A dynamic discrete-continuous choice model for car ownership and usage

Estimation procedure

Aurélie Glerum

Emma Frejinger

Anders Karlström

Muriel Beser Hugosson

Michel Bierlaire

April 2014

STRC | **14th Swiss Transport Research Conference**

Monte Verità / Ascona, May 14-16, 2014

A dynamic discrete-continuous choice model for car ownership and usage

Aurélie Glerum, Michel Bierlaire
School of Architecture, Civil and Environmental Engineering (ENAC), Transport and Mobility Laboratory (TRANSP-OR)
École Polytechnique Fédérale de Lausanne (EPFL)
Lausanne, Switzerland

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

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

April 2014

Abstract

This research focuses on the formulation of a dynamic discrete-continuous choice model to explain and predict individuals' choices regarding car ownership, choice of fuel type and choice of annual driving distance. Its main contribution is to integrate dynamic choice modeling for multiple-car households and discrete-continuous choice modeling. This paper presents the formulation of the model and its estimation on synthetic data.

Keywords

Dynamic discrete-continuous choice modeling, dynamic programming, discrete choice modeling, constant elasticity of substitution, car ownership and usage modeling.

Front picture: Shell Spirit and Motor Oils, René Vincent, 1926.

1 Introduction

Numerous governments have in the past implemented policies aiming at reducing green house gas emissions and favoring the introduction of alternative fuel vehicles in the market. In this context, quantitative models play an important role in understanding and predicting the changes in demand in response to policy changes. The literature on car related choice models is vast but there appears to be a consensus that car ownership (number of cars) and car usage (distance driven with each car) are interrelated in household decisions and should be modeled simultaneously. Moreover, a car is a highly durable good that can be used over a long period of time. It hence is essential to account the forward-looking behavior of households. Dynamic discrete choice models (DDCM) (Aguirregabiria and Mira, 2010, Rust, 1987) address that feature by explicitely accounting for the expected discounted utility of future actions.

Modeling the discrete-continuous choice of car ownership and usage while accounting for forward-looking households is essential to predict the evolution of car demand. However the literature is lacking such methods.

In this research we develop a *dynamic discrete-continuous choice model (DDCCM)* that jointly models car replacement decisions, choice of fuel type and usage of each car within a household. The contribution of this research is to bring together the two advanced methodologies of discrete-continuous choice modeling and DDCM, to jointly model household's decisions regarding ownership and usage.

The aim of this paper is to provide an operational estimation framework for a complex choice problem. To validate the approach, the model is estimated on a synthetic data set. The latter is generated based on the distributional properties of the population and car registers in Sweden from 1999 to 2008. An earlier version of the model is presented in Glerum *et al.* (2013).

The paper is structured as follows. Section 2 presents the model framework and estimation procedure of the DDCCM. Section 3 provides an example of application of the model on a synthetic data set which is generated based on the statistical properties of the car fleet and population of Sweden. Section 4 discusses potential extensions of the DDCCM. Section 5 concludes the paper by outlining the next steps of this research.

2 The dynamic discrete-continuous choice modeling framework

In this section we present the DDCCM framework. We start by stating the main assumptions on which the model is based. Then we describe the model structure, from the base components to the specification of the full model. One of the key elements of the choice variable is the annual mileage of each car and we explain in detail its specification. We end the section by discussing a possible estimation method of the model.

2.1 Main assumptions

The DDCCM is formulated as a discrete-continuous choice model that is embedded into a *dynamic programming (DP)* framework. We model the joint decision of vehicle transactions, mileage and fuel type, based on the following assumptions.

Decisions are made at a household level. In addition, we assume that each household can have at most two cars. Larger household fleets may also be considered but at the cost of increased complexity. As pointed out by de Jong and Kitamura (2009), it may be relevant to consider three car households for prediction even though the current share in several markets (typically European markets) is low.

The choice of vehicle transaction and fuel type(s) is *strategic*, that is, we assume that households take into account the future utility of the choice of these variables in their decision process.

We consider an infinite-horizon problem to account for the fact that households make long-term decisions in terms of car transactions and fuel type. For example, individuals are assumed to strategically choose the fuel type of the car they purchase according to their expectation of fuel prices in the next years, or they decide to purchase one only car at present but already know that they might add another car in the future years.

We make the simplifying assumption that when households decide how much they will drive their car for the upcoming year, they only consider the utility of this choice for that particular year, but that they do not account for whether the residual value of their car is affected by usage. In other words, 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.

Similarly to de Jong (1996) we make the reasonable assumption that the choice of mileage(s) is conditional on the choice of the discrete decision variables (i.e. the transaction type and the fuel type).

2.2 Definition of model components

The DP framework is based on four fundamental elements: the *state space*, the *action space*, the *transition function* and the *instantaneous utility*. In this section, we describe each of these in detail.

The *state space* S is constructed based on the following variables.

- The *age* y_{ctn} of car c of household n in year t. We set an upper bound for the age $\bar{Y}$, assuming that above this upper bound, changes in age do not affect the utility or transition from one state to another. This implies that we have $y_{ctn} \in Y = \{0, 1, \ldots, \bar{Y}\}$.
- The *fuel type* f_{ctn} of car c of household n in year t. A car c can have any fuel type $f_{ctn} \in F = \{0, 1, \ldots, \bar{F}\}$, where $1, \ldots, \bar{F}$ is the list of available fuel types in the market of interest. The level 0 indicates the absence of a car.

As described in Section 2.1, each household can have at maximum two cars. Each state $s_{tn} \in S$ can hence be represented as

$$s_{tn} = (y_{1tn}, f_{1tn}, y_{2tn}, f_{2tn}), \tag{1}$$

where the car denoted by the index 1 is the car which has been in a household n's fleet for the longest time, and the car denoted by the index 2 is the car which entered the household in a later stage.

The *action space* A is constructed based on the following variables.

- The *transaction* h_{tn} in household n's composition of the car fleet in year t. Every year, the household can choose to increase, decrease or replace all or part of the fleet, or do nothing. We additionally make the simplifying assumption that a household cannot purchase more than one car per time period. The enumeration (see Figure 1) leads to nine possible transactions. Therefore we have $h_{tn} \in H = \{1, \ldots, 9\}$.
- The *annual mileage* $\tilde{m}_{ctn} \in \mathbb{R}^{+}$ of each car c chosen by household n.
- The *fuel type* $\tilde{f}_{ctn} \in F$ of each car c chosen by household n.

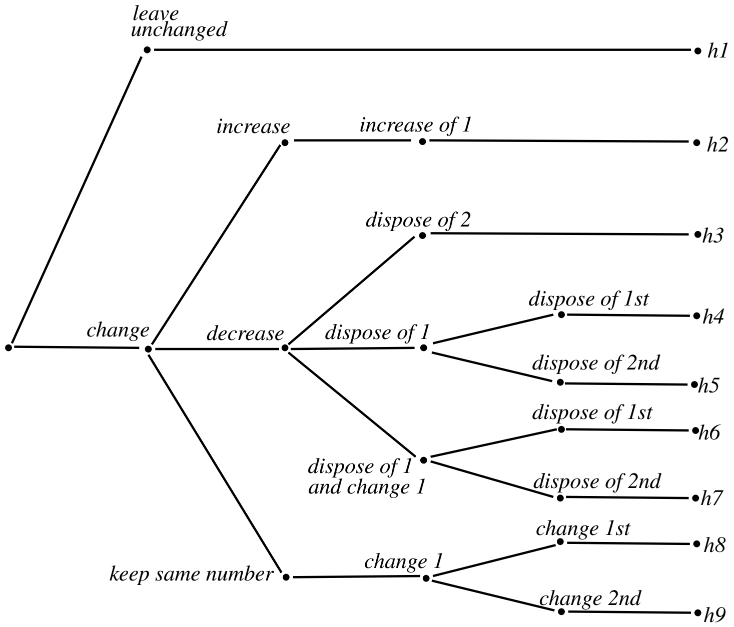

Figure 1: The nine possible transactions in a household fleet

Each action $a_{tn} \in A$ can be represented as

$$a_{tn} = (h_{tn}, \tilde{m}_{1tn}, \tilde{f}_{1tn}, \tilde{m}_{2tn}, \tilde{f}_{2tn}). \tag{2}$$

It is worth noting that we have a completely discrete state space, while the action space is discrete-continuous. From some particular states s_{tn}, not all actions are available. Hence, we implicitly have $a_{tn} \in A(s_{tn})$ and the total number of discrete actions must be obtained by enumerating all possible actions from each particular state.

Given that a household n is in a state s_{tn} and has chosen an action a_{tn}, the *transition function* $f(s_{t+1,n}|s_t, a_{tn})$ is defined as the rule mapping s_{tn} and a_{tn} to the next state $s_{t+1,n}$.

Assuming that $a^D_{tn} = (h_{tn}, \tilde{f}_{1tn}, \tilde{f}_{2tn})$ gathers the discrete components of an action a_{tn} and $a^C_{tn} = (\tilde{m}_{1tn}, \tilde{m}_{2tn})$ gathers the continuous components, the *instantaneous utility* is defined as:

$$u(s_{tn}, a^C_{tn}, a^D_{tn}, x_{tn}, \theta) = v(s_{tn}, a^C_{tn}, a^D_{tn}, x_{tn}, \varepsilon_C(a^C_{tn}), \theta) + \varepsilon_D(a^D_{tn}), \tag{3}$$

where variable x_{tn} contains socio-economic information relative to the household, θ is a vector of parameters to be estimated. Expression $v(s_{tn}, a^C_{tn}, a^D_{tn}, x_{tn}, \varepsilon_C(a^C_{tn}), \theta)$ is a deterministic term, $\varepsilon_D(a^D_{tn})$ is a random error term for the discrete actions and $\varepsilon_C(a^C_{tn})$ captures the randomness inherent to the continuous decision(s). Similarly as proposed by Rust (1987), the instantaneous utility has an additive-separable form.

2.3 Value function

As in a DDCM case (see e.g. Cirillo and Xu, 2011, Aguirregabiria and Mira, 2010), the value function of the DDCCM is defined as:

$$
\begin{aligned}
V(s_{tn}, x_{tn}, \theta) &= \max_{a_{tn}\in A}\{u(s_{tn}, a_{tn}, x_{tn}, \theta) + \beta \sum_{s_{t+1,n}\in S} \bar{V}(s_{t+1,n}, x_{t+1,n}, \theta) f(s_{t+1,n}|s_{tn}, a_{tn})\} \\
&= \max_{a_{tn}\in A}\{v(s_{tn}, a^C_{tn}, a^D_{tn}, x_{tn}, \varepsilon_C(a^C_{tn}), \theta) + \varepsilon_D(a^D_{tn}) + \beta \sum_{s_{t+1,n}\in S} \bar{V}(s_{t+1,n}, x_{t+1,n}, \theta) f(s_{t+1,n}|s_{tn}, a_{tn})\}
\end{aligned}
\tag{4}
$$

In order to obtain a version of the Bellman equation that does not depend on the random utility error term $\varepsilon_D(a^D_{tn})$, we consider the *integrated value function* $\bar{V}(s_{tn}, x_{tn}, \theta)$, given as follows.

$$
\bar{V}(s_{tn}, x_{tn}, \theta) = \int V(s_{tn}, x_{tn}, \theta) dG_{\varepsilon_D}(\varepsilon_D(a^D_{tn})) \tag{5}
$$

where G_{ε_D} is the CDF of ε_D.

In the case where all actions are discrete and the random terms $\varepsilon_D(a^D_{tn})$ are i.i.d. extreme value, it corresponds to the logsum (see e.g. Aguirregabiria and Mira, 2010). We aim at finding a closed-form formula in the case where the choices are both discrete and continuous too. In fact, a closed-form formula is possible in the special case where the choice of mileage of each car in the household is assumed myopic. This implies that individuals choose how much they wish to drive their car(s) every year, without accounting for the expected discounted utility of this choice for the following years[1].

Under the hypothesis of myopicity of the choice of annual driving distance(s), the integrated value function is obtained as follows.

$$
\begin{aligned}
\bar{V}(s_{tn}, x_{tn}, \theta) &= \int V(s_{tn}, x_{tn}, \theta) dG_{\varepsilon}(\varepsilon_D(a^D_{tn})) \\
&= \int \max_{a_{tn}\in A}\{u(s_{tn}, a_{tn}, x_{tn}, \theta, \varepsilon(a_{tn})) + \beta \sum_{s_{t+1,n}\in S} \bar{V}(s_{t+1,n}, x_{t+1,n}, \theta) f(s_{t+1,n}|s_{tn}, a_{tn})\} dG_{\varepsilon}(\varepsilon_D(a^D_{tn})) \\
&= \int \max_{a^D_{tn}}\{\max_{a^C_{tn}}\{v(s_{tn}, a^C_{tn}, a^D_{tn}, x_{tn}, \varepsilon_C(a^C_{tn}), \theta)\} + \varepsilon_D(a^D_{tn}) + \beta \sum_{s_{t+1,n}\in S} \bar{V}(s_{t+1,n}, x_{t+1,n}, \theta) f(s_{t+1,n}|s_{tn}, a_{tn})\} dG_{\varepsilon}(\varepsilon_D(a^D_{tn})) \\
&= \log \sum_{a^D_{tn}} \exp\{\max_{a^C_{tn}}\{v(s_{tn}, a^C_{tn}, a^D_{tn}, x_{tn}, \varepsilon_C(a^C_{tn}), \theta)\} + \beta \sum_{s_{t+1,n}\in S} \bar{V}(s_{t+1,n}, x_{t+1,n}, \theta) f(s_{t+1,n}|s_{tn}, a_{tn})\}
\end{aligned}
\tag{6}
$$

[1] This assumption was also made in the unpublished work by Anders Munk-Nielsen, University of Copenhagen. We make this reasonable hypothesis here too.

Similarly as in the case of a DDCM, the value function is obtained by iterating on Equation (6).

2.4 The utility function in details

Since our aim is to model both acquisition and usage, we assume that expression $v(s_{tn}, a^C_{tn}, a^D_{tn}, x_{tn}, \varepsilon_C(a^C_{tn}), \theta)$ of Equation (6) is the sum of a utility linked with the acquisition the vehicles v^D_{tn} and a utility linked with the usage of the car v^C_{tn}:

$$v(s_{tn}, a^C_{tn}, a^D_{tn}, x_{tn}, \varepsilon_C(a^C_{tn}), \theta) = v^D_{tn}(s_{tn}, a^D_{tn}, x_{tn}, \theta) + v^C_{tn}(s_{tn}, a^D_{tn}, a^C_{tn}, x_{tn}, \varepsilon_C(a^C_{tn}), \theta) \quad (7)$$

2.4.1 Optimal mileage for households owning two cars with different fuel types

By assumption, each household can have at maximum two cars. This implies that for two-car households, the annual mileage of each car must be decided every year. Expression $v(s_{tn}, a^C_{tn}, a^D_{tn}, x_{tn}, \varepsilon_C(a^C_{tn}), \theta)$ of Equation (6) must hence be maximized with respect to the two annual driving distances. Given the additive form of Equation (7), we only need to maximize expression $v^C_{tn}(s_{tn}, a^D_{tn}, a^C_{tn}, x_{tn}, \varepsilon_C(a^C_{tn}), \theta)$ with respect to a^C_{tn}.

However, if a household owns two cars, we observe from the data that one car is generally driven more than the other one, i.e. one is used for long distances while the other is used for shorter trips. We therefore make the assumption that the choice a household actually makes is not the independent choices of how much each car will be driven, but rather the repartition of the total mileage that it plans to drive across the two cars. Moreover, the use of both cars in the household is highly dependent on fuel prices. Hence in the common case of a two-car household owning both a car of a fuel f_1 and a car of a fuel f_2 (e.g. a diesel and a gasoline car), the repartition of mileages might fluctuate depending on this economic feature.

In the abovementioned case, this motivates the use of a CES utility function for the choice of mileages for cars of different fuel types within a same household, since it allows to evaluate how likely households substitute the use of one car with the other, when the difference between the fuel prices is changing.

Let us denote the mileages of the chosen cars with fuels f_1 and f_2 as $\tilde{m}_{f_1tn}$ and $\tilde{m}_{f_2tn}$, respectively. They are defined as follows.

$$\tilde{m}_{f_1tn} := \tilde{m}_{1tn} \cdot I_{1f_1tn} + \tilde{m}_{2tn} \cdot I_{2f_1tn} \quad (8)$$

and

$$\tilde{m}_{f_2tn} := \tilde{m}_{1tn} \cdot I_{1f_2tn} + \tilde{m}_{2tn} \cdot I_{2f_2tn}, \tag{9}$$

where I_{cf_1tn} is equal to c if car $c \in \{1, 2\}$ is a car driven with fuel f_1 (e.g. gasoline), 0 otherwise, and $I_{cf_2tn} = 1 - I_{cf_2tn}$ is an indicator of whether the car c is driven with fuel f_2 (e.g. diesel).

The deterministic utility of driving is given by the following CES function.

$$v^C_{tn}(s_{tn}, a^D_{tn}, a^C_{tn}, x_{tn}, \varepsilon_C(a^C_{tn}), \theta) = (\tilde{m}^{\rho}_{f_1tn} + \tilde{m}^{\rho}_{f_2tn})^{\frac{1}{\rho}} \tag{10}$$

For simplicity, we assume that the formulation does not include any randomness related to the choice of the continuous variable. Therefore we will omit the variable $\varepsilon_C(a^C_{tn})$. Nevertheless the framework could be extended to include randomness linked with the choice of annual driving distance.

Parameter ρ with $\rho \leq 1$ and $\rho \neq 0$, is related to the elasticity of substitution σ between the two cars, given by:

$$\sigma = \frac{1}{1 - \rho} \tag{11}$$

The choice of $\tilde{m}_{f_1tn}$ and $\tilde{m}_{f_2tn}$ must be made such that the budget constraint of the household holds:

$$p_{f_1tn}\tilde{m}_{f_1tn} + p_{f_2tn}\tilde{m}_{f_2tn} = \text{Inc}_{tn}, \tag{12}$$

where $p_{ftn} := \text{cons}_{ftn} \cdot \text{pl}_{ft}$ is the cost per km of driving a car with fuel $\tilde{f} \in \{f_1, f_2\}$, that is the product of the car consumption cons_{ftn} and the price of a liter of fuel pl_{ft} for that car. Variable Inc_{tn} is the share of the household's annual income which is used for expenses related to car fueling.

The optimal value of mileages for both cars is obtained by solving the following maximization problem.

$$\max_{\tilde{m}_{f_1tn}, \tilde{m}_{f_2tn}} v^C_{tn}, \text{ such that } p_{f_1tn}\tilde{m}_{f_1tn} + p_{f_2tn}\tilde{m}_{f_2tn} = \text{Inc}_{tn} \tag{13}$$

The above formulation of the CES utility fonction with the budget constraint has the following advantages. First, the constraint enables us to solve the maximization problem according to one dimension only. Such an approach has been considered by Zabalza (1983), in a context of trade-off between leisure and income. Second, the use of a CES function is also convenient, since the elasticity of substitution is directly obtained from the estimate of parameter ρ.

An analytical solution for $\tilde{m}_{f_2tn}$ can be obtained by solving the above optimization problem:

$$\tilde{m}^*_{f_2tn} = \frac{\text{Inc}_{tn} \cdot p_{f_2tn}^{(1/(1-\rho))}}{p_{f_1tn}^{(\rho/(\rho-1))} + p_{f_2tn}^{(\rho/(1-\rho))}}. \tag{14}$$

We can then infer the value of the optimal mileage for the car with the other fuel type:

$$\begin{aligned} \tilde{m}^*_{f_1tn} &= \frac{\text{Inc}_{tn}}{p_{f_1tn}} - \frac{p_{f_2tn}}{p_{f_1tn}}\tilde{m}^*_{f_2tn} \\ &= \frac{\text{Inc}_{tn}}{p_{f_1tn}} - \frac{p_{f_1tn}}{p_{f_1tn}} \cdot \frac{\text{Inc}_{tn} \cdot p_{f_2tn}^{(1/(1-\rho))}}{p_{f_1tn}^{(\rho/(\rho-1))} + p_{f_2tn}^{(\rho/(1-\rho))}} \end{aligned} \tag{15}$$

Consequently, we obtain the optimal value for the deterministic utility of the continuous actions:

$$v^{C*}_{tn} = \left(\left(\frac{\text{Inc}_{tn}}{p_{f_1tn}} - \frac{p_{f_2tn}}{p_{f_1tn}} \cdot \frac{p_{f_2tn}^{(1/(\rho-1))} \cdot \text{Inc}_{tn}}{p_{f_1tn}^{(\rho/(\rho-1))} + p_{f_2tn}^{(\rho/(\rho-1))}} \right)^{\rho} + \left(\frac{\text{Inc}_{tn} \cdot p_{f_2tn}^{(1/(\rho-1))}}{p_{f_1tn}^{(\rho/(\rho-1))} + p_{f_2tn}^{(\rho/(\rho-1))}} \right)^{\rho} \right)^{\frac{1}{\rho}} \tag{16}$$

Then v^{C*}_{tn} can be inserted back in Equation (7). The Bellman equation (6) becomes:

$$\begin{aligned} \bar{V}(s_{tn}, x_{tn}, \theta) &= \log \sum_{a^D_{tn}} \exp\{v^D_{tn}(s_{tn}, a^D_{tn}, x_{tn}, \theta) + v^{C*}_{tn}(s_{tn}, a^D_{tn}, a^{C*}_{tn}, x_{tn}, \theta) \\ &+ \beta \sum_{s_{t+1,n} \in S} \bar{V}(s_{t+1,n}, x_{t+1,n}, \theta) f(s_{t+1,n} | s_{tn}, a_{tn})\}, \end{aligned} \tag{17}$$

where $a^{C*}_t = (\tilde{m}^*_{1,t}, \tilde{m}^*_{2,t})$. The integrated value function $\bar{V}$ can then be computed by value iteration. Let us note that the optimal mileage(s) for car 1 and car 2 are obtained by the following mappings.

$$\tilde{m}^*_{1tn} = \tilde{m}^*_{f_1tn} \cdot I_{1f_1tn} + \tilde{m}^*_{f_2tn} \cdot I_{1f_2tn} \tag{18}$$

$$\tilde{m}^*_{2tn} = \tilde{m}^*_{f_1tn} \cdot I_{2f_1tn} + \tilde{m}^*_{f_2tn} \cdot I_{2f_2tn} \tag{19}$$

2.4.2 Optimal mileage for one-car households and households with two cars of the same fuel

The computation of the optimal mileage(s) for one-car households and for households owning two cars with the same fuel types is a result of the optimization problem (13) in two special cases.

For one-car households with a car of fuel f_1, the optimization problem reduces to the following problem.

$$\max_{\tilde{m}_{f_1tn}} v^C_{tn}, \text{ such that } p_{f_1tn}\tilde{m}_{f_1tn} = \text{Inc}_{tn} \tag{20}$$

The optimal mileage $\tilde{m}^*_{1tn}$ for the only car in the household is hence given by:

$$\tilde{m}^*_{1tn} = \tilde{m}^*_{f_1tn} = \frac{\text{Inc}_{tn}}{p_{f_1tn}} \tag{21}$$

Consequently the optimal utility of driving is given by:

$$v^{C*}_{tn} = \tilde{m}^*_{1tn} \tag{22}$$

For two-car households where both cars have the same fuel f_1, we assume perfect substitutability between the two cars, implying that ρ is equal to 1. This special case leads to infinitely many solutions to the optimization problem (13). We hence select the solution where both cars are driven the same distance, i.e. $\tilde{m}_{1tn} = \tilde{m}_{2tn} = \tilde{m}_{f_1tn}$. The optimization problem hence becomes:

$$\max_{\tilde{m}_{f_1tn}} v^C_{tn}, \text{ such that } 2 \cdot p_{f_1tn}\tilde{m}_{f_1tn} = \text{Inc}_{tn} \tag{23}$$

Therefore the optimal mileages for each car $\tilde{m}^*_{1tn}$ and $\tilde{m}^*_{2tn}$ are given by:

$$\tilde{m}^*_{1tn} = \tilde{m}^*_{2tn} = \tilde{m}^*_{f_1tn} = \frac{\text{Inc}_{tn}}{2 \cdot p_{f_1tn}}, \tag{24}$$

where p_{f_1tn} is the cost per km of driving either car (under the assumption that both cars have the same consumption).

The optimal utility of driving is hence given by:

$$v_{tn}^{C*} = 2 \cdot \tilde{m}_{1tn}^{*} \tag{25}$$

In both cases presented in this section, v_{tn}^{C*} is consequently introduced in the Bellman equation (6).

2.5 Maximum likelihood estimation

The parameters of the DDCCM are obtained by maximizing the following likelihood function.

$$\mathcal{L}(\theta) = \prod_{n=1}^{N} \prod_{t=1}^{T_n} P(a_{tn}^{D} | s_{tn}, x_{tn}, \theta), \tag{26}$$

where N is the total population size, T_n is the number of years household n is observed and $P(a_{tn}^{D} | s_{tn}, x_{tn}, \theta)$ is the probability that household n chooses a particular discrete action a_{tn}^{D} at time t. This probability is obtained as follows.

$$P(a_{tn}^{D} | s_{tn}, x_{tn}, \theta) = \frac{v_{tn}^{D}(s_{tn}, a_{tn}^{D}, x_{tn}, \theta) + v_{tn}^{C*}(s_{tn}, a_{tn}^{D}, a_{tn}^{C*}, x_{tn}, \theta) + \beta \bar{V}(s'_{t+1,n}, x_{t+1,n}, \theta)}{\sum_{\tilde{a}_{tn}^{D}} \left\{ v_{tn}^{D}(s_{tn}, \tilde{a}_{tn}^{D}, x_{tn}, \theta) + v_{tn}^{C*}(s_{tn}, \tilde{a}_{tn}^{D}, \tilde{a}_{tn}^{C*}, x_{tn}, \theta) + \beta \sum_{s_{t+1,n} \in S} \bar{V}(s_{t+1,n}, x_{t+1,n}, \theta) f(s_{t+1,n} | s_{tn}, a_{tn}) \right\}} \tag{27}$$

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. Our DP problem is quite simple because of the transition function being deterministic and we will adopt this approach here.

3 Application

In order to the validate the proposed modeling approach, we specify a DDCCM with arbitrary parameters and estimate it on synthetic data. The synthetic data is generated based on the distributional properties of the registers of vehicles and individuals in Sweden in years 1999 to 2008.

In this section, we describe simplifying assumptions relative to this example, the model specification, the data set generation procedure and the estimation results.

3.1 Assumptions

In this example of application, we consider the two following fuel types f_{ctn}: gasoline or diesel. Alternative-fuel vehicles are not included due to a too low frequency in the years of the data.

Moreover we assume a deterministic transition function, implying that $s_{t+1,n}$ can be inferred deterministically from s_t and a_t. The transition function $f(s_{t+1}|s_t, a_t)$ hence reduces to the following indicator.

$$f(s_{t+1}|s_t, a_t) = \begin{cases} 1 & \text{if } s_t \text{ and } a_t \text{ lead to state } s_{t+1} \\ 0 & \text{otherwise} \end{cases} \tag{28}$$

This implies that Equation (6) can be simplified as follows due the deterministic formulation of the transition.

$$\bar{V}(s_{tn}, x_{tn}, \theta) = \log \sum_{a^D_{tn}} \exp\{\max_{a^C_{tn}}\{v(s_{tn}, a^C_{tn}, a^D_{tn}, x_{tn}, \theta)\} + \beta \bar{V}(s'_{t+1,n}, x_{t+1,n}, \theta)\}, \tag{29}$$

where $s'_{t+1,n}$ is the state deterministically reached from s_{tn} if action a_{tn} is chosen. We note that this simplification decreases considerably the computational time.

3.2 Model specification

As an example, we consider a simple specification of the deterministic (instantaneous) utility relative to the choice of the discrete variables:

$$v^D_{tn}(s_{tn}, a^D_{tn}, x_{tn}, \theta) = \tau(a^D_{tn}) + \beta_{\text{Age}}(a^D_{tn}, s_{tn}) \cdot \max(\text{Age1}_{tn}, \text{Age2}_{tn}), \tag{30}$$

where Age1_{tn} is the age of the first of the two cars in household n and Age2_{tn} is the age of the second car. Expressions $\tau(a^D_{tn})$ and $\beta_{\text{Age}}(a^D_{tn}, s_{tn})$ are parameters that would typically be estimated on data. The transaction cost is meant to capture the unobserved costs (e.g. search costs) of actions involving the acquisition of a new car, i.e. actions with transactions h_2 (increase of 1),

h_6 (dispose of 1st and change 2nd), h_7 (dispose of 2nd and change 1st), h_8 (change 1st) or h_9 (change 2nd).

In order to illustrate the application of the DDCCM, we choose values for the parameters of Equation (30). The signs and values are chosen in order to match a priori expectations. For example, in a one-car household, the older the car is, the more likely the household is to dispose of it. Hence, we give a positive sign to parameter β_{Age}. The other chosen parameter values are reported in Table 1. Parameter β_{Age} depends on the size of the household, on whether the first or the second car is the oldest and on the transaction type. The transaction cost τ varies according to the different transactions types.

		β_{Age}			τ
Transaction name	Case	0 car	1 car	2 cars	all households
h_1: leave unchanged		0	0	0	0
h_2: increase 1		0	0	-	-3
h_3: dispose 2		-	-	0	0
h_4: dispose 1st	1st car is oldest	-	1.5	1.5	0
	2nd car is oldest	-	-	0	0
h_5: dispose 2nd	1st car is oldest	-	-	0	0
	2nd car is oldest	-	-	1.5	0
h_6: dispose 1st & change 2nd		-	-	0	-4
h_7: dispose 2nd & change 1st		-	-	0	-4
h_8: change 1st	1st car is oldest	-	1.5	1.5	-4
	2nd car is oldest	-	-	0	-4
h_9: change 2nd	1st car is oldest	-	-	0	-4
	2nd car is oldest	-	-	1.5	-4

Table 1: Parameters for the deterministic (instantaneous) utility relative to the discrete actions

Likewise, we choose values for the parameters of the deterministic (instantaneous) utility relative to annual mileage. The elasticity of substitution ρ is set to 0.5. Moreover we fix the discount factor in the Bellman equation (29) to 0.5.

3.3 Generation of the data set

We generate a synthetic data set of 5000 observations (with 500 observations for each year from 1999 to 2008). In order to provide a realistic example, the attributes of each observations are generated such that they reproduce the distributional properties of the Swedish car fleet and population. In Table 2 we report summary statistics of each variable in the real market.

The only attribute which we did not generate according to the data from the registers is the number of cars per household. In order to obtain more variability in the data, we have assumed an equal proportion of 0-car, 1-car and 2-car households ($\cong 33.3\%$).

We note that the percentages of gasoline and diesel cars are hence computed relatively to the population of cars with such fuel types only. Similarly, the proportion of households with 0 to 2 cars result from the normalization over the population of households that do not own more than 2 cars. To reduce the computational time of this estimation, we further assume that cars cannot become older than 9 years. We note that the percentages of cars with $0, \ldots, 9$ years result from a normalization over the population of cars no older than 9 years.

To build the synthetic data set, the age of each car and its fuel type are generated according to the proportions in Table 2. The households' incomes are generated according to a normal distribution truncated at 0, with mean and standard deviation indicated in the table.

The choice probabilities relative to each household are predicted based on the postulated parameters of Table 1. The choice indicator of each observation is generated following the inverse transform method.

3.4 Estimation results

In this section, we present the results of the estimation of the model described in Section 3.2 on the synthetic data. The estimation procedure is implemented in the C++ programming language and is run on a 24-core computer with processors Intel(R) Xeon(R) CPU X5680 3.33 GHz. An estimation lasts approximately 2 hour and 40 minutes.

To show that the model can consistently be estimated over several data sets, we generate 10 synthetic data sets based on the procedure described in Section 3.3. The model is estimated on each data set.

The parameter estimates for each data set are reported in Table 3. As expected, for each run, the parameters are not different from their true values, showing the consistency of the estimation procedure. For the 10th data set, t-tests could not be computed since the algorithm was already at the optimal point, according to the chosen stopping criterion (gradient norm < 0.01).

All parameters are significantly different from 0, showing that they significantly affect the choice.

Variable type	Variable name	Level	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
Market	Average fuel price [SEK]	Gasoline	8.4	9.56	9.52	9.37	9.46	10.05	11.13	11.55	11.65	12.54
		Diesel	6.66	8.44	8.69	8.36	7.92	8.61	10.48	11.27	10.88	13.12
Car fleet	Car age [% households]	0 year	12.77	12.57	9.02	8.64	8.57	8.21	7.72	7.16	7.42	5.45
		1 year	11.55	13.53	13.45	9.72	9.18	9.25	9.03	8.64	8.24	8.73
		2 years	11.14	11.80	13.81	13.54	9.89	9.46	9.56	9.45	9.22	9.06
		3 years	9.70	11.70	12.42	14.11	13.71	10.40	9.99	10.21	10.16	10.28
		4 years	8.90	9.73	11.99	12.46	13.82	13.66	10.62	10.35	10.63	11.01
		5 years	8.42	8.71	9.52	11.52	11.73	13.22	13.21	10.42	10.34	10.90
		6 years	6.70	8.08	8.43	9.06	10.72	11.10	12.64	12.76	10.25	10.43
		7 years	8.42	6.38	7.76	7.98	8.38	10.03	10.49	12.13	12.41	10.23
		8 years	10.18	7.95	6.08	7.30	7.34	7.82	9.42	9.98	11.74	12.30
		9 years	12.20	9.55	7.53	5.66	6.67	6.84	7.31	8.90	9.58	11.61
	Fuel type [% households]	Gasoline	97.23	97.04	97.10	97.05	97.01	96.96	96.91	96.47	95.37	94.44
		Diesel	2.77	2.96	2.90	2.95	2.99	3.04	3.09	3.53	4.63	5.56
Household	Number of cars [% households]	0 car	43.25	42.93	42.97	42.93	42.91	42.86	43.01	45.04	45.15	45.41
		1 car	44.96	44.76	44.69	44.65	44.57	44.44	44.20	42.54	42.35	42.11
		2 cars	11.79	12.30	12.35	12.41	12.52	12.70	12.79	12.42	12.50	12.48
	Income [SEK]	Mean	185’508	197’706	201’695	210’277	214’197	218’315	226’946	232’715	254’452	259’523
		SD	321’885	667’570	631’202	429’462	298’663	237’607	224’982	465’895	338’340	981’006

Table 2: Summary statistics of the variables included in the synthetic data set (market, fleet and population attributes) per year.

Run	ρ				β_{Age} 1 car Dispose / change				β_{Age} 2 cars Dispose / change oldest car				τ Increase of 1				τ Dispose of 1 and change the other / change 1				LL
	Value	**SD**	***t*-test 0**	***t*-test true value**	**Value**	**SD**	***t*-test 0**	***t*-test true value**	**Value**	**SD**	***t*-test 0**	***t*-test true value**	**Value**	**SD**	***t*-test 0**	***t*-test true value**	**Value**	**SD**	***t*-test 0**	***t*-test true value**	
1	0.49	0.03	15.48	-0.28	1.51	0.04	34.51	0.20	1.49	0.04	36.86	-0.15	-3.02	0.06	-48.70	-0.27	-3.99	0.06	-71.74	0.10	0.96
2	0.51	0.03	18.78	0.19	1.47	0.04	34.42	-0.61	1.50	0.04	36.60	0.11	-2.97	0.06	-53.24	0.45	-4.00	0.05	-73.07	0.07	0.96
3	0.50	0.03	17.60	-0.13	1.52	0.04	35.38	0.57	1.49	0.04	36.74	-0.17	-2.99	0.06	-51.84	0.18	-3.98	0.06	-71.77	0.43	0.96
4	0.50	0.03	19.95	0.02	1.49	0.04	34.83	-0.19	1.48	0.04	36.21	-0.58	-2.98	0.05	-55.85	0.42	-3.96	0.05	-73.09	0.79	0.97
5	0.50	0.03	17.80	-0.10	1.49	0.04	34.60	-0.19	1.48	0.04	36.56	-0.44	-2.97	0.06	-52.00	0.52	-4.01	0.06	-72.61	-0.26	0.95
6	0.50	0.03	17.23	-0.10	1.48	0.04	34.56	-0.40	1.49	0.04	36.50	-0.28	-3.01	0.06	-52.32	-0.22	-4.02	0.05	-73.88	-0.33	0.95
7	0.50	0.02	20.94	0.06	1.45	0.04	34.47	-1.21	1.50	0.04	36.79	-0.08	-2.97	0.05	-55.22	0.62	-3.99	0.05	-73.13	0.19	0.97
8	0.49	0.03	17.14	-0.27	1.46	0.04	33.78	-0.84	1.51	0.04	36.87	0.34	-2.96	0.06	-50.84	0.74	-4.01	0.06	-71.54	-0.19	0.97
9	0.50	0.03	17.80	-0.07	1.44	0.04	34.23	-1.38	1.48	0.04	36.67	-0.42	-2.93	0.06	-51.68	1.27	-3.96	0.05	-73.62	0.69	0.96
10	0.5	-	-	-	1.5	-	-	-	1.5	-	-	-	-3	-	-	-	-4	-	-	-	-
Average value	0.50				1.48				1.49				-2.98				-3.99				0.96
True value	0.5				1.5				1.5				-3				-4				
Initial value	0.5				1.5				1.5				-3				-4				

Table 3: Model estimation results, including parameter values, standard deviations (SD), values of t-test against 0 and against the true values, and final log-likelihoods (normalized by the number of households).

This is expected, since the choice variable is generated based on the true values of the postulated parameters.

Since all parameters are not significantly different from their true values and significantly different from 0, we can conclude that the proposed estimation framework leads to consistent estimates.

4 Discussion

In this section, we discuss the limitations of the approach presented in this paper and potential extensions of the presented framework.

First of all, we have integrated the choice of fuel type in the action space, while we did not integrate the consumption(s) of the chosen car(s). In terms of specification, this implies that the consumption for the chosen car is assumed to be constant, independently of the car type which is chosen. This is a rather restrictive assumption, which could be relaxed in future research. Anders Munk-Nielsen takes another approach which has the advantage of including fuel efficiently in the state space. This yields a more complex model but is a less restrictive assumption than ours.

In this paper, we have chosen to model choices regarding transaction type, fuel type and annual driving distance. However, the framework could be extended to incorporate other acquisition decisions such as purchasing a new versus a second-hand car, or selecting a company car, if this option is possible. If decisions regarding ownership status (i.e. the acquisition of a company versus a private car) are modeled, more research should then be performed in order to identify the individuals who really have access to a company car or not.

It is computationally demanding to estimate the model using the nested fixed point algorithm. In future research, we plan to adopt the approach by Aguirregabiria and Mira (2002) which reduces the number of times the DP problem needs to be solved.

The present approach does not account for correlation between observations of the same household. In future research we plan to integrate such effects to improve the model realism.

5 Conclusion

This paper presents a methodology to model jointly car ownership, usage and choice of fuel type. We have developed an operational estimation framework and shown its consistency by estimating it on sythetic data. One of the main properties of the model is that it accounts for the forward-looking behavior of individuals. This is crucial in the case of demand for durable goods such as cars, since the purchase of a car is affected by the utility gained from that car for the present and future years of ownership.

Though decisions involving discrete and continuous components are frequently occurring, this aspect is scarcely addressed in a dynamic models with forward-looking agents. To address this difficulty, we have formulated the problem of modeling ownership and usage decisions as a dynamic discrete-continuous choice model.

The present research has the following important contributions. First, in order to obtain a realistic dynamic model, we account for households' decisions rather than individual ones, and model transaction decisions for single- and two-car households. Second, we specify a CES utility function to capture substitution patterns that occur when two-car households decide on the annual mileages of the cars. Hence annual driving distances are not only treated as continuous, but the dependence between each other is also accounted for. Third, we consider a comprehensive choice variable, that accounts for decisions that are usually jointly made, such as car ownership, choice of fuel type and annual mileage.

Future works involve the estimation of the DDCCM on the Swedish register data, in order to assess its benefits over a simple static discrete choice model. The estimation on real data will allow for interesting analyses. The latter include the assessment of the impact of policies implemented during the years of the data on the dynamics of the Swedish car fleet and the prediction of the effect of policy scenarios that have been defined in the planning process of the Swedish government for the upcoming years.

Acknowledgements

This research was partially funded by the Center for Transport Studies, Stockholm. The authors would like to thank Roger Pyddoke for sharing part of the data. Previous results of this research have been presented at the 13th Swiss Transport Research Conference, TRISTAN VIII, the third International Choice Modelling Conference and the Second Symposium of the European

Association for Research in Transportation. The work presented in this paper has greatly benefited from comments received from the participants of these conferences.

6 References

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

Aguirregabiria, V. and P. Mira (2010) Dynamic discrete choice structural models: A survey, *Journal of Econometrics*, **156** (1) 38–67.

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

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

de Jong, G. C. and R. Kitamura (2009) A review of household dynamic vehicle ownership models: holdings models versus transactions models, *Transportation*, **36** (6) 733–743.

Glerum, A., E. Frejinger, A. Karlstrom, M. Beser Hugosson and M. Bierlaire (2013) A dynamic discrete-continuous choice model of car ownership, usage and fuel type, paper presented at the *Proceedings of the Second Symposium of the European Association for Research in Transportation*, Stockholm, Sweden.

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

Zabalza, A. (1983) The CES utility function, non-linear budget constraints and labour supply. results on female participation and hours, *The Economic Journal*, **93** (370) 312–330.