from odometer readings. In addition we have information on all households types except unmarried individuals living together without children. Part of this data (without car characteristics) was used by Pyddoke (2009).

There is variability in important attributes of this extensive register data and we therefore expect to be able to identify the parameters related to policy variables (such as registration taxes or fuel prices) inducing demand shifts. Moreover geographic information is available at a detailed level, which allows to analyze the impact of policy changes at a regional level.

3 A dynamic discrete-continuous choice model formulation

The DDCCM is formulated as a dynamic programming (DP) model and we model the joint decision of vehicle transactions, mileage, fuel type, use of a company car (if available) and purchase of a new or second-hand car.

We make the following assumptions. First, decisions are taken at a household level and we assume that each household can have at most two cars, since only a very small share of the Swedish households has more than two cars. Second, we consider an infinite-horizon problem to account for the fact that households make long-term decisions in the context of car acquisition. Third, the choice of mileage(s) is conditional on the choice of the discrete decision variables (i.e. the transaction type, the type of ownership, the fuel type and the car state). Fourth, the choice of the discrete decision variables is *strategic*, that is, we assume that households take into account the future utility of the choice of these variables in their decision process. Fifth, the choice of mileage(s) is *myopic*, that is, households do not take into account the future utility of the choice of the current annual driving distance(s) in their decision process.

3.1 Definition of the components of the model

The *state space* S is constructed based on the following variables: the age $y_{c,t}$ of car c at year t, a discrete variable $I_{c,t}$ indicating whether car c is owned privately, by sole proprietorship or by another type of company, the fuel type $f_{c,t}$ of car c. Therefore, each state $s_t \in S$ can be represented as $s_t = (y_{1,t}, I_{1,t}, f_{1,t}, y_{2,t}, I_{2,t}, f_{2,t})$.

The *action space* A is constructed based on the following variables: the transaction h_t in the household composition of the car fleet at year t (replacement, increase, decrease, etc.), the annual mileage $\tilde{m}_{c,t}$ for car c, the choice $\tilde{I}_{c,t}$ of taking company car, the fuel type $\tilde{f}_{c,t}$, the choice $\tilde{r}_{c,t}$ to buy a new or second-hand car. Each action $a_t \in A$ can hence be represented as $a_t = (h_t, \tilde{m}_{1,t}, \tilde{I}_{1,t}, \tilde{f}_{1,t}, \tilde{r}_{1,t}, \tilde{m}_{2,t}, \tilde{I}_{2,t}, \tilde{f}_{2,t}, \tilde{r}_{2,t})$.

Based on the above definitions, estimations of the size of the state space and the discrete part of the action space lead to reasonable sizes for the DP problem to be computationally feasible (i.e. in the order of 1400 and 80, respectively).

Given that a household is in a state s_t and has chosen an action a_t, the *transition function* $f(s_{t+1}|s_t, a_t)$ is defined as the rule mapping s_t and a_t to the next state s_{t+1}. In our case, s_{t+1} can be deterministically inferred from s_t and a_t.

3.2 Specification and estimation

At each time period, the household chooses an action $a_t \in A$ in order to maximize its current utility and expected utility for the future years. Assuming that $a_t^D = (h_t, \tilde{I_{1,t}}, \tilde{f_{1,t}}, \tilde{r_{1,t}}, \tilde{I_{2,t}}, \tilde{f_{2,t}}, \tilde{r_{2,t}})$ gathers the discrete components of a_t and $a_t^C = (\tilde{m}_{1,t}, \tilde{m}_{2,t})$ gathers the continuous components, the *instantaneous utility* can be defined as $u(s_t, a_t^C, a_t^D, x_t, \theta) = v(s_t, a_t^C, a_t^D, x_t, \varepsilon_C(a_t^C), \theta) + \varepsilon_D(a_t^D)$, where $v(s_t, a_t^C, a_t^D, x_t, \varepsilon_C(a_t^C), \theta)$ is the deterministic term relative to the choice of the discrete actions and $\varepsilon_C(a_t^C)$ and $\varepsilon_D(a_t^D)$ are error terms relative to the the continuous and discrete actions, respectively.

Variable x_t gathers the explanatory variables of households' choices which are not part of the state or action spaces. We consider three types of explanatory variables. First, it is important to capture transaction costs, which are major elements driving the car replacement process within a household. Second, the registration data of individuals in Sweden allow to consider a large range of socio-economic variables which potentially affect the action a household takes, such as the workplace distances of the heads of household, the total household income or the number of children at home. Third, policy-relevant attributes such as fuel price and vehicle circulation tax should be included.

Under the assumption that the choice of mileage(s) is myopic and conditional on the choice of the discrete actions[1], the integrated value function is obtained by iterating on the following definition of the Bellman equation:

$$\bar{V}(s_t, x_t, \theta) = \log \sum_{a_t^D} \exp\{\max_{a_t^C}\{v(s_t, a_t^C, a_t^D, x_t, \varepsilon_C(a_t^C), \theta)\} + \beta \sum_{s_{t+1} \in S} \bar{V}(s_{t+1}, x_{t+1}, \theta) f(s_{t+1}|s_t, a_t)\} \quad (1)$$

where $\beta \in (0, 1)$ is a discount factor, θ is a set of parameters to estimate.

The DDCCM is solved by maximizing the log of the likelihood function

$$\mathcal{L} = \prod_{n=1}^{N} \prod_{t=1}^{T_n} P(a_{n,t}^D | s_{n,t}, x_{n,t}, \theta), \quad (2)$$

where N is the total population size, T_n is the number of years household n is observed and $P(a_{n,t}^D | s_{n,t}, x_{n,t}, \theta)$ is the probability that household n chooses a particular discrete action $a_{n,t}^D$ at time t. The simplest way to estimate this type of model is using the nested fixed point algorithm proposed by Rust (1987) where the DP problem is solved for each iteration of the non-linear optimization algorithm searching of the parameter space. Due to the complex structure of the model, the computational time might be rather long. We hence plan to consider the alternative approach proposed by Aguirregabiria and Mira (2002) to reduce the number of times the DP problem needs to be solved.

[1]The same assumption was also made by Anders Munk-Nielsen, University of Copenhagen.

4 Expected results and conclusion

Given the durability of goods such as cars, this research highlights the importance of considering forward-looking agents, when car ownership, fuel type and usage are modeled. Using the unique register data of all vehicles and individuals in Sweden, the DDCCM can capture the heterogeneity underlying the individuals' decisions. Moreover the model allows to explain and predict the simultaneous choice of car transactions and usage.

This is work in progress and we are currently finalizing the implementation of the estimator of the DDCCM. In the presentation, we will describe results from an exploratory analysis performed on the Swedish register data as well as an illustrative example of the application of the DDCCM, using chosen parameter values. Some other related results from this ongoing research project have been presented at the Swiss Transport Research Conference 2013 and will be presented at the International Choice Modelling Conference 2013.

References

V. Aguirregabiria and P. Mira. Swapping the nested fixed point algorithm: A class of estimators for discrete markov decision models. *Econometrica*, 70(4):1519–1543, 2002.

S. Berry, J. Levinsohn, and A. Pakes. Automobile prices in market equilibrium. *Econometrica*, 63(4):841–890, 1995.

C. Cirillo and R. Xu. Dynamic discrete choice models for transportation. *Transport Reviews*, 31(4):473–494, 2011.

G. De Jong. A disaggregate model system of vehicle holding duration, type choice and use. *Transportation Research Part B: Methodological*, 30(4):263–276, 1996.

K. Gillingham. Selection on anticipated driving and the consumer response to changing gasoline prices. Technical report, Yale University, 2012.

R. Pyddoke. Empirical analyses of car ownership and car use in Sweden. Technical report, VTI, 2009.

J. Rust. Optimal replacement of gmc bus engines: An empirical model of Harold Zurcher. *Econometrica*, 55(5):999–1033, 1987.

P. Schiraldi. Automobile replacement: a dynamic structural approach. *The RAND Journal of Economics*, 42(2):266–291, 2011.

R. Xu. *Dynamic discrete choice models for car ownership modeling*. PhD thesis, University of Maryland, 2011.