A model of car ownership and use incorporating quality choice and ownership of multiple cars

Jan Rouwendal and John Pommer
Department of Spatial Economics
Free University
De Boelelaan 1105
1081 HV Amsterdam
Netherlands

This version: April 30, 2003

Key words: car ownership, demand for car kilometers, multiple car ownership, quality choice.

Abstract

In this paper we develop and estimate a discrete-continuous model for car ownership and use that incorporates quality choice and the decision to own multiple cars. The basic model, used for instance in De Jong (1991), treats all cars as being equal (no differences in quality) and considers ownership of a single car. In order to introduce quality into the model we assume that the marginal utility of driving is increased by a latent variable (interpreted as quality) that is related to the fixed cost of the automobile, possibly through observable characteristics. We show that this formulation enables us to identify the (subjective) quality concept used by the car owners. Next, we show how to extend the basic model to situations in which 0,1 or 2 cars can be owned. In order to do this we introduce a sub-utility function for mobility whose value is determined by the kilometers driven by both cars (if two are owned) and is identical to the kilometers driven by the car (if one car is owned). The possibility to substitute the kilometers to be driven by the two cars for each other plays an important role in our formulation. Third, we study the model that results when both extensions are used simultaneously.

Our empirical application uses data from the Dutch Consumer Union on (fixed and variable) costs associated with cars. We use an extension of the survey that pays special attention to the ownership of more than one car in the household.

1 Introduction

The automobile is one of the most important consumption goods in modern societies. It may even be said that widespread automobile ownership is one of the defining characteristics of such societies. Participation in social life is difficult for those who do not own a car.

Although the market for automobiles has been studied extensively, most attention seems to have been devoted – at last in recent years – to the functioning of the market as a whole and the exercise of monopoly power. Far less attention has been paid to the micro-characteristics of automobile demand. One of the topics that seems to have received only sparse attention is the ownership of two – or even more cars – by households. This is somewhat surprising, since much of the increase in automobile demand seems to originate from second cars. One would therefore expect analyses of the decision to hold multiple cars. However, such studies seem to be scarce.

During the nineteen eighties there has been a lot of interest in micro-models of automobile demand, but only a few papers have considered the multiple car ownership. Probably the first study to consider ownership of two cars explicitly is the paper by Train and Lohrer (1983) that later appeared as a chapter in Train (1986). This analysis used the framework of discrete choice models. Every car make is considered as a separate alternative. If a household owns two cars, every possible combination of car makes is a separate alternative. It is therefore clear that the number of alternatives to be considered becomes very large, unless some aggregation takes place.

Some studies have considered the way cars are used in households, while taking the number of cars as given. An early example is Mannering (1983) who considered two linear simultaneous equations referring to the use of the two cars in a household. He only considered two-car households. More recently, Golob and co-authors (see Golob, Kim and Ren (1996), Golob and McNally (1997)) used more extensive systems of linear equations – called structural models – in order to study activity participation and car use by household members. In these analyses the number of cars own by the household remains exogenous. These studies make clear that the activity patterns of household members are related to each other. In single car households, coordination between these patterns is necessary if the intervals overlap during which different household members want to participate in activities for which the car is needed. If two cars are available (and the number of licensed drivers exceeds one), these problems can be mitigated. The benefit of owning two cars is therefore related to the constraints that are imposed on household members by the absence of a second car. In the model to be developed below we capture this insight in a simple way by linking ownership of two cars to the possibilities for substituting kilometers driven by the second car for those driven by the first car.

In the present paper we present an analysis of car ownership that considers the choice between one or two cars to be owned and the choice of car quality. The model is developed as a discrete-continuous model. In such models the utility of car ownership is derived from the utility associated with car use. The discrete element of the decision process is to own two cars, one car or no car at all, the continuous element is the number of kilometers to be driven with the car(s). De Jong (1990) develops a discrete-continuous model of car-ownership and use that considers only the decision to own and use a single

car. We extend this approach in two directions. The first is that we introduce ownership of two cars into the model by considering 'car mobility' as a composite commodity. The quantity consumed of this commodity is determined by the number of kilometers driven by the two cars that can be owned by the household. For a household that owns a single car, the number of kilometers of one car is restricted to be equal to zero.
The price of the composite commodity 'car kilometers' is lower than the variable cost of either of the two cars owned by the household, unless these two commodities are perfect substitutes. Perfect substitutability can be identified with a situation in which the utility of the household is independent of the way in which the total number of kilometers to be driven is distributed over the two cars. Since owning two cars is costly, the household will in this situation never decide to own two cars. If the car kilometers are imperfect substitutes, there is something to be gained by owning two cars instead of one. Imperfect substitutability occurs as soon as the household would like to use both cars at the same time. An obvious example is a dual earner households where both workers are employed at different locations and want to use the car for commuting purposes. The gain in utility associated with ownership of two cars appears in the model as a lower price for the composite good. The size of this gain is traded off against the higher fixed cost associated with owning two cars.
The second extension that we introduce concerns the choice of the quality of the cars. We do this by means of a latent variable, 'quality of car kilometers.' We introduce this quality in the same way as is done in the repackaging model and can consider the consumer as choosing the number of car quality units. We assume that the quality of car kilometers is an increasing function of the fixed cost associated with ownership of the car.
The analysis proceeds as follows. In section 2 we discuss the discrete-continuous model of car ownership and use, while paying special attention to the contribution of De Jong (1990). In sections 3 and 4 we consider extensions with respect to the number of cars owned and heterogeneity among cars. ….

2 A discrete-continuous model

People own cars because they want to drive. This obvious statement stresses the close connection between car ownership and car use. If we want to learn about the determinants of automobile demand, one cannot avoid studying automobile use. The literature contains a number of examples of models that explicitly derive the demand for cars from the demand for car kilometers. A positive demand for automobile kilometers does not automatically imply that a car will be owned. The reason is that fixed costs have to be overcome. The associated disadvantage (less money is available for spending on other goods) has to be traded off against the benefits associated with car use. These benefits have to exceed a certain threshold in order to make it beneficial for a household to own a car.
Elaborating this very basic idea leads to so-called discrete-continuous models of automobile demand. To see how this works, consider the following model of car ownership and use. An individual maximizes a utility function with two arguments: car mobility x and a composite consumption good q:

$$u = u(q, x) \tag{2.1}$$

Car mobility is interpreted as the number of kilometers driven with one's own car. Utility is maximized subject to a budget restriction. If no car is owned, this restriction is simply:

$$y = \pi q \tag{2.2}$$

with y income (or consumption expenditure) and π the price of the composite good. Maximum utility is in this case equal to:

$$u^0 = u\left(\frac{y}{\pi},0\right) \tag{2.3}$$

If a car is owned, the restriction is:

$$y = R + px + \pi q \tag{2.4}$$

with R the fixed cost associated with ownership and p the variable cost associated with car use.
Maximization of (1) subject to (4) leads to a demand function for automobile kilometers:

$$x = x(y, p, \pi) \tag{2.5}$$

and an analogous demand function for the composite good. Substitution of these demand function into the utility function (1) leads to the indirect utility function, which gives maximum attainable utility as a function of income and prices:

$$u^1 = v(y, p, \pi) \tag{2.6}$$

A car will be owned if u^1 is larger than u^0. If a car is owned, the demand for kilometers is described by (5).

An example of this model structure can be found in De Jong (1991) who estimated the model on cross section data. The indirect utility function he used is:

$$v = \frac{1}{\beta}\exp(\delta - \beta p) + \frac{1}{1-\alpha}(y - R)^{1-\alpha} \tag{2.7}$$

The value for u^0 can be found by taking the limit for $p \to \infty$:

$$u^0 = \frac{1}{1-\alpha} y^{1-\alpha} \tag{2.8}$$

The parameter δ is assumed to be a random variable. The demand function for automobile kilometers follows from (7) by Roy's identity:

$$\ln(x) = \alpha \ln(y - R) - \beta p + \delta \qquad (2.9)$$

This model is relatively easy to estimate. De Jong (1989) showed that all coefficients are identified by cross section data (that do not contain price variation).
The model leaves several questions unanswered. A major gap is that it does not consider the ownership of two or more cars. De Jong (1991) discussed the possible generalization of the model to ownership of two cars, by specifying a utility function:

$$v = \frac{1}{\beta_1}\exp(\delta_1 - \beta_1 p_1) + \frac{1}{\beta_2}\exp(\delta_2 - \beta_2 p_2) + \frac{1}{1-\alpha}(y - R_1 - R_2)^{1-\alpha} \qquad (2.10)$$

where the suffixes refer to the first and second car, respectively. This utility function implies that the decision to own a second car does not depend on the decision to own a first car. Indeed, the cross-price elasticities of the demands for kilometers by the two cars are equal to zero. De Jong states that applying the complete model (explaining the ownership of 0,1 or 2 cars) would be computationally infeasible (p. 396) and estimated two models. For households with one driving license the choice between 0 and 1 car is modeled, for those with 2 or more driving licenses the choice between 1 and 2 cars. Below, in section 3, we propose a different generalization.
Another issue that De Jong did not consider is the heterogeneity among car makes. He proceeds as if the fixed and variable costs of all cars are identical. For the purpose of analyzing multiple car ownership, this approach seems unsatisfactorily. It is well known that the quality of the second car is usually different from that of the first car. Indeed, the fact that in many cases there is such a clear distinction between the first and the second car in a household point to the relevance of incorporating the heterogeneity among cars into the model. In section 4 we consider how to do this.

3 Multiple car ownership

In this section we generalize the model of section 1 so as to incorporate the decision to own two cars. In order to do so, we will consider x now as an indicator of a (possibly composite) good called automobile mobility. This mobility is a function of the number of kilometers that will be driven with two cars, to be denoted as x_1, and x_2, respectively. We define:

$$x = \left(\delta_1 x_1^{\rho} + \delta_2 x_2^{\rho}\right)^{1/\rho} \qquad (3.1)$$

In this equation the δ's are 0-1 variables that indicate if the household owns a first and a second car (without loss of generality we assume that a single car owned by the household is always car 1). If the parameter ρ is equal to 1, x is simply the sum of x_1 and x_2, that is the total number of kilometers driven. In this special case the kilometers driven by car 1 and car 2 are perfect substitutes. This means that the consumer's utility is independent of the way the total demand for kilometers x_1+x_2 is distributed over the two cars. Since owning two cars is costly, a household will in this situation never decide to do so. If ρ is smaller than 1, kilometers driven by the two cars are not perfect substitutes and it is therefore not possible to transfer car kilometers from one car to the other at a constant level of utility. Less than perfect substitutability between the two demands for car kilometers provides therefore a rationale for owning two cars.

Perfect substitutability may occur if the husband uses the car for home-to-work travelling while the wife uses the car for shopping purposes on Saturday and both use the car simultaneously for joint social and recreational events. Less than perfect substitutability occurs if both partners have to use the car for commuting purposes. Although combining the home-work trip is sometimes possible, it is always costly in terms of time and coordination effort needed. It is, of course, no coincidence that the first example refers to a 'traditional' household, whereas the second refers to a dual earner situation.

When (3.1) is used, the household's utility maximization problem can be interpreted as a two-stage problem. In the first stage an amount z for the total variable cost of automobile mobility is determined, in the second stage it will be distributed over the two cars. Optimal spending on z can be derived from maximizing x in (3.1) under the budget restriction:

$$p_1 x_1 + p_2 x_2 = z \qquad (3.2)$$

which leads to demand equations of the form:

$$x_i = \frac{p_i^{1/(\rho-1)}}{p_1^{\rho/(\rho-1)} + p_2^{\rho/(\rho-1)}} z \qquad (3.3)$$

Substitution of these equation in the group utility function for mobility gives:

$$\begin{aligned} x &= \frac{z}{\left(p_1^{\rho/(\rho-1)} + p_2^{\rho/(\rho-1)}\right)^{(\rho-1)/\rho}} \\ &= \frac{z}{P} \end{aligned} \qquad (3.4)$$

The variable P, defined as the denominator in the first line of (3.4) can be interpreted as a price index for transport. If $0 \le \rho \le 1$ it follows immediately that:

$$0 \le P \le \min\{p_1, p_2\}. \qquad (3.5)$$

This means that the price for the composite good x is smaller that that if the two goods that are its ingredients. This reflects the benefit of owning two cars instead of one and these benefits are larger if ρ is closer to zero. When the two prices (=variable cost per kilometer) are equal : $P = p2^{(\rho-1)/\rho}$. Clearly P approximates 0 when ρ gets close to 0. These properties of P are in line with the interpretation of the parameter ρ as an indicator of the difficulties involved in substituting the kilometers driven by the two cars for each other.

When will a consumer own 2 cars instead of 1? The simple answer is of course: if that offers him more utility. Let z^* denote total expenditure on automobile mobility (fixed plus variable cost). If 1 car is owned: $z=z^*-R$, if two cars are owned: $z=z^*-2R$. If one car is owned $x=(z^*-R)/p_1$, if two cars are owned $x=(z^*-2R)/P$. Two cars will therefore certainly be preferred to one if:

$$\frac{z^*-2R}{P} > \frac{z^*-R}{p_1} \tag{3.6}$$

This is a sufficient condition. An alternative, necessary and sufficient condition uses the indirect utility function v associated with u: two cars lead to a higher utility than one if:

$$v(\pi, P, y-2R) > v(\pi, p, y-R) \tag{3.7}$$

The model just discussed can be extended to situation in which the mobility of car 1 has another marginal utility than that of car 2. A possible specification is:

$$x = \left(\delta_1 a_1 x_1^{\rho} + \delta_2 a_2 x_2^{\rho}\right)^{1/\rho}. \tag{3.8}$$

It is convenient to change quantities and prices in the following way: $x_i^* = a_i x_i$ and $p_i^* = p_i / a_i$. The derivations for the adapted variables are identical to those given above for. Note, however, that the two adapted prices are unequal. If only one car is owned, it is optimal to choose that with the highest value of a_i.

4 Quality choice

The model discussed above ignores differences among cars. Cars are treated as a homogeneous commodity of which either 0 or 1 unit can be owned. In reality there is a large variety of car makes available to the consumer and it is therefore of some interest to see if quality choice can be introduced into the model. In this section a simple way to do this will be discussed. For simplicity, the discussion refers to quality choice in the context of owning a single car.

We start by interpreting car quality as a scalar variable k. It is probable that a higher quality of a car increases the marginal utility of car kilometers and one way to introduce this effect is to reformulate the utility function as:

$$u = u(q, kx) \tag{4.1}$$

The product $kx=x'$ can be interpreted as the number of quality-adjusted car kilometers. The price of a quality-adjusted car kilometer is $p'=p/k$. The budget restriction can be reformulated as:

$$p'x' + \pi q = y - R \tag{4.2}$$

In the short run, the consumer determines the number of quality adjusted car kilometers and other commodities while taking the quality of the car, and the associated fixed cost as given. Utility maximization leads to demand equations and an indirect utility function that can be written as:

$$u = v(\pi, p', y - R) \tag{4.3}$$

The quality of car kilometers is related to the fixed cost of the car. Indeed, the main reason for buying a more expensive car is that it offers a higher quality of driving. It is also possible that there is a trade-off between fuel efficiency and fixed car costs, that is: consumers may be willing to accept higher fixed cost in order to gain lower variable cost because of improved fuel efficiency, lower maintenance cost, et cetera. These considerations suggest that there is a relation between the price per quality-adjusted kilometer and the fixed cost:

$$p' = p'(R) \tag{4.4}$$

with *p'* a decreasing function.
Optimal quality choice implies optimal choice of the fixed cost *R*. The first order condition is:

$$\frac{\partial v}{\partial p'}\frac{\partial p'}{\partial R} = \frac{\partial v}{\partial (y - R)} \tag{4.5}$$

Using Roy's identity, this may be rewritten as:

$$\left(-\frac{\partial p'}{\partial R}\right) x'(\pi, p', y - R) = 1 \tag{4.6}$$

This equation says that the marginal effect of higher fixed cost on the price of a quality adjusted car kilometer should be inversely proportional to the demand for such kilometers. We can rewrite this equation as:

$$\left(-\frac{\partial \ln p'}{\partial R}\right) p' x'(\pi, p', y - R) = 1 \tag{4.7}$$

which says that the marginal effect of fixed cost on the logarithm of the price of a quality adjusted car kilometer should be inversely proportional to (planned) expenditure on such kilometers, that is to variable car cost.
If there would be a linear relationship between variable costs and fixed cost we should have:

$$-\frac{\partial \ln p'}{\partial R} = \frac{1}{a' + b'R} \tag{4.8}$$

implying:

$$p' = (a + bR)^c \tag{4.9}$$

5 Data description

The data set used for estimation can best be looked upon as an amalgamation of three different data sources. We discuss these below. The data describe car ownership, car type choice, car attributes, household and person characteristics for the years 1992, 1993 and 1994. The car panel conducted by the Dutch consumers' association (Consumentenbond) constitutes the foundation of our data set. The car panel contains about 6000 passenger cars. Participants are recruited among the readers of the two-weekly magazine published by the consumers' association. It frequently inserts an advertisement in its magazine to interest readers to join the car panel. Readers who consider participating, can apply by either writing or phoning the consumers' association. Each applicant has to answer some questions first to check whether the car meets the entrance criteria. These are:

The car has to be a member of the sample of car types under study.

The odometer must not exceed 150000 kilometres.

The car is at most 10 years old.

The car covers at least 5000 but at most 40000 kilometres a year.

The consumers' association strives for a preset amount of 50 observations for every car in the sample of car types under study. If a shortage of observations threatens, a repeated advertisement is put in the magazine. However, there's no guarantee for reaching the desired amount and indeed our sample contains car types represented by less than 50 observations.

Participants have to leave the panel when the car gets too old, when the odometer exceeds 150000 kilometres or if the car is no longer under study. If one of these events occur, participants do not receive a next questionnaire. Otherwise, they are allowed to stay in. This means that in spite of the last entrance criterion mentioned, once participating, cars can drive less than 5000 or more than 40000 kilometres.

The car panel holds 214 make/model combinations of passenger cars. These are selected beforehand by the Dutch consumers' association. Very popular vehicle types are under represented and somewhat less popular cars are over represented. The set of cars under study is being updated frequently, making the set continuously to consist of recent make/model combinations.

<u>The car cost survey</u>

Twice a year the panel car owners are requested by the Dutch consumers' association to fill in a form with questions about repair cost – this is the main purpose of the car panel, number of kilometres recorded and attributes of the panel car. The respondent is also inquired for the duration - in months - the panel car is in possession. Mileage is recorded at four points in time each year: January 1st and May 31st in the questionnaire covering the first five months, June 1st and December 31st in the questionnaire covering the final seven months of the year. We translate car use demand to average kilometres per month during the first half and second half of the year respectively.

<u>The household survey</u>

The household survey is conducted on behalf of a publicity campaign which started towards the end of 1992, initiated by the Dutch agency for energy and environment (Novem). The campaign was aimed at achieving a more fuel-efficient car use and acquisition behaviour among passenger car drivers. The data are collected among the participants of the car panel of the Dutch consumers' association in October/November of the consecutive years 1992, 1993 and 1994. The respondents were requested to answer questions about (i) personal and household characteristics, (ii) car use, (iii) car attributes (iv) holding duration, (v) attributes of the former car and (vi) attributes of the second car if present. We use the household survey to add information to the car cost survey, mainly with respect to those items that do not concern the panel car itself: (i) household and panel car owner characteristics, (ii) attributes of the panel car's predecessor, (iii) second car attributes and (iv) mileage of the panel car in case of absence of this information in the car cost survey. As for the latter: if a car is replaced, the car cost survey does not provide us with information on car use during the time interval preceding the transaction moment. For instance, a panel car observed in the questionnaire on May 31st, may have been acquired in March. In that case we lack information on car use from January till March. To reconstruct this data we use the information on the predecessor of the panel car provided by the household survey.

<u>The technical car attributes</u>

The third source contains technical information on the 214 cars included in this study. The data are provided by the Dutch motorists' association (ANWB). The information regards (i) car attributes, (ii) fuel consumption, (iii) fixed and variable depreciation rates, (iv) market prices of new and second hand versions of each of the 214 make/model combinations and (v) the time period in which the version was available for purchase on the market.

We forge together the three data sources described in the previous section – car cost survey, household survey and technical car attributes – into six waves of panel data. The first, third and fifth wave provide information on car attributes, household characteristics and car use, referring to the first six months of the consecutive years 1992, 1993 and 1994. Likewise, the second, fourth and sixth wave provide the same information for the second half of the years mentioned. We assign the kilometres recorded during the first six months to the car that is in possession of the household on June 30th. Similarly, all kilometres driven during the second half of the year are supposed to be covered by the car that is in possession on December 31st.

When deleting the observations from the sample which suffer from item non response, we lose a great number of observations (around 50%). In order to be able to estimate the car ownership model for two cars, we must have the same information for the second car as for the panel car. This information requirement is decisive for a further loss of observations, resulting in the sub sample (NT = 17174) which is used for the estimation exercises performed in this paper. Table 1 gives a comparison of some items between the sample and the sub sample.

Table 1: Some (sub) sample means (std dev)

Characteristics	Sample	Sub sample
Observations		
NT	34980	17174
N = Households [#]	7909	4747
T = Waves [#]	4.42 (1.75)	3.62 (1.51)
Transactions [#] / Observed [#]	3409 / 3409	2093 / 936
Multiple cars owning households [#]	1610	519
Panel car attributes		
Mileage per 6 months [km]	8539 (4646)	8343 (4300)
Age [years]	3.04 (2.18)	3.23 (2.18)
Engine power [cc]	1481 (275)	1480 (2.63)
Weight [kg]	932 (133)	932 (130)
Purchase price [€]	12444 (3782)	12251 (3550)
Household characteristics		
Employees [#]	1.17 (0.87)	1.09 (0.86)
Persons [#]	2.57 (1.09)	2.53 (1.08)
Driving licence owners [#]	1.83 (0.60)	1.78 (0.57)
Income [brackets 1-7]	5.14 (1.32)	5.03 (1.30)
Cars [#]	1.22 (0.46)	1.10 (0.33)
Company cars [#]	0.03 (0.17)	0.02 (0.13)
Panel car owner characteristics		
Age [years]	48.8 (13.4)	49.7 (13.7)
Gender [1=male, 0=female]	0.88 (0.33)	0.89 (0.31)
Education [levels 1-5]	3.37 (1.01)	3.37 (1.00)
Second car attributes		
Mileage per 6 months [km]	6595 (6246)	7800 (6886)
Age [years]	5.51 (3.78)	4.43 (3.56)
Engine power [cc]	1350 (451)	1418 (447)
Weight [kg]	871 (228)	906 (232)
Purchase price [€]	8704 (6704)	10383 (6883)

We observe a drop out of multiple cars owning households. These are households that own more than one car during at least one period.
In the sample the average percentage of households owning more than one car during the observed time horizon (1992-1994) is 15%. In the sub sample this share is 9%.

Percentage of Dutch households holding none, one, two, three or more cars							
	1991	1992	1993	1994	1995	1998	2002
None	26.2	26.9	25.2	25.5	24.9	23.5	
One	60.2	59.4	60.6	60.2	60.3	59.1	59.1
Two	12.4	12.7	13.0	13.2	13.8	16.3	
Three or more	1.2	1.0	1.2	1.1	1.0	1.1	
Two or more, given at least one	18.4	18.7	19.0	19.2	19.2	22.7	

Source: CBS Statline and TNO Inro

The percentage of Dutch households owning at least two cars, given that they possess at least one car, amounts to an average of 19% during the observed time horizon (1992-1994). These figures are given at the bottom of the table. They can be calculated using the percentages in the table above the dash. Concluding, the sample includes fewer multiple cars owning households than the Dutch population. Obviously, this holds even stronger for the sub sample.

Furthermore, note that the second car is on average smaller, older and less used than the panel car. This suggests that within the household a hierarchy exists between the two cars: most of the time the panel car is the principal car, the second car is the auxiliary car. Most likely, this is caused by the way the set of cars under study is composed, namely recent make/model combinations, so relative young cars, and the way the participants in the car panel are recruited: through an advertisement in the magazine of the consumers' association yielding mainly male participants which bring along the bigger cars. The loss of observations on multiple cars households make the second car attributes in the sub sample to lie closer to those of the panel car, although the fore mentioned hierarchy is still present. This implicit and troubled hierarchy is very unsatisfactory for our analysis. We therefore introduce an explicit and clear hierarchy by labelling the most used car as the first car. If we do so, Table 1 must be revised. The result is presented in Table 2.

From Table 1 we can derive an indication for the average holding. If 7909 households perform 3409 car replacements during on an average 4.4 periods, we calculate: 7909/3409 * 4.4 = 10.2 waves which is the same as 5.1 years.

Actually, the observed unit is ‘car’, not ‘household’. However, there are very few households in the sample who have multiple cars participating in the car panel. To be specific, 38 households in the sample are represented with two cars, 1 household with three cars - so strictly, 7869 households participate in the sample (with 7909 cars). In sub sample 1, 24 households are represented with two cars, 1 household with three cars – so the number of households in sub sample 1 amounts to 5626. Finally, 15 households have two cars participating in sub sample 2, which leaves us with a total of 4736 households.

Concluding, though not entirely similar, we allow ourselves to both use ‘car’ as well as ‘household’ to indicate the observed unit.

Though not verifiable, this seems a very plausible figure. It is probably biased upwards because company (lease) cars are under represented in our sample. These cars tend to have a short holding duration as lease contracts are often closed for a three years period.

Table 2: Some sub sample characteristics if the first car is the most used car within the household

Characteristics	Mean (std dev)
Observations	
NT	1584
N = Households [#]	519
T = Waves [#]	3.05 (1.34)
First car attributes	
Mileage per 6 months [km]	11929 (6268)
Age [years]	2.88 (1.95)
Engine power [cc]	1578 (274)
Weight [kg]	972 (134)
First car owner characteristics (NT = 967)	
Age [years]	46.0 (11.5)
Gender [1=male, 0=female]	0.88 (0.32)
Education [levels 1-5]	3.63 (0.95)
Second car attributes	
Mileage per 6 months [km]	4568 (2917)
Age [years]	4.75 (3.44)
Engine power [cc]	1257 (369)
Weight [kg]	824 (179)
Second car owner characteristics (NT = 617)	
Age [years]	44.9 (11.2)
Gender [1=male, 0=female]	0.57 (0.50)
Education [levels 1-5]	3.52 (1.08)

In this paper we use the sub sample (NT = 17174) under our imposed hierarchy in a static context. It means among other things that we presume that observations for the same household over time are independent. We realize this is a rather stringent presumption.

To get a grasp of the quality of our data set, we compare it with national statistics published by the Netherlands bureau of statistics yearly. Some features are given in Table 3 and Table 4 in the Appendix. They show a close equivalence in value between overall mean mileage per year of panel cars compared with the Dutch car fleet. Entering into the figures more detailed, we see that petrol cars cover more kilometres in our sample whereas diesel en LPG cars drive approximately the same distance compared with their counterparts in the Dutch car fleet. In our sample the number of diesel and LPG cars are under represented. This is not surprising since those cars tend to cover many kilometres. Recall that cars covering more than 40000 kilometres a year do not enter the car panel of the Dutch consumers’ association.

Our sample represents the middle segment of the Dutch car fleet with respect to mileage. This leads - for one thing - to an under representation of diesel, LPG and company cars. These cars tend to move in the upper mileage segment. On an average, the cars in the middle segment cover less kilometres than the Dutch car fleet as a whole. On the other hand, our sample consists of younger cars than the Dutch car fleet. As younger cars tend to cover more kilometres than older cars, the two counteracting circumstances - middle segment versus young cars – make overall mean mileage of the sample to correspond with the Dutch car fleet.

In conclusion, the following marginal notes can be made:
- The cars in our sample are younger than the Dutch car fleet.
- Because of counteracting effects, mean mileage in our sample corresponds with the Dutch car fleet.

And not reported here:
- The households in our sample hold slightly more members than the Dutch households (2.6 in our sample versus 2.4 in the Dutch population 1992-1994).
- The households in our sample possess fewer company cars (0.3 in our sample versus 0.12 in the Dutch population).
- The car owners in our sample are somewhat older than the Dutch population (over eighteen) (49-50 in our sample versus 46 in the Dutch population 1992-1994).

6 Estimating the model with multiple cars

In this section we consider the question how the model that has been developed in section 3 can be estimated on the data discussed in the previous section.

De Jong (1991) treated the parameter δ as a random variable in order to take into account differences in demand for automobile kilometers among households that are not related to differences in observable characteristics of the household. In the present model it seems likely that there are also differences in the substitutability of kilometers driven by the two automobiles that cannot be related to differences in observable characteristics. This suggests that the parameter ρ should also be treated as a random variable.

Note that:

- if δ is very small (less than zero and large in absolute value), no car will be better than either one or two cars,
- if δ is large, at least one car will be owned,
- if ρ is close to 0, two cars will always be preferred to one car,
- if ρ is close to 1, one car will always be preferred to two cars.

This suggests that there will always be a positive probability for all three events that 0,1 or 2 cars will be owned.

For any given pairs of values of δ and ρ we can determine:

$$u^0 = \frac{1}{1-\alpha} y^{1-\alpha} \qquad (6.1)$$

$$u^1 = \frac{1}{1-\alpha}(y-R)^{1-\alpha} + \frac{1}{\beta}\exp(\delta - \beta\, p_1) \tag{6.2}$$

$$u^2(\rho) = \frac{1}{1-\alpha}(y-R_1-R_2)^{1-\alpha} + \frac{1}{\beta}\exp(\delta - \beta\, P(\rho)) \tag{6.3}$$

Note that only u^2 depends on ρ. Let u^c=min(u^1,u^2). Since u^1 and u^2 are bot increasing in δ, u^c will also be increasing in δ. There is a critical level, to be denoted as δ^*, at which owning one or two cars will become more attractive than having no car.
Since all the households in our sample own at least one car, we have only households with $\delta > \delta^*$. Note that δ^* may depend on ρ.

For any given value of ρ there is a second critical value of δ, to be denoted as δ^{**}, at which owning two cars becomes more attractive than owning one car. This second critical value depends on ρ too. The probability p_1 that a household in the sample owns one car is:

$$p_1 = \int_0^1 \Pr(\delta < \delta^{**}(\rho) \mid \delta > \delta^*(\rho)) f(\rho) d\rho \tag{6.4}$$

The probability in this equation may be equal to zero: our model does not guarantee that δ^{**} exceeds δ^*. However, if ρ is close enough to 1, owning one car will become more attractive than owning two cars.
The probability that a household owns two cars is defined analogously.
We adopt the following specification for δ:

$$\delta = \mu_\delta + \sigma_\delta \varepsilon \tag{6.5}$$

with ε a standard normal distributed random variable. This equation says that δ is normal distributed with mean μ_δ and standard deviation σ_ε. The critical values for δ imply critical values for ε, that will be denoted analogously. We find:

$$\varepsilon^{**} = \frac{1}{\sigma_\delta}\ln\left(\frac{\beta}{1-\alpha}\,\frac{(y-R_1)^{1-\alpha} - (y-R_1-R_2)^{1-\alpha}}{\exp(\mu_\delta)\left(\exp(-\beta\, P(\rho)) - \exp(-\beta\, p_1)\right)}\right) \tag{6.6}$$

and for the other critical value:

$$\varepsilon^* = \min\left\{\varepsilon^{01}, \varepsilon^{02}\right\} \tag{6.7}$$

with:

$$\varepsilon^{01} = \frac{1}{\sigma_\delta}\ln\left(\frac{\beta}{1-\alpha}\,\frac{(y)^{1-\alpha} - (y-R_1)^{1-\alpha}}{\exp(\mu_\delta)\left(\exp(-\beta\, p_1)\right)}\right) \tag{6.8}$$

and:

$$\varepsilon^{02} = \frac{1}{\sigma_\delta} \ln\left(\frac{\beta}{1-\alpha} \frac{(y)^{1-\alpha} - (y - R_1 - R_2)^{1-\alpha}}{\exp(\mu_\delta)(\exp(-\beta P(\rho)))} \right). \tag{6.9}$$

For any given value of ρ (and the parameters) the critical values of ε can be computed easily. Since ε is standard normal distributed, we can also easily compute the probability under the integral sign in (3.12) and the analogous equation for ownership of two cars.

We can draw randomly from the distribution of ρ and use the random drawings to approximate the integral in 3.12 as:

$$p_1 \approx \frac{1}{S} \sum_{s=1}^{S} \Pr\left(\delta < \delta^{**}(\rho^s) \mid \delta > \delta^{*}(\rho^s)\right) \tag{6.10}$$

where ρ^s denotes the s-th drawing from the distribution of ρ and S the total number of drawings.

For f(ρ) we can use the beta distribution:

$$f(\rho) = \frac{(\alpha + \beta + 1)!}{\alpha! \beta!} \rho^\alpha (1-\rho)^\beta \tag{6.11}$$

For the numbers of kilometers driven we have to distinguish between the situation in which one car is owned and that in which two cars are owned. In the former case we have the (expected) demand equation:

$$\ln(x) = \alpha \ln(y - R_1) - \beta p + \delta_1 \tag{6.12}$$

with:

$$\delta_1 = E(\delta \mid \delta^* < \delta < \delta^{**}) \tag{6.13}$$

Assuming a normally distributed error term e (which is independent of δ), the likelihood of observing a demand for kilometers x^* is:

$$\phi\left(\frac{\ln(x^*) - \alpha \ln(y - R_1) + \beta p - \delta_1^*}{\sigma_e} \right) \tag{6.14}$$

where φ denotes the standard normal density function.

If two cars are owned, the demand equation for x is:

$$\ln(x) = \alpha \ln(y - R_1 - R_2) - \beta P + \delta_2 \quad (6.15)$$

with:

$$\delta_2 = E(\delta \mid \delta^{**} < \delta) \quad (6.16)$$

From (3.3) we have:

$$\ln(x_i) = \frac{1}{\rho - 1} \ln(p_1) - \ln(P) + \ln(z) \quad (6.17)$$

Since $z=Px$, we can rewrite this as:

$$\ln(x_i) = \frac{1}{\rho - 1} \ln(p_1) + \alpha \ln(y - R_1 - R_2) - \beta P + \delta_2 \quad (6.18)$$

We have to add an error term to each equation. We should allow for correlation between the two error terms e_1 and e_2. The bivariate normal distribution can be used for this purpose. The realized values of the errors are the differences between observed and predicted values.

The likelihood for a household owning one car is:

$$l_1 \approx \frac{1}{S} \sum_{s=1}^{S} \Pr(1\,\text{car} \mid \text{at least one car}) \Pr(e_1 \mid 1\,\text{car}) \quad (6.19)$$

and for a household owning two cars:

$$l_2 \approx \frac{1}{S} \sum_{s=1}^{S} \Pr(2\,\text{cars} \mid \text{at least one car}) \Pr(e_1, e_2 \mid 2\,\text{cars}) \quad (6.20)$$

The probability that one car is used, given that at least one car will be used was discussed above, see (3.18). The probability that two cars will be used, given that at least one will be used can be derived analogously. The probabilities that the e's have a particular value should be interpreted as the values of the appropriate (univariate or bivariate) normal density functions.

7 Estimating the model with quality choice

In the discrete-continuous model discussed in chapter 2, fixed and variable cost of car kilometers were considered as exogenous variables, whose values were taken as given by the consumers. In section 4, it was assumed that car quality is chosen by the consumer and that it is related to fixed car cost. Fixed car cost thus became an endogenous variable. However, it can still be argued that the demand for kilometers is determined by a consumer while taking the characteristics of the currently owned car as given. Even if the complete model is meant to explain car quality as well, it may reasonably be expected that this quality is predetermined when the decision about car use is taken.

This leaves us with the question how quality choice should be modeled. In section 4 we have suggested a formulation that avoids explicit use of this latent variable and focuses attention on the choice of fixed car cost. Ideally, one would derive a reduced form equation from the model in which no other dependent variables appear and estimate its coefficients. However, it is often impossible to find an analytical reduced form expression for *R* as determined by the model. To see this, return to equation (4.5) and use, for instance, the indirect utility function (2.7). Observe, moreover that *p'* is in general a nonlinear function of *R*, for instance the one specified in (4.9). Solving the resulting expression for *R* is impossible.
In order to solve this problem, we proceed as follows. Even though we can't find an equation for *R*, we can solve the equation numerically for any values of the parameters of the model, to be denoted as θ. The resulting value of *R* will be denoted as $R^*(\theta)$. We kan now assume:

$$R = R^*(\theta) + \varepsilon \tag{7.1}$$

where *R* is the actual value of fixed car cost and ε is a disturbance term. This model can be estimated by nonlinear least squares. The target function is:

$$T = \sum_i \left(R - R^*(\theta)\right)^2 \tag{7.2}$$

and the first order conditions are:

$$\sum_i -2\left(R - R^*(\theta)\right)\frac{\partial R^*(\theta)}{\partial \theta} = 0 \tag{7.3}$$

Solving these conditions results in parameter estimates.

8 Conclusion

In the previous sections we have discussed two extensions of a discrete-continuous model of car ownership and use. These extensions allow us to incorporate ownership of more than one car and quality choice into this model. We have also discussed the data set on which these models will be estimated and how to estimate the models. Future versions of the paper will contain estimation results as well.

References

Golob, T.F., S. Kim and W. Ren (1996) How Households Use Different Types of Vehicles: A Structural Driver Allocation and usage Model *Transportation Research* **30A** 103-118.

Golob, T.F. and M.G. McNally (1997) A Model of Activity Participation and Travel Interaction Between Household Heads *Transportation Research* **31B** 177-194.

Hensher, D. (1985) An Econometric Model of Vehicle use in the Household Sector *Transportation Research* **19B** 303-313.

Jong, G.C. de (1989) Some Joint Models of Car Ownership and Use, PhDF thesis, University of Amsterdam.

Jong, G.C. de (1991) A Micro-Economic Model of the Joint Decision on Car Ownership and Car Use, p. 387-403 in: *Methods for Understanding Travel Behavior in the 1990s* Proceedings, Conference International sur les Comportements de Deplacements, Quebec.

Mannering, F.L. (1983) An Econometric Analysis of Vehicle Use in Multivehicle Households *Transportation Research* **17A** 183-189.

Train, K. (1986) *Qualitative Choice Analysis* MIT Press, Cambridge (Ma)

Train, K. and M. Lohrer (1983) Vehicle Ownership and Usage: An Integrated System of Disaggregate Demand Models, paper presented at TRB Annual Meeting.