# Explaining household decisions on car ownership and use using an approach based on an indirect utility function

## Abstract

In this paper, we present a model that can be viewed as an extension of the traditional Tobit model. In contrast to that specific model, our model also accounts for the fixed costs of car ownership. This extension is required because being carless is an option for many households in societies that have good public transportation systems, the main reason being that carless households wish to save the fixed costs of car ownership. As yet, no existing model is able to map the impact of these fixed costs on car ownership adequately. By using the modelling framework, we are able to overcome this limitation. Our model is based on an indirect utility function corresponding to a log-linear Marshallian demand function; it includes the fixed costs of car ownership. By using this model, we are able to evaluate the effect of policies intended to influence household behaviour with respect to car ownership *and* use, which can be of great interest to policy makers. Our model enables us to compute the effect of policies such as taxes on fuel or on car ownership on both the share of carless households and the average driving distance. We calibrated the model using data on Swiss private households in order to be able to forecast responses to policies. We will compare the results with those of a model based on a direct utility function calibrated using the same data.

Dr. Reto Tanner
Transportbeleid en Logistieke Organisatie
Faculteit Techniek Bestuur en Management
Jaffalaan 5
2628 BX Delft
Netherlands
+31 (0)15 278 20 15
R.Tanner-1@tudelft.nl /
retanner@gmx.ch

Prof. Dr. Gerard de Jong
Institute for Transport Studies
University of Leeds
34-40 University Road
LEEDS
LS2 9JT
+44 (0)113 34 36610 (direct)
+44 (0)113 34 35357 (secretary's office)
G.C.Dejong@its.leeds.ac.uk

## Introduction

It is of great interest to policy makers to be able to examine the effect of policies intended to influence household behaviour with respect to car ownership and use. So far, several attempts have been made to model household decisions on car ownership and use in order to simulate the outcome of such policies. Unfortunately, all existing modelling techniques have significant drawbacks. The purpose of this new modelling approach is to overcome some of these drawbacks.

Overviews of models for car ownership can be found in de Jong et al. (2004) and Anowar et al. (2014). Both of these reviews also contain a section on models that simultaneously explain car ownership and car use as a discrete-continuous problem.

The drawbacks of existing modelling techniques can be summarised as follows: the OLS fails to map carless households. The Tobit model is unable to map the impact of fixed costs. The sample selection model fails due to the lack of an instrumental variable: there is no variable that influences only the choice of whether or not to own a car whilst not influencing the demand for driving at the same time. An interesting candidate for solving this problem is the Discrete-Continuous Choice model, introduced by Dubin and McFadden (1984). This model can be used to explore the ownership of certain car types and their use. Unfortunately, the model only allows the choice of being carless to be captured if the annual mileage travelled using public transport is given in the dataset. Since this information is not available in most micro-census datasets, this model cannot be applied. Another interesting candidate is the Multiple Discrete-Continuous Extreme Value Model (MDCEV), as introduced by Bhat (2005). This model consists of a direct utility function and a budget restriction. It is assumed that it maps the utility maximisation process of a household, and is based on the assumption that a household chooses certain amounts of goods from a set of goods of different qualities in which the researcher may be interested, e.g. different types of wine. The model includes the possibility of a household choosing only the numeraire good, which is a basket of all the remaining goods apart from the goods of specific interest. Bhat and Sen (2006) adapted this particular modelling framework to the case where the different goods represent annual kilometres driven by different car types. There are two main drawbacks to their model: first, it ignores the fact that car ownership itself implies fixed costs. Second, it ignores the fact that car ownership and use have an impact on the household's remaining budget that can be spent on other goods. Instead, they assume that the household's total annual mileage is fixed and that it would simply choose how to split these kilometres amongst different car types. Both of these drawbacks violate the basic assumptions of microeconomic models.

De Jong (1990) applied a modelling framework as used in the paper by Dubin and McFadden (1984), which presented the so-called "Discrete-Continuous choice model" for the first time. This model captures a joint decision of deciding on one type of capital good and the intensity of using this capital good. Examples of such decisions are the choice of type of heating system and then the choice of room temperature that will define the energy costs, as examined by Dubin and McFadden (1984), or the choice of car type and the annual mileage driven. In contrast to their model, De Jong's model is able to capture the case where households decide not to use a car. De Jong used a modelling framework that contains just two goods, the numeraire good and driving distance by car. This implies that the model can capture car ownership and use only, but not the type of cars or the ownership and use of several cars. Note that, as in Dubin and McFadden, de Jong assumes a log-linear Marshallian demand function and an indirect utility function, which are linked by Roy's Theorem. Similarly to Dubin and McFadden, de Jong also captures the fixed costs of car ownership. Since de Jong adds an extra (second) random variable to the Marshallian demand function, which is not included in the derivation using Roy's identity, this model is not fully compatible with the basic assumptions of a microeconomic model either.

The purpose of our model is to overcome this drawback: we use the same modelling framework as de Jong (1990), but our model is consistent with microeconomic theory in that we do not add an extra random term.

In the following, we first describe our assumptions on household behaviour when they decide on their choice and use of a type of capital good. We then present the microeconomic demand system that maps this behaviour. Next, we will state the assumptions on the error term and will show how the parameters of this model can be estimated. We will then show how the elasticities of driving demand and car ownership can be computed. Finally, we present the empirical results using data on Swiss private households.

## 1. Introduction to the model

In this section, we present the microeconomic demand system, which should map household behaviour. The basic idea behind this model is that the household computes its utility for two cases: case (a), when it decides not to own a car and spends all its income on other goods, and case (b) when it decides to own a car and to drive a certain annual distance. The household will then choose the case that yields the higher utility.

We start by describing how the Marshallian demand function relates to the level of utility given case (b) where the household chooses to own a car and bear the fixed costs of car ownership. Given the choice of owning a car, it is assumed that a household chooses the annual driving distance $x_2$ that provides the highest utility, given its income $y$ net the fixed costs of car ownership $k_2$ and marginal driving costs $p_2$. Good one $x_1$ is a composite good containing all goods except for the capital good. The price of this composite good $p_1$ is regarded to be numeraire. Thus, utility $x_1$ can also be regarded as the utility of the remaining income after having paid for the expenditures for car ownership and its use, irrespective of whether this remaining income has been spent entirely on the composite good or whether it has been saved. Note that we assume that only car driving provides utility, not car ownership itself. We assume that the household decision corresponds to a microeconomic modelling framework that corresponds to a log-linear Marshallian demand function (2a):[1]

$$V_b = v\left(p_2, y-k_2, \alpha, \beta, \gamma s, \varepsilon_c\right) = -\frac{1}{\alpha}\cdot e^{\gamma s+\alpha p_2+\varepsilon} + \frac{1}{1-\beta}\cdot\left(y-k_2\right)^{1-\beta}, \qquad (1)$$

$$X_2 = x_2\left(p_2, y-k_2, \alpha, \beta, \gamma s, \varepsilon_c\right) = e^{\gamma s+\alpha p_2+\varepsilon}\cdot\left(y-k_2\right)^{\beta}. \qquad (2)$$

The Marshallian demand function can also be written in natural logarithms:

$$\ln\left(X_2\right) = \alpha p_2 + \beta\cdot\ln\left(y-k_2\right) + \gamma s + \varepsilon. \qquad (2a)$$

Note that the indirect utility function (1) and the Marshallian demand function (2) are linked by Roy's identity, and that there is only a very limited set of Marshallian demand functions for which the corresponding indirect utility function is of a known and explicit form, enabling quick computation.[2] Here, $s$ reflects socio-demographic variables of the household. The random variable $\varepsilon$ contains unobserved socio-demographic variables, with parameter vector $\gamma$. Relevant unobserved household attributes could be the preference for car driving or a disability that prevents one member of the household from using public transportation. The Marshallian demand function (2) describes which driving distance the household would choose in case (b) and which utility level (1) it would reach in that case.

In the alternative case (a), the household chooses not to own a car. In this case, the complete income $y$ is available to the household and the demand for car driving is zero by definition. The utility level of this case cannot be computed straightforwardly, since the direct utility function is unknown. Thus, we first have to compute the marginal cost of driving demand $p_2$ that corresponds to a Marshallian demand of driving (2) of zero for the case where the household's budget is equal to its total income $y$. We then have to plug this value into the indirect utility function (1):[3]

$$V_a = \frac{1}{1-\beta}\cdot y^{1-\beta}. \qquad (3)$$

---

[1] In fact, formulas (1) and (2) only hold if $\varepsilon > \varepsilon_0$, where $x_2\left(p_2, y, \alpha, \beta, \gamma, \varepsilon_0\right) = 0$. Later we will show that it is not necessary to consider the case $\varepsilon < \varepsilon_0$, and thus we do not show what (1) and (2) would be in that case.

[2] Roy's identity is defined as follows: $x_2\left(p_1, p_2, y\right) = -\partial v\left(p_1, p_2, y\right)/\partial p_2 / \partial v_i\left(p_1, p_2, y\right)/\partial y$. Applying Roy's identity to (1) yields: $\partial v\left(p_1, p_2, y\right)/\partial p_2 = -\alpha\cdot e^{\gamma s+\varepsilon+\alpha p_2}$, $\partial v\left(p_1, p_2, y\right)/\partial y = y^{-\beta}$ and thus $x_2\left(p_1, p_2, y\right) = \alpha\cdot e^{\gamma s+\varepsilon+\alpha p_2}\cdot y^{\beta}$.

[3] In the first step, I set (2) to zero: $x_2\left(p_2, y, \alpha, \beta, \gamma s, \varepsilon_c\right) = 0 \Leftrightarrow e^{\gamma s+\varepsilon+\alpha p_2}\cdot y^{\beta} = 0$. Note that the fixed costs are now zero, $k_2 = 0$, since the household no longer owns a car. In this case, the Marshallian demand function is zero when $p_2$ goes to infinity: $p_2 \to \infty$. Note that $\alpha$ is negative, since the impact of the price on driving distance is negative.

$$V_a = v\left(\infty, y, \alpha, \beta, \gamma s, \varepsilon_c\right) = \frac{1}{1-\beta}\cdot y^{1-\beta}.$$

The household will now decide to be carless if $V_a > V_b$ and the random variable $X_2$ is defined as follows:

$$X_2 = \begin{cases} V_a < V_b : 0 \\ V_a \geq V_b : e^{\gamma s + \varepsilon + \alpha p_2} \cdot \left(y - k_2\right)^{\beta} \end{cases}, \tag{4}$$

where $V_a$ and $V_b$ are defined in (1) and (3) and we choose $\varepsilon$ to be normally distributed with zero mean and standard deviation $\sigma$. From this follows the following probability function of $X_2$:

$$f\left(z\right) = f\left(z, y, p_2, e_c, \theta\right) = \begin{cases} z = 0 : \Phi\left(\dfrac{e_c}{\sigma}\right) \\ 0 < z < x_{2,c} : 0 \\ z \geq x_{2,c} : \dfrac{1}{z} \cdot \dfrac{1}{\sigma} \cdot \phi\left(\dfrac{\ln\left(z\right) - \alpha p_2 - \beta \cdot \ln\left(y - k_2\right) - \gamma s}{\sigma}\right) \end{cases} \tag{5}$$

with $\theta = \left(\alpha, \beta, \delta, \sigma\right)$,

where $P_a\left(y, k_2, p_2, \alpha, \beta, \gamma\right) = P\left(V_a > V_b\right) = P\left(X_2 < x_{2,c}\right) = P\left(\varepsilon < e_c\right) = \Phi\left(\dfrac{e_c}{\sigma}\right)$, (5a)[4]

and $e_c = \ln\left(\dfrac{\alpha}{1-\beta} \cdot \left(\left(y - k_2\right)^{1-\beta} - y^{1-\beta}\right)\right) - \alpha p_2 - \gamma s$,[5]

and $x_{2,c} = -\dfrac{\alpha}{1-\beta} \cdot \left(y - k_2\right)^{\beta} \cdot \left(y^{1-\beta} - \left(y - k_2\right)^{1-\beta}\right)$.[6]

$F\left(\bullet\right)$ is the cumulated density function corresponding to the random variable $\varepsilon$, e.g. $F\left(e\right) = \Phi\left(e/\sigma\right)$ for the case where $\varepsilon$ is normally distributed with zero mean and standard deviation $\sigma$, where $\Phi\left(\bullet\right)$ is the cumulative distribution function of the standard normal distribution.

A very nice feature of this model is that both $e_c$ and $x_{2,c}$ can be computed explicitly. This enables very fast computation, opening the door to many extensions of the model when estimations are based on Simulated Maximum Likelihood Estimation.

Another pleasing and useful result is that the expectation value can also be computed explicitly. This is particularly useful when simulating the effects of changes in parameters or economic variables.

[4] The relation $V_a > V_b \Leftrightarrow \varepsilon < e_c$ follows straightforwardly from (2a).

[5] This can be computed by solving $v_a\left(...\right) = v_b\left(...\right)$ for $\varepsilon$.

[6] This follows from plugging in $e_{c=...}$ into (2).

$$E(X_2) = \exp\left(-\frac{1}{2}\cdot\frac{m^2-(m+\sigma^2)^2}{\sigma^2}\right)\cdot\left(1-\Phi\left(\frac{\ln(x_{2c})-(m+\sigma^2)}{\sigma}\right)\right) \tag{6}$$

with $m = \alpha p_2 + \beta\cdot\ln(y-k_2)+\gamma s$.

## 2. Estimation of the model parameters

The parameters will be estimated such that the model explains the real word data as effectively as possible. To this end, the maximum likelihood estimation (MLE) method will be applied. Unfortunately, the fact that, according to this model, there is no observation $x_2$ in the interval $0 < x_2 < x_{2,c}$, such observations would cause a problem when applying the MLE method. This is because if any observation $x_2$ is in the interval $0 < x_2 < x_{2,c}$, the MLE function will be zero and thus, in such cases, the MLE function cannot be maximised. Note that, in real data as displayed the observations $x_2 > 0$ in the histogram in Figure 1 below, there are always some observations in the interval $0 < x_2 < x_{2,c}$ due to households misreporting or if they simply have an unusual preference for owning a car but use it sparingly. The following figure shows that such households exist, but are very rare. This is reflected in the observation in the histogram in the bin (0 ... 5,000) and parts of the observations in the bin [5,000 ... 10,000).

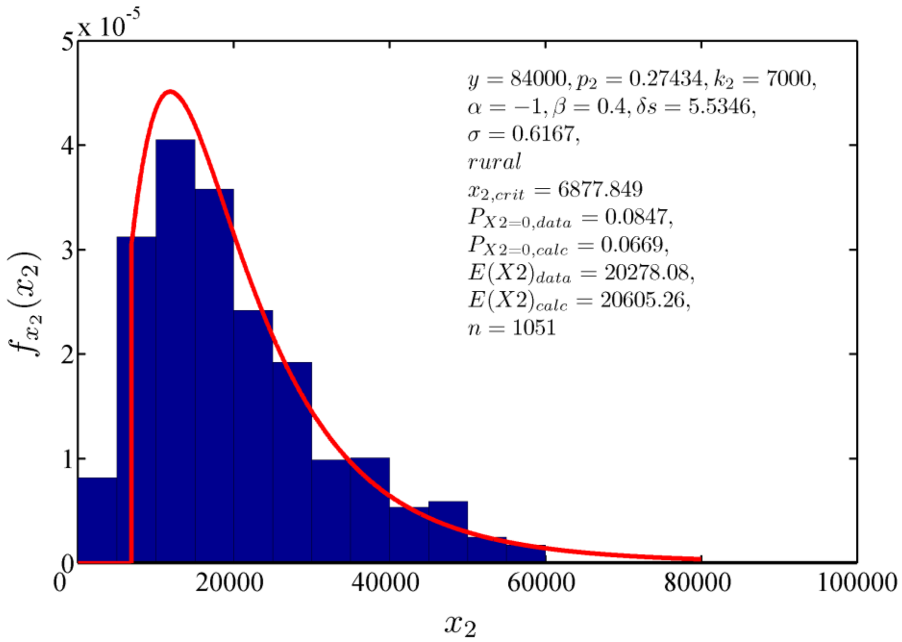


**Figure 1:** Empirical (histogram) and theoretical (red line) distribution for urban households with an income of CHF 108,000[7]

[7] Note that the height of the bars is normalised by factor $1/n$ so that the total surface of all bars equals one.

Formula (5a) shows that the minimum driving distance depends only on parameters $\alpha$ and $\beta$ and the fixed cost $k_2$ of car ownership. Thus the estimation of parameters $\alpha$ and $\beta$ plays a crucial role with respect to the minimum driving distance. Changing parameters $\alpha$ and $\beta$ has an impact on whether some observations $x_2$ fall in the interval $0 < x_2 < x_{2,c}$, which leads to an MLE function equal to zero, meaning that parameter $\gamma$ cannot be estimated. We circumvent this problem by applying the following estimation routine:

1. Choose values for $\alpha$ and $\beta$.
2. Compute $x_{2,c}$ for each observation $n$ $x_{2,c}$.
3. Eliminate all observations where $0 < x_2 < x_{2,c}$.
4. Estimate parameters $\delta$ and $\sigma$ by MLE conditional on $\alpha$ and $\beta$. Compute a penalty function that depends a) positively on the proportion of eliminated datasets, b) positively on the relative error of the difference between the average simulated proportion of carless households, c) positively on the actual proportion of carless households and d) on the difference between the average simulated expectation value of driving demand and the actual average driving distance. Note that the actual proportion of carless households and the actual average driving distance refer to the measures based on the dataset after eliminating the observations according to Step 3.
5. Repeat Steps 1 - 4 for a number of different values for $\alpha$ and $\beta$ (grid search). Choose values $\alpha$ and $\beta$ so that the lowest value of the penalty function is yielded.

For the MLE estimation, we use the following log ML function:

$$L(x, y, p_2, s, \theta) = \sum_{i=1}^{n} \ln\left(f(x_i)\right) =$$

$$= \sum_{i=1}^{n} I(x_i = 0) \cdot \ln\left(\Phi\left(\frac{e_{c,i}}{\sigma}\right)\right) + I(x_i > 0) \cdot \ln\left(\frac{1}{z} \cdot \frac{1}{\sigma} \cdot \phi\left(\frac{\ln(z) - \alpha p_2 - \beta \cdot \ln(y - k_2) - \gamma s}{\sigma}\right)\right), \tag{7}$$

where $e_{c,i}$ and $\theta$ are defined in (5).

As the penalty function we chose

$$\begin{aligned} Q = c_1 \cdot \frac{1}{N} \cdot \sum_{i \in yc} \sum_{j \in rc} \left( n_{i,j} \cdot \frac{P_{sim,i,j}\left(X_2 = 0 \mid yc, rc\right) - \text{mean}\left(x_2 = 0 \mid yc, rc\right)}{P_{sim,i,j}\left(X_2 = 0 \mid yc, rc\right)} \right)^2 + \ldots \\ \ldots + c_2 \cdot \frac{1}{N} \cdot \sum_{i \in yc} \sum_{j \in rc} \left( n_{i,j} \cdot \frac{E_{sim,i,j}\left(X_2 \mid yc, rc\right) - \text{mean}\left(x_2 \mid yc, rc\right)}{E_{sim,i,j}\left(X_2 \mid yc, rc\right)} \right)^2 + \ldots \\ \ldots + \left( \frac{\#\text{ elim. observations}}{\text{size of initial datasets}} \right)^2, \end{aligned} \tag{8}$$

where *rc* and *yc* are the income and rural/urban segments of the data, $rc = \{"urban","rural"\}$ and $yc = \{18,36,60,84,108,132,156,180,228\}$ in CHF 1000 per year and $n_{i,j}$ denotes the number of observations within an income/(urban, rural) segment. $P_{sim}$ is the average of the simulated probabilities, $P_{real}$ is the actual proportion of carless households in the dataset, $E_{sim}(X_2)$ is the average of the simulated expectation values of driving distance and $\text{mean}(x_2)$ is the mean of the actual driving distance in the dataset. Expressions "$\left(E_{sim}(X_2) - \text{mean}(x_2)\right)/\text{mean}(x_2)$" and "$(P_{sim} - P_{real})/P_{real}$" are the relative errors of the average of the simulated values, which could be called "replication errors". Note that we use the simulated values $P_{sim,i,j}(X_2 = 0 \mid yc, rc)$ and $E_{sim,i,j}(X_2 \mid yc, rc)$ in the denominator for weighting differences between the simulated and actual values of the individual segments. We do this because some segments contain only a small number of observations. Only a very small share of carless households could therefore be contained in one of these segments, which would lead to a very high weight of the difference between the simulated and actual values. Since this could have a strong impact on the penalty function, this could distort the entire optimisation process. In contrast, the simulated values do not exhibit this problem, since their behaviour is rather "smooth" when income increases. In order to reduce this problem further, we set the empirical value of the proportion of carless households to 0.05. Here "dataset" relates to the dataset after eliminating the dataset from observations where $0 < x_2 < x_{2,c}$. Expression "$\#\,\text{elim. observations/size of initial datasets}$" corresponds to the percentage of eliminated datasets with respect to the initial number of datasets. Parameters $c_1$ and $c_2$ are weighting parameters. We chose $c_1 = 0.5$, which means that both types of replication errors should be weighted about equally, and $c_2 = 0.5$.

## 3. Data

The data we used to estimate the parameters is the micro-census data on the travel behaviour of Swiss households, Swiss Federal Statistical Office SFSO (2006). 33,000 households were interviewed. The dates of when the interviews were conducted were more or less distributed evenly over the year 2005. This dataset contains a vast number of information on travel behaviour, ownership of cars, motorbikes and bicycles, and information on the households. Since the purpose of this study is to investigate fuel demand, we will use the information on total kilometres driven by cars. Since in the present model we do not consider the choice of different car types, we will use the total annual kilometres driven by all households as a proxy for fuel demand. Since we are basically interested in the effect of fuel prices on the distance travelled and the decision of whether or not to own one or several cars, we will only use the household variables that appeared to have the most important impact on travel distance or fuel demand in other models. In this case, we will only use the income and the place of residence as explanatory variables, namely whether the households live in a rural area or in a non-rural area, which we denote as "urban areas". As in Bhat (2008)[8], we choose the price of the composite good $x_1$ to be the numeraire.[9] Since $p_1$ is one, amount $x_1$ is nothing but income *y* minus the amount spent on driving, $k_2 + p_2 x_2$, since we assume that households spend all of their income and do not save anything. Of course, it is a simplification to assume that households will spend all of their income on consumption, but no data on savings is available in the dataset. Furthermore, savings can also be regarded as providing utility since having savings allows for future consumption and contributes to a positive feeling of having money set aside.

Note that, since our model captures only one car type, it is considered to be an "average car". The fixed costs of maintaining a car and the marginal costs of driving are thus assumed to be equal to those of an average car owned by a Swiss household. The values we retain for $k_2$ and $p_2$ were taken from the Swiss touring club TCS (2007) and comprise:[10]

$$k_2 = 7000 \text{ and } p_2 = 0.1601 + 0.0778 \cdot p_{fuel}, \text{ all units are in CHF.} \tag{9}$$

Note that the annual fixed costs $k_2$ mainly consist of depreciation, which is unrelated to the car's use, such as rusting, and loss in value due to the technical progress of new cars, capital costs, taxes on car ownership and parking costs. Since we neglect such costs as evaluation and registration costs, we assume that owning a car is similar to renting a car and that households can switch from owning a car to being carless without any cost. The costs dependent on the number of kilometres driven consist of fuel costs $0.0778 \cdot p_{fuel}$ and non-fuel-related variable costs such as the wear of tyres and mechanical components, which account for CHF 0.1601 per kilometre. The fuel price $p_{fuel}$ is the average fuel price from the last twelve months prior to interviewing the household to which the information on annual driving distance refers.[11]

---

[8] "If an outside good is present, label it as the first good which now has a unit price of one," Bhat (2008: 290). Note that Bhat denotes an "outside good" as a good that is always chosen.

[9] This is reasonable since the price of a composite good is a price index, and a price index is scale-free.

[10] According to TCS (2007), the total annual costs of an average car amounted to CHF 11,600 when the annual distance driven was 15,000 kilometres (km). 17.4% of these costs, namely CHF 2,018.4, were fuel costs. Based on the average fuel price paid for petrol 98 octane of CHF 1.729/litre in 2007 (SFSO 2009), it can be computed that the TCS (2007) based this fuel cost on a fuel consumption of 7.7825 litres/100 km: (CHF 2,018.4/15,000 km) / (CHF 1.729/litre) = 7.7825 litres/100 km. The fuel costs of an average car per kilometre are therefore 7.7825 litres/100 km/100 multiplied by the fuel price per litre paid by households. Non-fuel-related marginal costs of a car were calculated to be 20.7% of the total costs, 0.207 · CHF 22,600 = CHF 3,312, amounting to CHF 3,312/15,000 km = CHF 0.1601/km, see TCS (2007).

[11] The computation of $p_{fuel}$ is based on the monthly average price of petrol 98 octane, as published by the SFSO (2009a).

| | $\text{mean}(x=0)$ | $\text{mean}(x)$ | $\text{sdev}(x)$ | $\min(x)$ | $\max(x)$ |
|---|---|---|---|---|---|
| $x_2$ | 0.1890 | 13,890 | 12,195 | 0 | 59,731 |
| $y$ | -- | 80,187 | 43,373 | 18,000 | 228,000 |
| $k_2$ | -- | 7,000 | 0 | 7,000 | 7,000 |
| $p_2$ | -- | 0.2745 | 0.0036 | 0.2692 | 0.2838 |
| *rural* | 0.7717 | 0.2283 | 0.2745 | 0 | 1 |

**Table 1:** Summary statistics of the data, SFSO (2006) [12]

Note that observations of driving distance $x_2$ with more than 60,000 kilometres were eliminated, since these observations are considered to have been reported incorrectly. Note that these observations would strongly influence the MLE value and would therefore cause biases in the parameter estimation.

## 4. Results

We used a very simple specification where the dummy variable "rural" was the only socio-demographic variable. This is because we wanted to be able to split the data into the segments "rural"/"urban" for each income category. This will allow us to compare the probability function to the histogram of observations of driving distance $x_2$ for each such segment, giving us an intuitive idea of the model's fit to the data. For parameters *a* and *b,* we chose grids with ranges $\alpha = -2.5,..,-0,4$ and $\beta = 0.1,..,2.0$.[13] We computed the standard deviation of the parameters and the elasticities by the bootstrapping method. We randomly sampled 20 times each 5000 draws from the dataset and estimated the parameters and the elasticities.

[12] Note that these values refer to the complete dataset. In the estimation of the model given optimal values for *α* and *β*, some observations will be omitted. Around 11.8% of all observations will be omitted.

[13] The values of the grid were: $\alpha = (0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2.0, 2.2, 2.5)$ and $\beta = (0.1, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45, 0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.9, 0.95, 1.05, 1.1, 1.2, 1.3, 1.4, 1.5, 1.7, 2.0)$.

| | point estimate | sdev | t-value |
|---|---|---|---|
| $\alpha_{\text{opt}}$ | -1 | 0.020 | -49.796 |
| $\beta_{\text{opt}}$ | 0.4 | 0.014 | 29.582 |
| $\delta_0$ | 5.2367 | 0.150 | 34.853 |
| $\delta_{\text{rural}}$ | 0.29785 | 0.008 | 37.081 |
| $\sigma$ | 0.6167 | 0.006 | 102.929 |
| $N$ | 18408 | - | - |
| $(c_1,c_2)$ | $(0.5,0.5)$ | - | - |
| dropout rate | 11.8% | - | - |
| $(P_{sim}-P_{real})/P_{real}$ | -17.9% | - | - |
| $(E_{sim}(X_2)-\text{mean}(x_2))/\text{mean}(x_2)$ | -0.47% | - | - |
| Penalty value | 0.0299 | - | - |
| $\varepsilon_{E(X_2),y}$ | 0.517 | 0.018 | 28.520 |
| $\varepsilon_{E(X_2),p_2}$ | -0.322 | 0.007 | -47.715 |
| $\varepsilon_{E(X_2),p_{fuel}}$ | -0.135 | 0.003 | -47.715 |
| $\varepsilon_{E(X_2),k_2}$ | -0.222 | 0.008 | -26.597 |
| $\varepsilon_{P(X_2=0),y}$ | -0.936 | 0.025 | -37.897 |
| $\varepsilon_{P(X_2=0),p_2}$ | 0.599 | 0.014 | 41.677 |
| $\varepsilon_{P(X_2=0),p_{fuel}}$ | 0.252 | 0.006 | 41.677 |
| $\varepsilon_{P(X_2=0),k_2}$ | 2.249 | 0.023 | 97.380 |

**Table 2:** Elasticities and additional results based on the dataset SFSO (2006).[14]

Our model assumes no costs when switching from owning a car to being carless, and vice versa, and thus our elasticities can be interpreted as long-term elasticities. Our value for the fuel price elasticity (-0.14) of driving demand is clearly lower in absolute value than the average values determined in international studies (-0.29), such as in Goodwin et al. (2004). However, it is not much lower than those determined for Switzerland by Baranzini et al. (2009) for fuel demand using Swiss time series data (-0.20) or Axhausen and Erath (2010) using Stated preference data (-0.15). Note that the elasticities of fuel demand are higher than the elasticities of driving demand in absolute terms, since the part of the

[14] Note that the symbol "$\varepsilon$" stand for elasticities. The symbol "$\varepsilon_{E(X_2),y}$" stands for the relative change of the expectation value of the driving distance given a one percent increase in income, $\varepsilon_{E(X_2),y}=\partial E(X_2)/\partial y\cdot\text{mean}(y)/E(X_2)$. The symbol "$\varepsilon_{P,y}$" stands for the relative change of the probability of being carless given a one percent increase in income, $\varepsilon_{P(X_2=0),y}=\partial P(X_2=0)/\partial y\cdot\text{mean}(y)/P(X_2=0)$. Note that elasticities $\varepsilon_{P(X_2=0),p_{fuel}}$ and $\varepsilon_{E(X_2),p_{fuel}}$ were computed from $\varepsilon_{P(X_2=0),p_2}$ and $\varepsilon_{E(X_2),p_2}$ by multiplying by a factor of 0.42. This follows from $\varepsilon_{E(X_2),p_{fuel}}=\partial E(X_2)/\partial p_{fuel}\cdot p_{fuel}/E(X_2)=\partial E(X_2)/\partial p_2\cdot p_2/E(X_2)\cdot\partial p_2/\partial p_{fuel}\cdot p_{fuel}/p_2=\varepsilon_{E(X_2),p_2}\cdot\partial p_2/\partial p_{fuel}\cdot p_{fuel}/p_2$. The expression $\partial p_2/\partial p_{fuel}\cdot p_{fuel}/p_2$ yields $\partial p_2/\partial p_{fuel}\cdot p_{fuel}/p_2=0.0778\cdot((0.2745\text{-}0.1601)/0.0778)/0.2745=0.4167$, which follows from (9) using the average price $p_2=0.2745$ from Table 2.

reduction effect of fuel consumption is due to a shift towards more fuel-efficient vehicles. Goodwin et al. (2004) determined values (-0.64) versus (-0.29).

The income elasticity of aggregate driving distance we obtained (0.52) is also slightly lower than the average values established in international studies (0.73) by both Graham and Glaister (2005) and Goodwin et al. (2004) or Axhausen and Erath (2010) using Stated preference data (0.829) for total long-run demand for fuel and for total long-run demand for petrol only (0.627). Note that in this case, too, the elasticities relating to fuel demand are higher than those relating to distance driven, since households with a higher income tend to buy larger cars that consume more fuel.

The value $\varepsilon_{P(X_2>0),\,p_{fuel}} = -0.053$ can be computed from $\varepsilon_{P(X_2=0),\,p_{fuel}}$ , and reflects the decrease in percentage of households that own a car if the fuel price increases by one percent.[15] In our modelling world, which consists of only zero or one-car households, this reflects the elasticity of total cars in the economy with respect to fuel price. Our value is roughly in the range indicated by Johansson and Schipper (1997) (-0.2..0.0, best guess -0.1). In contrast, the value we obtain for the elasticity of total cars in the economy with respect to income is $\varepsilon_{P(X_2>0),\,y} = 0.200$, which is quite lower than the values found by Johansson and Schipper (1997) (0.75..1.25, best guess 1.00) and Dargay (2001) (0.74). We explain this difference by the fact that our elasticities refer to the case of “at least one car”, and that the income elasticity for buying a second or even a third car can be assumed to be greater since the latter can be considered a luxury good. In contrast, $\varepsilon_{P(X_2>0),k_2} = -0.480$ is quite larger in absolute terms compared to the values determined by Dargay (2001) (-0.26) and Johansson and Shipper (1997) (-0.127..-0.063) for the elasticity of the car stock with respect to the car’s fixed costs. In their model, this tax was imposed by a tax on car purchase. Annualising one unit of this tax yields an increase in the fixed costs of car ownership of about 2%, yielding a 0.6% decrease in car stock. Thus, a 1% increase in fixed costs would reduce the vehicle stock by 0.3%. However, it is also important to note that the results found in international studies for the elasticity of car ownership vary greatly and thus it is hard to judge whether the values a model yields are plausible. In our eyes, our value $\varepsilon_{P(X_2>0),k_2} = -0.481$ is plausible for Swiss data, since households may easily switch to public transportation in Switzerland, if driving costs increase.

Although our model consists of only five parameters, namely $\alpha$, $\beta$, $\delta_0$, $\delta_{\text{rural}}$ and $\sigma$, the model fits very well to the data. Since we used only the dummy variable “rural” and since income is quantified in nine

---

[15] $\varepsilon_{P(X_2>0),\,p_{fuel}} = \partial P(X_2>0)/\partial p_{fuel} \cdot p_{fuel}/P(X_2>0) = -\partial P(X_2=0)/\partial p_{fuel} \cdot p_{fuel}/P(X_2=0) \cdot P(X_2=0)/P(X_2>0) = \ldots$
$\ldots = \varepsilon_{P(X=0),\,p_{fuel}} \cdot (-1) \cdot P(X_2=0)/P(X_2>0)$, and $P(X_2>0) = 1 - P(X_2=0)$. We used the simulated values for $P(X_2=0)$ which we calculated from Tables 1 and 2. This results in $(-1) \cdot P(X_2=0)/P(X_2>0) = -0.2135$.

discrete values of $y = \{18, 36, 60, 84, 108, 132, 156, 180, 228\}$ in CHF 1000, we can check whether the density functions fit the eighteen individual segments well.

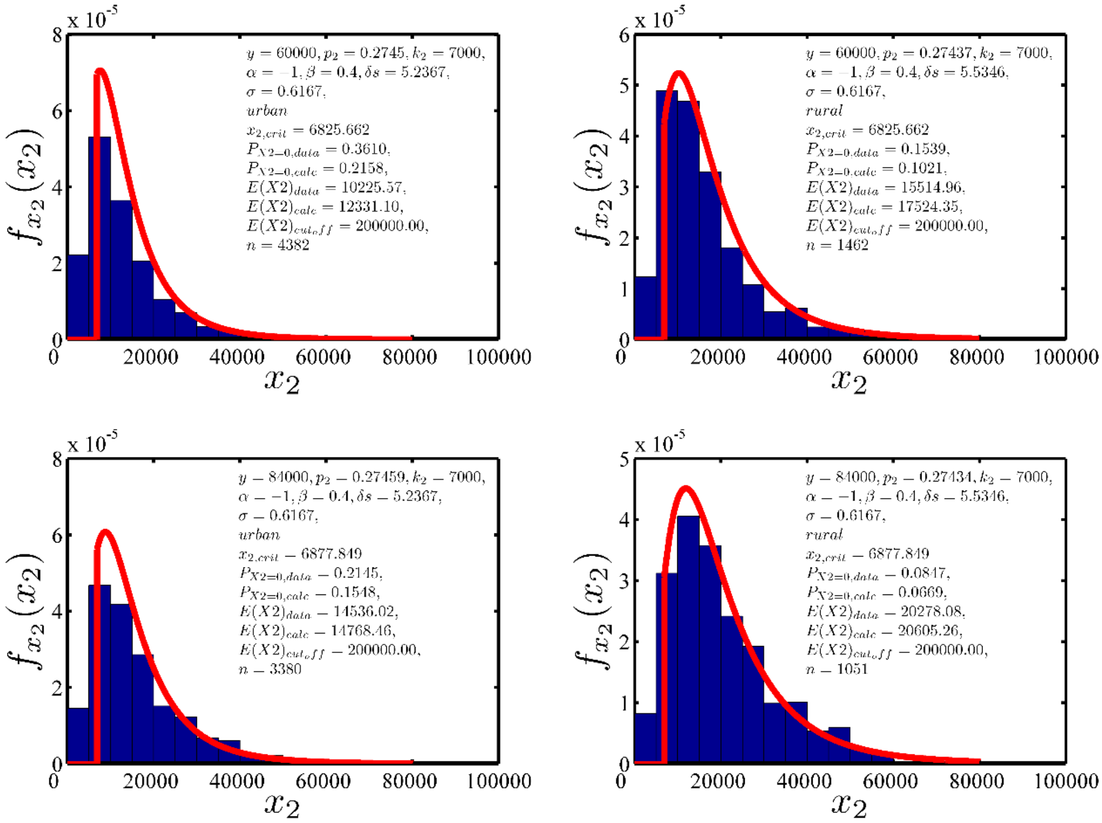

**Figure 2:** Model fit at an income level of CHF 60,000 and 84,000 to data, SFSO (2006)

These diagrams show that the model fits the data quite well. Note that these diagrams cover 10,275 observations, which is more than half of the total observations. In spite of the quite good fit of the model's density functions, there are differences between the empirical values and the model's probability of being carless and the expectation value of driving distance. These are summarised in the following diagram.

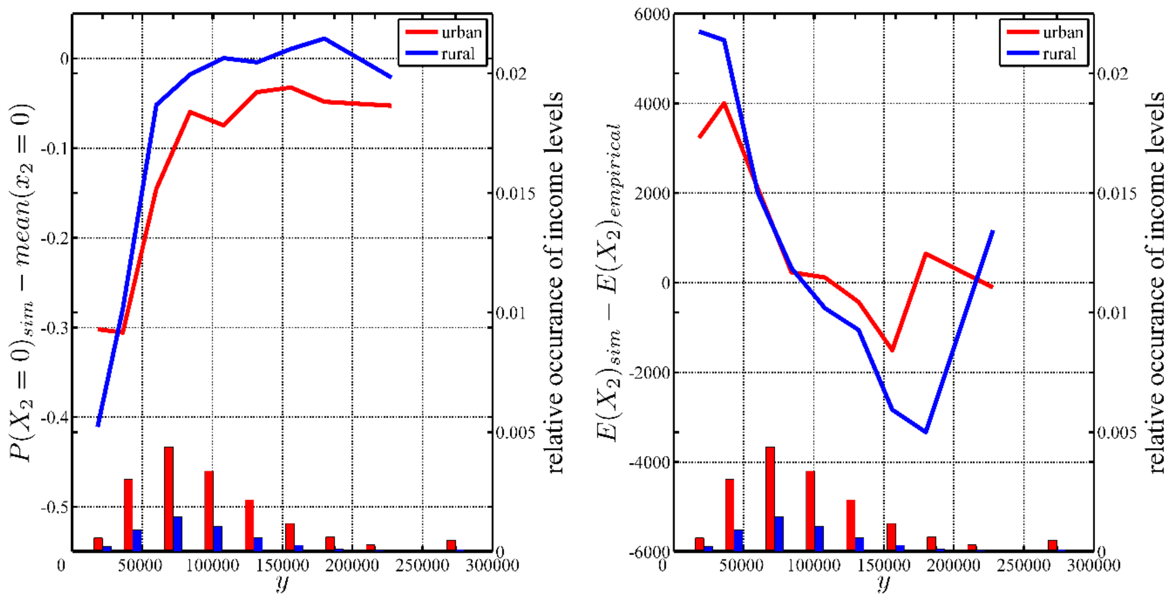

**Figure 3:** Model fit of all income levels, SFSO (2006)

These results show that in the range with the most observations, the difference between the model's expectation value and the empirical value of the corresponding income segments decreases. This implies that the model's elasticity with respect to income is too low, which might imply the difference between the model's elasticities and the elasticities found by other researchers.

An issue that could arbitrarily influence the results is the choice of parameters $c_1$ and $c_2$ of the penalty function. We therefore checked to see whether choosing different values for $c_1$ and $c_2$ would have a strong impact on the measures of interest, namely elasticities. In the first trial, we computed all results for the following combinations of $c_1$ and $c_2$: $(c_1, c_2) = \{(0.5, 1.0), (1.0, 0.5), (1.0, 1.0)\}$. In all these cases, the optimal values for $\alpha$ and $\beta$ were the same as in the benchmark case $(c_1, c_2) = (0.5, 0.5)$. The resulting optimal values $\alpha$ and $\beta$ and elasticities only change if we increase parameters $c_1$ and $c_2$ further, $(c_1, c_2) = \{(1.0, 2.0), (2.0, 1.0), (2.0, 2.0)\}$. In the most extreme cases $(c_1, c_2) = (2.0, 1.0)$ or $(c_1, c_2) = (2.0, 2.0)$, the measures of most interest, the elasticity of driving demand with respect to income and with respect to fuel price, increase by 14% and 42%, respectively, in absolute terms. However, since the dropout rate of observations in the dataset increases from 11.8% in the benchmark $(c_1, c_2) = (0.5, 0.5)$ to 18.5% in the cases $(c_1, c_2) = (2.0, 1.0)$ and $(c_1, c_2) = (2.0, 2.0)$, we do not consider choices $(c_1, c_2) = \{(1.0, 2.0), (2.0, 1.0), (2.0, 2.0)\}$ to be feasible, since the dropout rate is too high in our view.

Another interesting point to consider is comparing the results we obtain using this model with those yielded by the MDCEV of Tanner and Bolduc's (2014) model, which is based on a direct utility function.

| | Tanner and de Jong (2015) | Tanner and Bolduc (2014) | |
|---|---|---|---|
| Upper limit of integrating, $E(X_2)$ | $\infty$ | $\infty$ | 60,000 km |
| dropout rate | 11.8% | | 9% |
| $\varepsilon_{E(X_2),y}$ | 0.517 (0.018) | 1.17 (0.0031) | 0.75 (0.0069) |
| $\varepsilon_{E(X_2),p_2}$ | -0.322 (0.007) | − 1.07 (0.0069) | − 0.67 (0.0075) |
| $\varepsilon_{E(X_2),p_{fuel}}$ | -0.135 (0.003) | − 0.49 (0.0031) | − 0.28 (0.0031) |
| $\varepsilon_{E(X_2),k_2}$ | -0.222 (0.008) | − 0.16 (0.0030) | − 0.16 (0.0029) |
| $\varepsilon_{P(X_2=0),y}$ | -0.936 (0.025) | − 1.42 (0.0087) | − 1.41 (0.0081) |
| $\varepsilon_{P(X_2=0),p_2}$ | 0.599 (0.014) | 0.24 (0.0052) | 0.24 (0.0049) |
| $\varepsilon_{P(X_2=0),p_{fuel}}$ | 0.252 (0.006) | 0.14 (0.00005) | 0.11 (0.0028) |
| $\varepsilon_{P(X_2=0),k_2}$ | 2.249 (0.0023) | 1.27 (0.0102) | 1.33 (0.0111) |

**Table 3:** Elasticities compared to Tanner and Bolduc (2014)

Note that "60,000 km" means that, when computing the expectation value, the upper limit of the integral is limited to 60,000 km, due to a heavy tail problem of the distribution used in that model. In the following, we only consider this case when we refer to the model of Tanner and Bolduc (2014). The above table shows that both elasticities that are of most interest, $\varepsilon_{E(X_2),y}$ and $\varepsilon_{E(X_2),p_{fuel}}$ are substantially larger in Tanner and Bolduc (2014) "60,000", namely by about 50% and 100%, respectively. In contrast, both elasticities with respect to the cars' fixed costs, $\varepsilon_{E(X_2),k_2}$ and $\varepsilon_{P(X_2=0),k_2}$, are smaller in the case of Tanner and Bolduc (2014), namely by about 30% and 45%, respectively. Also, $\varepsilon_{P(X_2=0),p_2}$ and $\varepsilon_{P(X_2=0),k_2}$ are smaller in the case of Tanner and Bolduc (2014), namely by about 30% and 45%, respectively. In contrast, the elasticity of being carless with respect to income $\varepsilon_{P(X_2=0),k_2}$ is 50% larger in Tanner and Bolduc (2014). There is no straightforward explanation for these differences. One reason could well be that, in the case of Tanner and Bolduc (2014), the penalty function is designed differently, as the difference of the forecasted proportion of carless households and the expectation value of driving demand were each computed on the aggregate level and not as a weighted sum of the squared deviations at the segment level as in (9).

$$Q = c_1 \cdot \left( \frac{P_{sim} - P_{real}}{P_{real}} \right)^2 + c_2 \cdot \left( \frac{E_{sim} - E_{real}}{E_{real}} \right)^2 + \left( \frac{\text{\# elim. observations}}{\text{size of initial datasets}} \right)^2 . \tag{10}$$

Another reason could well be the different specification of the model, in particular, the different assumptions on the error term and the choice of a model based on a direct utility function. Despite the differences in results, most of the resulting elasticities are in the range determined by other studies. Note that the standard deviations of the elasticities of both this model and that of Tanner and Bolduc (2014) are rather small. Note that, in this model, standard deviations are computed using the bootstrapping method. Twelve 10% random samples were taken from all observations.

## Conclusions and further work

While the adaptation of the model's probability function to the empirical observations seems to be very good, see Figure 2, there is a potential to reduce the error of the forecasted driving distances for the different income segments. One way to improve this could be to introduce a fixed utility of car ownership. Unfortunately, the first trial to find such a solution failed. This was because no plausible measure could be found to determine the optimum level of utility to be added for car ownership. Using the value of the penalty function yielded no clear result. Another possibility is to choose the fixed costs of car ownership $k_2$ as an endogenous variable. We would expect the resulting value for $k_2$ to be lower than the economic value. The difference would denote a hedonic value for owning a car. The disadvantage would be that we would leave the pure microeconomic framework, where all economic variables reflect observed and thus true values, and the difference in behaviour between individuals given the same economic variables are explained by differences in preferences. The first trial also failed in this case because the use of the value of the penalty function yielded no clear outcome concerning which value would be the best for $k_2$.

Another option would be to use a model based on a Marshallian demand function that is linear in both income and driving costs. Since, in this case, it would not be realistic for the standard deviation of driving distance to be equal for small distances and large distances, we developed a modelling approach where the standard deviation depends on income, which is the variable that has the greatest impact on driving distance.

We will also recalibrate the model using stated preference data where the fuel price varies much more strongly than in the SFSO (2006) data.

Further research must also be conducted to determine the differences in results between this model and Tanner and Bolduc (2014).

We also need to conduct research into the problem that some segments contain only a small number of observations and thus the proportion of carless households may be very low for statistical reasons alone. We may need to adapt the penalty function in order to avoid data from such segments that are "outlyers" having too great an impact on the penalty function, which could bias parameter estimates and thus elasticities.

## Acknowledgements

We would like to thank Prof. Donzé (University of Fribourg (CH)) for sharing the data with us.

## Literature

Anowar, Sabreena, Naveen Eluru and Luis F. Miranda-Moreno, 2014, Alternative modeling approaches used for examining automobile ownership: a comprehensive review", Transport Reviews, Vol. 34, No. 4, 441-473.

Axhausen, Kay W. and Alexander Erath, 2010, "Long term fuel price elasticity: Effects on mobility tool ownership and residential location choice", Bundesamt für Energie, 2010, Publikation 29015. http://www.bfe.admin.ch/php/modules/enet/streamfile.php?file=000000010334.pdf&name=000000290158

Baranzini, Andrea, David Neto and Sylvain Weber, 2009, "Élasticité-prix de la demande d'essence en Suisse", 2009, Bundesamt für Energie, 2009, Publikation 290036. http://www.bfe.admin.ch/forschungewg/02544/02807/02808/index.html?lang=de&dossier_id=03910

Bhat, Chandra R., 2005, "A multiple discrete-continuous extreme value model: formulation and application to discretionary time-use decisions", Transportation Research Part B 39 (2005), 679-707.

Bhat, Chandra R., 2006, "Household vehicle type holdings and usage: an application of the multiple discrete-continuous extreme value (MDCEV) model", Transportation Research Part B 40 (2006), 35-53.

Bhat, Chandra R., 2008. "The multiple discrete-continuous extreme value (MDCEV) model: role of utility function parameters, identification considerations, and model extensions", Transportation Research Part B 42 (3), 274-303.

De Jong, Gerard C., 1990, "An indirect utility model of car ownership and private car use", European Economic Review, Elsevier, Vol. 34(5), 971-985, July.

De Jong, Gerard, C., Andrew Daly, James Fox, Marits Pieters and Remko Smit, 2004, "Comparison of car ownership models", Transport Reviews, Vol. 24, No. 4, 263-276.

Dargay, Joyce M., 2001, "The effect of income on car ownership: evidence of asymmetry", Transportation Research Part A: Policy and Practice, Volume 35, Issue 9, November 2001, 807-821.

Dubin, Jeffrey A. and Daniel L. McFadden, 1984, "An Econometric Analysis of Residential Electric Appliance Holdings and Consumption", Econometrica, Vol. 52, No. 2 (Mar., 1984), 345-362.

Graham, D.J. and Glaister, S., 2005, Decomposing the Determinants of Road Traffic Demand, Applied Economics, 37 (1), pp. 19-28.

Goodwin, P and Dargay, J and Hanly, M., 2004, “Elasticities of road traffic and fuel consumption with respect to price and income: a review”, Transport Reviews, 24 (3) , 275-292.Hausman, Jerry A., 1981, “Exact Consumer's Surplus and Deadweight Loss”, The American Economic Review, Vol. 71, No. 4. (Sep., 1981), 662-676.

Johansson, Olof and Lee Schipper, 1997, “Measuring the Long-Run Fuel Demand of Cars”, Journal of Transport Economics and Policy, September 1997.

Swiss Federal Statistical Office SFSO, 2006, “Mikrozensus 2005 zum Verkehrsverhalten”, Bundesamt für Statistik, Neuenburg (Switzerland) 2006.

Tanner, R.; Bolduc, D. (2014) , “The Multiple Discrete-Continuous Extreme Value Model (MDCEV) with Fixed Costs”, Procedia Social and Behavioral Sciences, vol. 111, February, pp. 390-399.

Touring Club der Schweiz (TCS), 2007, “Kosten eines Musterautos”. http://www.tcs.ch/main/de/home/auto_moto/kosten/kilometer/musterauto.html