A dynamic discrete-continuous choice model of car ownership, usage and fuel type

Aurélie Glerum * Emma Frejinger † Anders Karlström ‡

Muriel Beser Hugosson ‡ Michel Bierlaire *

Second Symposium of the European Association for Research in Transportation
4th-6th September 2013, Stockholm

Abstract

We formulate a dynamic discrete-continuous choice model (DDCCM) of car ownership, usage and fuel type. The framework embeds a discrete-continuous choice model (DCCM) into a dynamic programming (DP) framework to account for the forward-looking behavior of households in the context of car acquisition. More specifically, we model the of transaction type, the choice of fuel type, the ownership status (private versus company car), the choice of car state (new versus second-hand) and the annual driving distance for up to two cars in the household fleet. In this paper, we present the methodological framework and demonstrate the applicability of such a model by showing a concrete example of application.

Key words

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

*ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE (EPFL), School of Architecture, Civil and Environmental Engineering (ENAC), Transport and Mobility Laboratory (TRANSP-OR), {aurelie.glerum, michel.bierlaire}@epfl.ch

†UNIVERSITÉ DE MONTRÉAL, Faculté des arts et des sciences, Department of Computer Science and Operations Research, frejinge@iro.umontreal.ca

‡ROYAL INSTITUTE OF TECHNOLOGY (KTH), School of Architecture and the Built Environment, Department of Transport Science, {anders.karlstrom, muriel.hugosson}@abe.kth.se

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. In this paper we propose such a model. 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 and that is often sold a number of times at the second-hand market. However, due to various transaction costs households do not generally change their fleet frequently (e.g. on a yearly basis). As proposed by Schiraldi (2011) this motivates the use of dynamic discrete choice models (DDCM) (Rust, 1987; Aguirregabiria and Mira, 2010) that explicitly takes into account the forward-looking behavior of households.

In this paper we specify a dynamic discrete-continuous choice model (DDCCM) that jointly models car usage and replacement decisions, including choice of fuel type and ownership status of each car. One of the main issues is how to model the continuous choice variables capturing the annual driving distances for each car. We address this issue with a constant elasticity of substitution (CES) utility function. This is an exploratory study where the main contribution lies in the specification of an operational model for a complex choice problem. We use the case of Sweden as an illustration but we emphasize that the proposed methodology is general and can be adapted or extended to different choice settings (e.g. including the choice of other car characteristics than fuel type). This is ongoing research and in this paper we present some numerical results for an illustrative example (without estimating the parameters).

A comprehensive review of the vast literature related to car ownership, type and usage choices is out of the scope of this paper (for reviews we refer the reader to e.g. De Jong et al., 2004; De Jong and Kitamura, 2009). Even though many sophisticated studies have been published on static models, we focus here on the fairly scarce literature directly related to our work, which deals with dynamic models taking into account the forward-looking behavior of decision-makers. One of the most notable studies is Schiraldi (2011). It focuses on the estimation of transaction costs in a dynamic framework based on aggregate data and analyzes the effect of a scrappage policy in Italy. Moreover Schiraldi (2011) models the price on the second-hand market and provides an excellent review of related literature. In this work we do not attempt to model the second-hand market, we simply differentiate the transaction of buying a new or a used car. Similarly to Schiraldi (2011) we assume that a decision-maker maximizes the expected discounted lifetime utility modeled by a value function that is the solution to the Bellman equation. Moreover we make the same assumptions (actually dating back to Rust, 1987) to deal with the 'curse of dimensionality' and obtain an operational model. As opposed to Schiraldi (2011) where a consumer holding a car a given year decides whether to hold, sell or scrap the car (if the consumer does not hold a car he/she decides to continue that way or to buy one), we have a more complex choice setting because a household can hold more than one car and we also model the usage of each car.

Related to our work is also the one by Xu (2011) who develops a dynamic discrete choice model to explain car acquisition decisions and choice of fuel type. Car usage is however not considered so there are no continuous choice variables. The model is applied to stated preferences data collected in Maryland and corresponds to a optimal stopping model similar to the one by Rust (1987). Other modeling approaches have been considered in order to jointly model car ownership and usage. A very interesting approach is presented by Gillingham (2012) who models cars' monthly mileage conditional on vehicle type. He integrates consumers' expectations about the cars' future resale prices and future gasoline prices after a six-year period.

It is important to mention an ongoing research project at University of Copenhagen with an objective similar to ours, namely a discrete-continuous dynamic choice model for transaction decisions and usage. Their presentation at the IRUC seminar (Copenhagen, December 2012) inspired us to view the continuous choice variable of car usage as a myopic choice conditional on the discrete choice variables (unpublished work by Anders Munk-Nielsen presented at the IRUC seminar at University of Copenhagen on 17th December 2012).

There is also literature on duration models that model the time elapsed between two car transactions. For instance, De Jong (1996) presents an interesting study based on a system of models including a duration model for the time between car replacement decisions and a regression model for annual car usage. The main difference between the dynamic model presented here and a duration model are that households are assumed to be forward-looking. This means that they optimize their choices taking expected future utility into account, and socio-economic characteristics are not assumed constant between transactions and that we can model several choices jointly.

The paper is structured as follows. Section 2 presents some information on the Swedish car market that is used as illustration. Section 3 presents the methodological framework of the DDCCM. Section 4 provides an illustration of the application of the model. Section 6 concludes the paper by outlining the next steps of this research.

2 The case of Sweden

The definition of important choice variables obviously depends on the application. In this study we use the case of Sweden mainly due to the fact that we have access to the entire Swedish population and car registers that allows us to follow cars, individuals and households from 1998 to 2008. Moreover, several different policies have been in place and act on different actors at different geographical levels (local, regional and national) with the objective of accelerating the introduction of clean cars in the fleet. Shifts in demand can be observed in response to these policies. As a first example, a significant increase of small diesel cars has been observed in the later years (Hugosson and Algers, 2012; Kageson, 2013). As a second example, fluctuations in the number of cars owned by each household have been observed. More precisely, the fraction of households without a car has been increasing from 2006 (see Figure 1), which is the year of the trial period of the Stockholm congestion tax, preceding the actual introduction of this system.

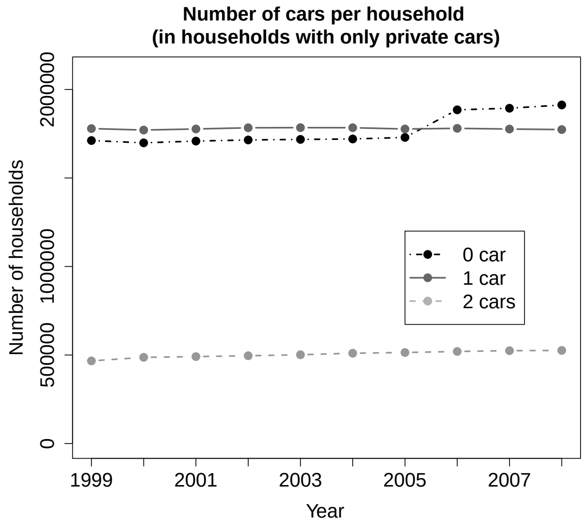


Figure 1: Evolution of car ownership in Sweden.

The registers are based on individuals. We have extensive socio-economic data such as net income, home and work locations, type of employment, etc. in addition to characteristics of each owned car (make, model, fuel type, fuel consumption, age, annual mileage from odometer readings, etc.). In addition we have information on all household types except for unmarried individuals living together without children. Part of this data (without car characteristics) was used by Pyddoke (2009).

A special characteristic of the Swedish car market is the high proportion of company-owned cars, which is a consequence of the fringe benefit taxation system. These cars are important to consider because they represent a large share of the new car sales and can be used privately by the households. However, from the registers we know if an individual pays a benefit tax and hence has access to a company car but we do not have any information on the car characteristics. The company owned cars therefore pose a number of challenges to the specification of our model which we attempt to address below.

3 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.

3.1 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, fuel type, use of a company car (if available) and purchase of a new or second-hand car, 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, since a very small share of the Swedish households has more than 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 the population is low.

The choice of vehicle transaction, fuel type(s), use of a company car(s) and selection of (a) new versus second-hand car(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, choice of car ownership status, fuel type and car state (new versus second-hand). 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, the type of ownership, the fuel type and the car state).

3.2 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[1].

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

- The *age* $y_{c,t}$ of car c at year t. We set an upper bound for the age $\bar{Y}$, assuming that above this upped bound, changes in age do not affect the utility or transition from one state to another. This implies that we have $y_{c,t} \in Y = \{0,1,\dots,\bar{Y}\}$.

- The *car ownership status* $I_{c,t}$. It consists of a discrete variable indicating whether car c is owned privately (level 1), by sole proprietorship[2] (level 2) or by another type of company (level 3) at year t. We have $I_{c,t} \in I_C = \{0,1,2,3\}$, where level 0 indicates the absence of car c.

- The *fuel type* $f_{c,t}$ of car c at year t. A car c can have one of the three following fuel types $f_{c,t}$: petrol, diesel or other fuel types (flexi fuel ethanol, CNG, hybrid, plug-in hybrid and electric car), denoted by 1, 2 and 3, respectively. Therefore we have $f_{c,t} \in F = \{0,1,2,3\}$, where level 0 indicates the absence of a car.

Each state $s_t \in S$ can hence be represented as

$$s_t = (y_{1,t}, I_{1,t}, f_{1,t}, y_{2,t}, I_{2,t}, f_{2,t}). \tag{1}$$

Due to the fact that we only have information about the age of the car and its fuel type for privately-owned cars and cars owned by sole proprietorship, we do not represent age and fuel type

[1]All the variables defined in Sections 3.2, 3.3 and 3.4 are household-specific. However, to simplify the notation, we omit the household-specific index n.

[2]By *sole proprietorship*, we mean a form of business that legally has no separate existence from its owner (Source: http://www.entrepreneur.com).

for company cars. Therefore, if we have $I_{c,t} = 3$, then this implies that we also have $y_{c,t} = 0$ and $f_{c,t} = 0$, respectively.

For households who have access to company cars, the size of the state space can be computed as

$$
\begin{aligned}
|S| &= (|Y| \times (|I_C| - 2) \times (|F| - 1) + 1)^2 && (2)\\
&+ (|Y| \times (|I_C| - 2) \times (|F| - 1) + 1) && (3)\\
&+ 1. && (4)
\end{aligned}
$$

The first term (2) consists of the number of possible states for two-car households. The element $|Y| \times (|I_C| - 2) \times (|F| - 1)$ of (2) is the number of states for housholds with privately owned cars or cars owned by sole proprietorship, while the element 1 of (2) is the number of states for households with company cars. For these households, we indeed only have the information of whether a company car is chosen or not. The exponent 2 stands for the two cars in the household. The second term (3) is the number of possible states for one-car housholds and the last term (4) stands for the absence of cars in a household. Assuming that cars can be at maximum 10 years old and given the above definitions of I_C and F, the size of the state space reaches the reasonable size of $3'783$. It is important to keep the size as low as possible since we need to solve the DP problem repeatedly when estimating the model parameters.

We note that not all households have access to company cars and some states of S are then unavailable.

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. 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 2) leads to nine possible transactions. Therefore we have $h_t \in H = \{1, \ldots, 9\}$.

- The *annual mileage* $\tilde{m}_{c,t} \in \mathbb{R}^+$ for each car c.

- The choice $\tilde{I}_{c,t} \in I_C$ of *car ownership status*.

- The *fuel type* $\tilde{f}_{c,t} \in F$.

- The *car state* $\tilde{r}_{c,t}$, i.e. the decision to purchase a new or second-hand car. We hence have $\tilde{r}_{c,t} \in R = \{0, 1, 2\}$, where level 0 means that no car has been bought, level 1 means that car c is bought new and level 2 means that car c is bought second-hand.

Note that if $I_{c,t} = 1, 2$, then we model the choice of mileage $\tilde{m}_{c,t}$, fuel type $\tilde{f}_{c,t}$ and car state $\tilde{r}_{c,t}$ for a car c for the next year. However, for $I_{c,t} = 3$, no information on the choice of mileage, fuel type or state of the car (new or second-hand) for the next year is available from the data. Therefore we do not model such decisions.

Each action $a_t \in A$ can be represented as

$$
a_t = (h_t, \tilde{m}_{1,t}, \tilde{I}_{1,t}, \tilde{f}_{1,t}, \tilde{r}_{1,t}, \tilde{m}_{2,t}, \tilde{I}_{2,t}, \tilde{f}_{2,t}, \tilde{r}_{2,t}). \tag{5}
$$

It is worth noting that we have a completely discrete state space, while the action space is discrete-continuous. From some particular states s_t, not all actions are available. Hence, we implicitly have $a_t \in A(s_t)$ and the total number of discrete actions must be obtained by enumerating all possible actions from each particular state. Table 1 summarizes the number of discrete actions that can be attained for households with 0, 1 or 2 cars, depending on the type of transaction which

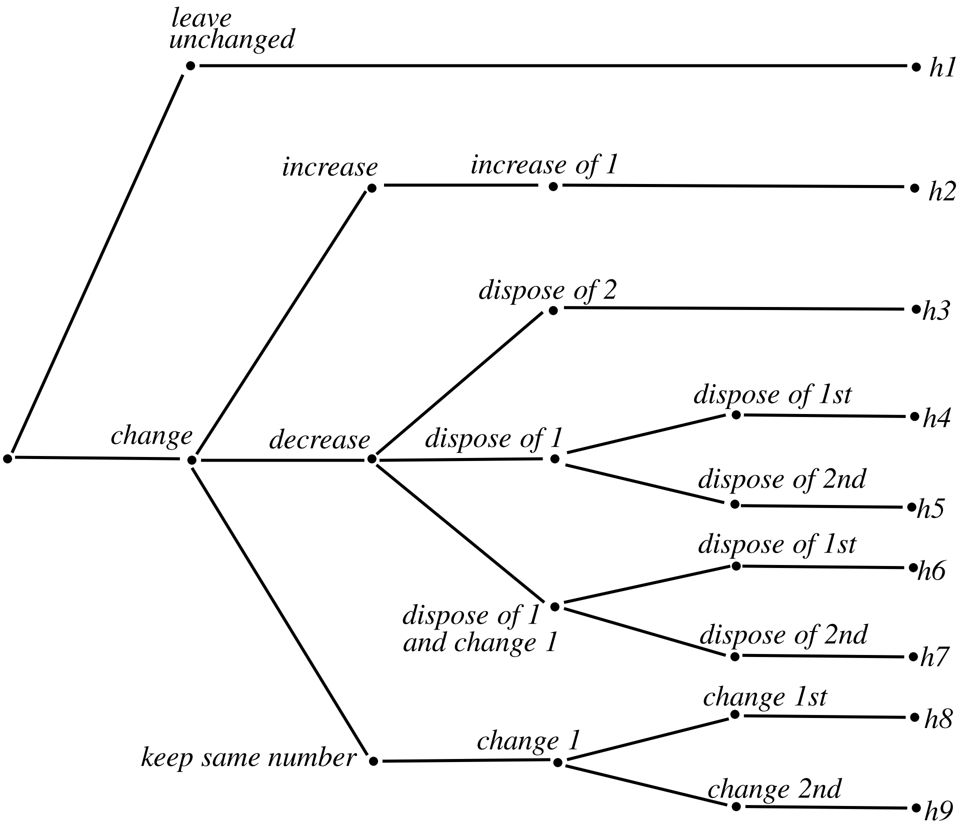


Figure 2: The nine possible transactions in a household fleet

is chosen. For example, a 1-car household that decides to increase the fleet of 1 car has the choice between 3 types of car ownership status, 3 types of fuel and 2 types of car state, leading to 18 possible actions. In the row 'Sum', the total number of possible discrete actions for households with respectively 0, 1 or 2 cars are reported.

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 inferred deterministically from s_t and a_t. Function $f(s_{t+1}|s_t,a_t)$ is hence defined as follows:

$$f(s_{t+1}|s_t,a_t) = \begin{cases} 1 & \text{if } s_t \text{ and } a_t \text{ lead to state } s_{t+1} \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

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 an action a_t and $a_t^C = (\tilde{m}_{1,t}, \tilde{m}_{2,t})$ gathers the continuous components, the *instantaneous utility* is defined as:

$$u(s_t, a_t^C, a_t^D, x_t, \theta) = v(s_t, a_t^C, a_t^D, x_t, \varepsilon_C(a_t^C), \theta) + \varepsilon_D(a_t^D), \tag{7}$$

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

Transaction name	0 car	1 car	2 cars
h_1: leave unchanged	1	1	1
h_2: increase 1	18	18	-
h_3: dispose 2	-	-	1
h_4: dispose 1st	-	1	1
h_5: dispose 2nd	-	-	1
h_6: dispose 1st and change 2nd	-	-	18
h_7: dispose 2nd and change 1st	-	-	18
h_8: change 1st	-	18	18
h_9: change 2nd	-	-	18
Sum	19	38	76

Table 1: Number of possible actions for households with 0, 1 or 2 cars (in the action space generated by the discrete components of the choice variable).

3.3 Value functions

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

$$\begin{aligned} V(s_t, x_t, \theta) &= \max_{a_t \in A} \{u(s_t, a_t, x_t, \theta) + \beta \sum_{s_{t+1} \in S} \bar{V}(s_{t+1}, x_{t+1}, \theta) f(s_{t+1}|s_t, a_t)\} & (8) \\ &= \max_{a_t \in A} \{v(s_t, a_t^C, a_t^D, x_t, \varepsilon_C(a_t^C), \theta) + \varepsilon_D(a_t^D) + \beta \sum_{s_{t+1} \in S} \bar{V}(s_{t+1}, x_{t+1}, \theta) f(s_{t+1}|s_t, a_t)\} & (9) \end{aligned}$$

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

$$\bar{V}(s_t, x_t, \theta) = \int V(s_t, x_t, \theta) dG_{\varepsilon_D}(\varepsilon_D(a_t^D)) \tag{10}$$

where G_{ε_D} is the CDF of ε_D.

In the case where all actions are discrete and the random terms $\varepsilon_D(a_t^D)$ 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[3].

Under the hypothesis of myopicity of the choice of annual driving distance(s), the integrated

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

value function is obtained as follows:

$$
\begin{aligned}
\bar{V}(s_t,x_t,\theta) &= \int V(s_t,x_t,\theta)dG_\varepsilon(\varepsilon_D(a_t^D)) \\
&= \int \max_{a_t\in A}\{u(s_t,a_t,x_t,\theta,\varepsilon(a_t))+\beta\sum_{s_{t+1}\in S}\bar{V}(s_{t+1},x_{t+1},\theta)f(s_{t+1}|s_t,a_t)\}dG_\varepsilon(\varepsilon_D(a_t^D)) \\
&= \int \max_{a_t^D}\{\max_{a_t^C}\{v(s_t,a_t^C,a_t^D,x_t,\varepsilon_C(a_t^C),\theta)\}+\varepsilon_D(a_t^D)+\beta\sum_{s_{t+1}\in S}\bar{V}(s_{t+1},x_{t+1},\theta)f(s_{t+1}|s_t,a_t)\}dG_\varepsilon(\varepsilon_D(a_t^D)) \\
&= \log\sum_{a_t^D}\exp\{\max_{a_t^C}\{v(s_t,a_t^C,a_t^D,x_t,\varepsilon_C(a_t^C),\theta)\}+\beta\sum_{s_{t+1}\in S}\bar{V}(s_{t+1},x_{t+1},\theta)f(s_{t+1}|s_t,a_t)\} \qquad (11)
\end{aligned}
$$

Similarly as in the case of a DDCM, the value function is obtained by iterating on Equation (11). In our case, Equation (11) can be simplified as follows due the deterministic formulation of the transition.

$$\bar{V}(s_t,x_t,\theta)=\log\sum_{a_t^D}\exp\{\max_{a_t^C}\{v(s_t,a_t^C,a_t^D,x_t,\varepsilon_C(a_t^C),\theta)\}+\beta\bar{V}(s'_{t+1},x_{t+1},\theta)\}, \qquad (12)$$

where s'_{t+1} is the state deterministically reached from s_t if action a_t is chosen.

However we emphasize on the fact that Equation (11) is also applicable to case where the transition is stochastic.

3.4 Constant Elasticity of Substitution (CES) utility function

We assume that expression $v(s_t,a_t^C,a_t^D,x_t,\varepsilon_C(a_t^C),\theta)$ of Equation (11) is the sum of the utility of the discrete actions v_t^D and the utility of the continuous actions v_t^C:

$$v(s_t,a_t^C,a_t^D,x_t,\varepsilon_C(a_t^C),\theta)=v_t^D(s_t,a_t^D,x_t,\theta)+v_t^C(s_t,a_t^D,a_t^C,x_t,\varepsilon_C(a_t^C),\theta) \qquad (13)$$

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 $v(s_t,a_t^C,a_t^D,x_t,\varepsilon_C(a_t^C),\theta)$ of Equation (11) must hence be maximized with respect to the two annual driving distances. Given the additive form of Equation (13), we only need to maximize expression $v_t^C(s_t,a_t^D,a_t^C,x_t,\varepsilon_C(a_t^C),\theta)$ with respect to a_t^C.

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.

This motivates the use of a *constant elasticity of substitution (CES)* utility function for the choice of mileage(s), since it allows to evaluate the rate of substitution of mileages $m_{1,t}$ and $m_{2,t}$:

$$v_t^C(s_t,a_t^D,a_t^C,x_t,\varepsilon_C(a_t^C),\theta)=(m_{1,t}^{-\rho}+\alpha\cdot m_{2,t}^{-\rho})^{-1/\rho} \qquad (14)$$

Parameter ρ is the elasticity of substitution of v_t^C. Expression α represents the weight of the mileage of one car relative to the other. It is a function of socio-economic characteristics about the household x_t and a random term $\varepsilon_C(a_t^C)$:

$$\alpha:=\exp\{\gamma x_t-\varepsilon_C(a_t^C)\}. \qquad (15)$$

Here, γ is a vector of parameters to estimate. To simplify the problem, we will assume in a first stage that α is only a constant to estimate. This implies that we assume no error term $\varepsilon_C(a_t^C)$. We will later relax this assumption and assume that α has the form of Equation (15).

The choice of $m_{1,t}$ and $m_{2,t}$ must be made such that the budget constraint of the household holds:

$$p_{1,t}m_{1,t} + p_{2,t}m_{2,t} = \mathrm{Inc}_t, \tag{16}$$

where $p_{c,t} := \mathrm{cons}_{c,t} \cdot \mathrm{pl}_{c,t}$ is the cost per km of driving car $c \in \{1,2\}$ in SEK/km, that is the product of the car consumption $\mathrm{cons}_{c,t}$ and the price of a liter of fuel $\mathrm{pl}_{c,t}$ for that car. Variable Inc_t is the share of the household's annual income which is used for expenses related to car fueling.

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. We however note that we use the fairly restrictive assumption that all households allocate the same percentage of their incomes to fuel expenses, since we do not observe these particular expenses from the data.

The optimal value of mileage $m_{1,t}$ is obtained by solving the following maximization problem:

$$\max_{m_{1,t},m_{2,t}} v_t^C, \text{ such that } p_{1,t}m_{1,t} + p_{2,t}m_{2,t} = \mathrm{Inc}_t \tag{17}$$

Assuming that we know what share of the household's income is spent on fuel[4], we can obtain an analytical solution for $m_{2,t}$:

$$m_{2,t}^* = \frac{\mathrm{Inc}_t \cdot p_{2,t}^{(-1/(\rho+1))}}{p_{2,t}^{(\rho/(\rho+1))} + p_{1,t}^{(\rho/(1+\rho))}\alpha^{(-1/(\rho+1))}}. \tag{18}$$

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

$$\begin{aligned} m_{1,t}^* &= \frac{\mathrm{Inc}_t}{p_{1,t}} - \frac{p_{2,t}}{p_{1,t}} m_{2,t}^* && (19) \\ &= \frac{\mathrm{Inc}_t}{p_{1,t}} - \frac{p_{2,t}}{p_{1,t}} \cdot \frac{\mathrm{Inc}_t \cdot p_{2,t}^{(-1/(\rho+1))}}{p_{2,t}^{(\rho/(\rho+1))} + p_{1,t}^{(\rho/(1+\rho))}\alpha^{(-1/(\rho+1))}} && (20) \end{aligned}$$

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

$$v_t^{C*} = \left(\left(\frac{\mathrm{Inc}_t \cdot p_{2,t}^{(-1/(\rho+1))}}{p_{2,t}^{(\rho/(\rho+1))} + p_{1,t}^{(\rho/(1+\rho))}\alpha^{(-1/(\rho+1))}} \right)^{-\rho} + \alpha \cdot \left(\frac{\mathrm{Inc}_t}{p_{1,t}} - \frac{p_{2,t}}{p_{1,t}} \cdot \frac{\mathrm{Inc}_t \cdot p_{2,t}^{(-1/(\rho+1))}}{p_{2,t}^{(\rho/(\rho+1))} + p_{1,t}^{(\rho/(1+\rho))}\alpha^{(-1/(\rho+1))}} \right)^{-\rho} \right)^{-1/\rho} \tag{21}$$

Then v_t^{C*} can be inserted back in Equation (13). The Bellman equation (11) becomes:

$$\bar{V}(s_t, x_t, \theta) = \log \sum_{a_t^D} \left\{ \exp\{v_t^D(s_t, a_t^D, x_t, \theta) + v_t^{C*}(s_t, a_t^D, a_t^{C*}, x_t, \theta)\} + \beta \bar{V}(s'_{t+1}, x_{t+1}, \theta) \right\}, \tag{22}$$

where $a_t^{C*} = (m_{1,t}^*, m_{2,t}^*)$.

The integrated value function $\bar{V}$ can then be computed by value iteration.

[4]For example, from 2006 to 2009, households in Sweden spent between 7.3 and 8.1 percent of their income on the operation of motor-cars (Source: Statistics Sweden).

3.5 Maximum likelihood estimation

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

$$\mathscr{L}(\theta) = \prod_{n=1}^{N} \prod_{t=1}^{T_n} P(a^D_{n,t} | s_{n,t}, x_{n,t}, \theta), \tag{23}$$

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

$$P(a^D_{n,t}|s_{n,t}, x_{n,t}, \theta) = \frac{v^D_{n,t}(s_{n,t}, a^D_{n,t}, x_{n,t}, \theta) + v^{C*}_{n,t}(s_{n,t}, a^D_{n,t}, a^{C*}_{n,t}, x_{n,t}, \theta) + \beta \bar{V}(s'_{n,t+1}, x_{n,t+1}, \theta)}{\sum_{\tilde{a}^D_{n,t}} \left\{ v^D_{n,t}(s_{n,t}, \tilde{a}^D_{n,t}, x_{n,t}, \theta) + v^{C*}_{n,t}(s_{n,t}, \tilde{a}^D_{n,t}, \tilde{a^{C*}_{n,t}}, x_{n,t}, \theta) + \beta \bar{V}(s'_{n,t+1}, x_{n,t+1}, \theta) \right\}} \tag{24}$$

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 in a first stage.

4 Illustrative example

As described in the previous sections, the discrete-continuous choice model is embedded into a dynamic programming framework in order to account for the expected discounted utility of each action. In this section, we present an illustration of the results of the value iteration algorithm, for imposed parameter values.

4.1 Example of specification

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

$$v^D_t(s_t, a^D_t, x_t, \theta) = \mathrm{C}(s_t) + \tau(a^D_t) + \beta_{\mathrm{Age}}(a^D_t, s_t) \cdot \max(\mathrm{Age1}_t, \mathrm{Age2}_t), \tag{25}$$

where $\mathrm{Age1}_t$ is the age of the first of the two cars and $\mathrm{Age2}_t$ is the age of the second car. Expressions $\mathrm{C}(s_t)$, $\tau(a^D_t)$ and $\beta_{\mathrm{Age}}(a^D_t, s_t)$ are parameters that would typically be estimated on data. We follow the approach proposed by Schiraldi (2011) and specify a constant $\mathrm{C}(s_t)$ relative to households owning at least one car and a transaction cost $\tau(a^D_t)$. The constant is included in order to capture differences of preferences between households owning at least one car and households without a car. The transaction cost is meant to capture the unobserved costs (e.g. search cost) 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 (25). 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}. Assuming that owning at least one car has a positive impact on choice, we set C to 5 if the household owns at least a car and to 0 otherwise. The other chosen parameter values are reported in Table 2. 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	-1	-1	0
h_2: increase 1		0	0	-	-3
h_3: dispose 2		-	-	1	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 and change 2nd		-	-	0	-4
h_7: dispose 2nd and 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 2: 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. As described in Section 3.4, we assume that α is a constant and fix its value to 0.3. The elasticity of substitution ρ is set to 0.5. Moreover we fix the discount factor in the Bellman equation (22) to 0.7.

4.2 Results from the value iteration

Using the parameters of Section 4.1, we iterate on the Bellman equation (22) to obtain the value function for an example observation. The latter consists of a household with an annual income of $530'000$ SEK that spends about 8% of its income on fuel (following the hypothesis described in Section 3.4).

The program is implemented in C++ and the running time to obtain the value function is about 2 minutes on a 20-core computer.

Figure 3 shows boxplots of the value function for ages of car ranging from 0 to 3[5]. As expected we observe that the value function decreases as the maximum of the ages of the two cars increases. This shows that the older the car is, the smaller its expected discounted utility becomes.

[5] We note that in this evaluation of the value function, we restricted the maximum of the age of the two cars to 3 years for a better visualization of the difference of the results, but the upper bound for age can be increased.

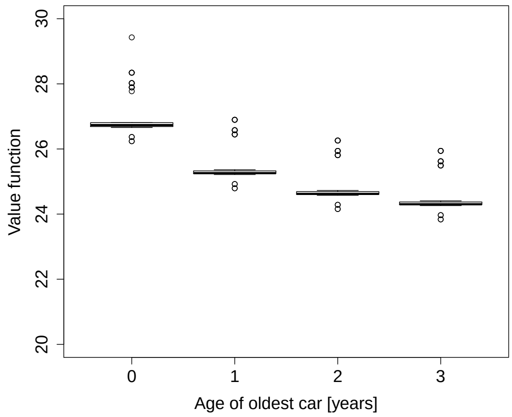
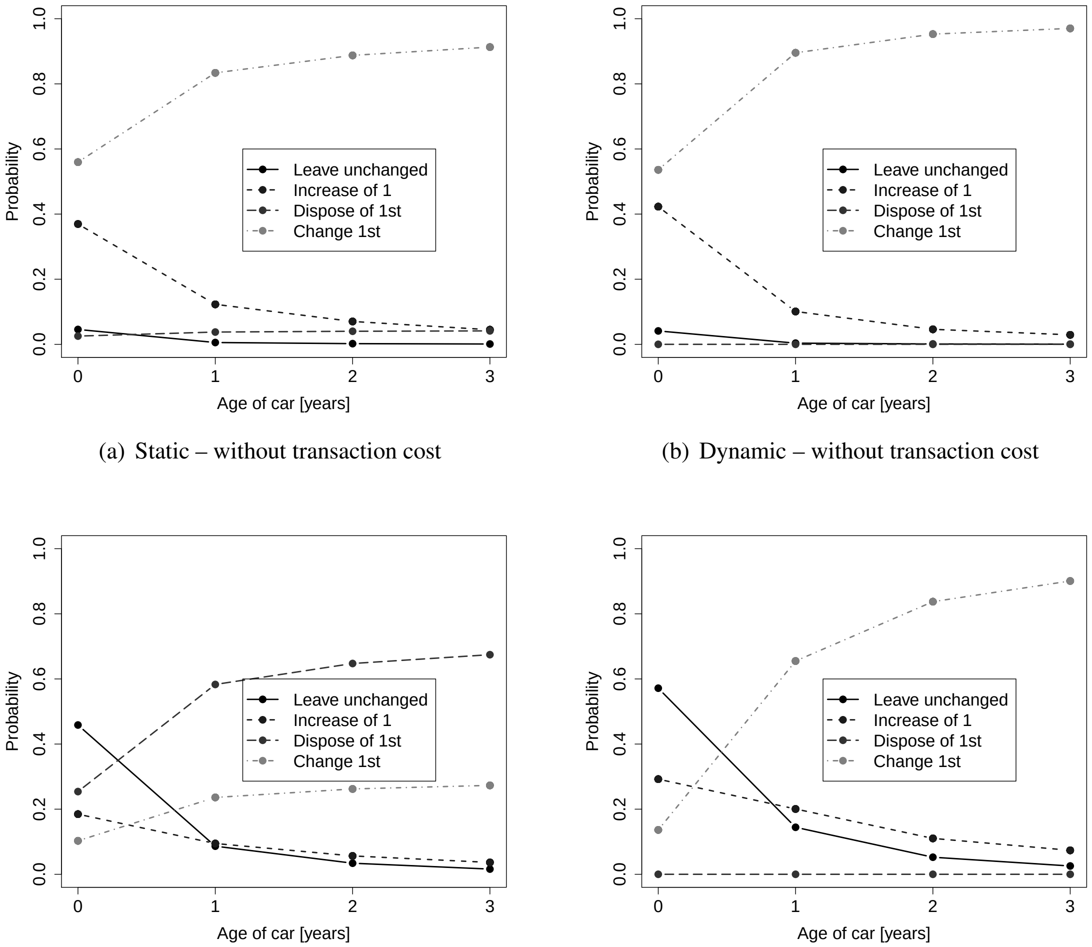


Figure 3: Value function

Probabilities of choosing a particular action can be computed using Equation (24). As an example, we evaluate these probabilities for the above mentioned household, assuming that it has one private diesel car. We analyze the probabilities of leaving the household car fleet unchanged, adding one car, disposing of the only car and changing the only car, as a function of the age of the car (Figure 4). For comparison purposes, we do not only report the probabilities for the above specified model (Figure 4(d)), but also for a model with the same specification but without transaction cost (Figure 4(b)), and for static models with (Figure 4(c)) and without (Figure 4(a)) transaction cost. By static models, we denote models based on the assumption that households do not account for the future utility of the choices of transaction, ownership status, fuel type and car state.



(a) Static – without transaction cost

(b) Dynamic – without transaction cost

(c) Static – with transaction cost

(d) Dynamic – with transaction cost

Figure 4: Choice probabilities for different transactions, as a function of the age of the car.

Figure 4 allows to analyze variations of particular probabilities as a function of the age of the car, e.g. the probability of replacing the only car in the household (denoted as changing 1st). As expected (from the imposed sign on the parameter) we note that the probability of changing the only car in the household increases as the age of the car increases.

Although we emphasize the fact that no behavioral interpretation can be made on the model at this point since all parameters have chosen values, we want to highlight that considering a dynamic model can result in different choice probabilities than in the static case and it is a key aspect to investigate. Moreover the introduction of a transaction cost also affects the trade-offs between the transactions.

5 Discussion of extensions

In this section, we discuss the limitations of the approach presented in this paper.

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 the model we presented in this paper, it is assumed that all individuals have access to company cars. In a later stage, this assumption should be relaxed, and more research should be performed in order to identify the individuals who can really decide whether to select a company car or not.

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

6 Conclusion

This paper presents a methodology designed to jointly model car ownership, usage and choice of fuel type together with an example of application. 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 car, since the purchase of a car affects the utility of an individual for the present and future years of ownership (Schiraldi, 2011).

In order to obtain a realistic model, we account for households' decisions rather than individual ones. We specify a CES utility function to capture substitution patterns that occur when two-car households decide on the annual mileages of the two cars. We also 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.

The next steps in this research are (1) to validate the DDCCM by estimating it on synthetic data generated from distributions of attributes of observations in the Swedish register of cars and individuals and (2) to estimate it on the full register data. In a later stage, we will assess the impact of policies implemented during the years of the data on the dynamics of the Swedish fleet and perform a forecasting analysis of several policy scenarios that have already 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 and the third International Choice Modelling Conference. The work presented in this paper has greatly benefited from comments received from the participants of these conferences.

References

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

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

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

De Jong, G., Fox, J., Daly, A., Pieters, M. and Smit, R. (2004). Comparison of car ownership models, *Transport Reviews* **24**(4): 379–408.

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

Hugosson, M. B. and Algers, S. (2012). Accelerated introduction of 'clean' cars in sweden, *in* T. I. Zachariadis (ed.), *Cars and Carbon*, Springer Netherlands, pp. 247–268.

Kageson, P. (2013). Dieselization in Sweden, *Energy Policy* **54**(0): 42 – 46.

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

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

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

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

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.