AMSTER

AER no. 1988/6

INSTITUTE OF ACTUARIAL SCIENCE & ECONOMETRICS

REPORT AE 6/88

11

JOINT PREDICTION OF AUTOMOBILE OWNERSHIP AND MILEAGE BY A CROSS-SECTION MODEL

G.C. de Jong, H.P. Boswijk and J.S. Cramer

University of Amsterdam.

Faculty of actuarial science & econometrics

Title: **Joint prediction of automobile ownership and mileage by a cross−section model**

Author: G.C. de Jong, H.P. Boswijk and J.S. Cramer

Address: University of Amsterdam
Institute of Actuarial Science and Econometrics
Jodenbreestraat 23
1011 NH Amsterdam

Date: July 1988

Series and number: Report AE 6/88
Revision and extension of Report AE 2/88

Pages: 20

Price: No charge

JEL Subject Classification: 211/921/933

Keywords: Discrete and conditional prediction/car ownership/car use

Abstract:
A previously constructed model, which explains car ownership and private car use simultaneously and which was originally estimated on the 1980 Dutch budget survey, is now applied to 1985 in order to obtain validation of this model. Two methods are used: post sample prediction and re−estimation. This exercise raises some general issues of predicting individual discrete choice and of conditional prediction in a simultaneous framework. The main result is that a model which performs rather well at the aggregate level may fail to explain and predict behaviour at the level of the individual household.

JOINT PREDICTION OF AUTOMOBILE OWNERSHIP AND MILEAGE BY A CROSS-SECTION MODEL

G.C. de Jong, H.P. Boswijk and J.S. Cramer

Department of Econometrics
University of Amsterdam
Jodenbreestraat 23,
1011 NH Amsterdam

Acknowledgement
We gratefully acknowledge support from the Netherlands Organization for Scientific Research (N.W.O.) and of the Dienst Verkeerskunde of Rijkswaterstaat.

1. Introduction

We have constructed a model of the simultaneous household demand for private automobile use, and fitted it to individual household data from the Dutch 1980 budget survey (De Jong and Cramer, 1987). The result is now applied to the survey of 1985 in order to obtain validation of the model by post sample prediction. This exercise raises some general issues of predicting individual discrete choice and of conditional prediction in a simultaneous framework. It also brings home that a model with a fair overall performance may well fail to explain or predict behaviour at the individual level.
The model and the data are described in section 2 and 3. The post sample prediction methods and – results can be found in section 4. Section 5 concerns another method of validation of the model, which is re-estimation.

2. The model

The central relation of the model is an Engel curve for the logarithm of household i's intended private automobile mileage, y_i^*, as a function of per capita income, family size, age, and of dummies for farmer households and for woman drivers. This function has a disturbance v. The random variable y_i^* governs private automobile ownership, for this occurs iff y_i^* exceeds a constant threshold value γ, which in this model is to be estimated. In the event, observed private mileage y_i (also in logarithms) will consist of y_i^* plus a further disturbance w. It is helpful to think of y_i^* as 'permanent mileage', with v representing such permanent unobservables as taste, while w stands for transitory factors that intervene after the decision to own a car has been taken. The two disturbances are assumed independent Normally distributed with zero mean. This model can be summarized as follows:

$$y_i^* = x_i'\beta + v_i \; , \tag{1}$$

$$y_i = y_i^* + w_i \quad \text{if } y_i^* > \gamma \; , \tag{2}$$

$$y_i \text{ not observed otherwise} \; , \tag{3}$$

$$v_i \text{ and } w_i \text{ independent and Normally distributed with zero means and variances } \sigma_v^2 \text{ and } \sigma_w^2 \; . \tag{4}$$

At times we shall also use

$$u_i = v_i + w_i \text{, a Normal variate with zero mean and variance } \sigma_u^2 = \sigma_v^2 + \sigma_w^2 \; . \tag{5}$$

It will be clear that the model provides expressions for the probability that a given household i owns a private automobile,

$$P_i = P(y_i^* > \gamma) = P(x_i, \theta) \; , \tag{6}$$

and that it then has a private annual mileage with logarithm y_i, given by

$$y_i \mid y^*_i > \gamma = x_i'\beta + v_i + w_i \; . \tag{7}$$

The complete set of parameters θ consists of the vector β, γ, and the variances σ_v^2 and σ_w^2. Once these have been estimated the expressions (1), (6) and (7) can be evaluated for any household with given values of the regressor variables. The result can in turn be used to predict aggregate or individual ownership and mileage.

The model applies to households who have a free choice in the matter of automobile ownership, in the sense that they are not compelled to own a car for business or professional use. If they are, their car is defined as a business car, regardless of whether the costs are borne by the employer, set off against tax allowances, or paid for out of gross income. The survey distinguishes between private and business cars, and also between single and multiple ownership. It also provides information about private mileage, that is (annual) mileage for private purposes, including the journey to and from work. In the case of multiple car ownership, private mileage is recorded for the principal car only. The analysis is therefore in the main restricted to (A) households without an automobile and (B) households with one private automobile. Households with more than one private car, or with a combination of private and business cars, have been omitted from the data set.

Households with one or more business cars (there is no distinction by number) form a third category (C). These households have access to an automobile, regardless of their demand for motoring, and their private mileage may be expected to be governed by the relation

$$y_i = x_i'\beta + v_i + w_i = x_i'\beta + u_i , \tag{8}$$

with the same parameters as in (1). This is the Engel curve for mileage from (1) and (2), without the ownership condition. The estimates of β and σ_u^2 may therefore be strengthened by including category (C) in the analysis.
In yet another version of the model the threshold γ is replaced by an endogenous γ_i , which depends on per capita income, household size, age and degree of urbanization.

— see here Table 1 —

The 1980 survey yields plausible and reasonably precise maximum likelihood estimates of the model's parameters θ, which are shown in Table 1. But residual variation is high, with σ_u, measured in logs, equal to .53. We therefore can not expect the model to reproduce individual houehold behaviour at all closely. For details of the estimation and of the data we refer to de Jong and Cramer (1987).
We call this model simultaneous for two reasons. First, it is estimated by maximizing a single likelihood function. In the second place car ownership is made dependent on permanent mileage, while observed car use is conditional on car ownership.

3. Two samples

We shall use the estimates from the 1980 survey to calculate predicted aggregate and individual ownership and mileage in both the 1980 and the 1985 survey. But for minor changes, the 1985 survey has the same design, the same format, and the same variables as the 1980 survey. 1985 Incomes must of course be expressed in 1980 guilders.

— see here Table 2 —

Table 1. Parameter estimates for 1980 (n= 2308)

	datasets A and B	
coefficient β of log mileage:		
intercept	2.25	(5.5)
log income	.25	(6.2)
log size	.25	(6.2)
age	-.02	(-5.1)
DA	-.01	(-.4)
DF	-.11	(-4.2)
threshold γ	4.578	(116.0)
σ_V	.19	
σ_W	.49	
value of log likelihood (not including constant)	-880	

t–ratios in brackets. Income per equivalent adult, size in number of equivalent adults, age in five–year classes; DA farmer household dummy, DF woman driver dummy.

Table 2. Some sample statistics

	1980	1985
	numbers*	
households		
(A) without car	763	553
(B) with one private car	1545	1791
(C) with business car(s)	245	185
	means	
income per equivalent adult, 1,000 1980 guilders per year	15.663	14.234
household size, number of persons	2.01	2.19
age of head of household, years	46.60	40.07
private mileage, 1,000 km per year, of		
- private car	13.86	12.55
- business car	7.93	9.36

*) after omitting a small number of incomplete observations

Table 2 shows that the households of the 1985 survey are younger, somewhat larger and poorer than those of 1980. We also see that private automobile ownership is more prevalent and that private mileage per car is lower.

4. Post sample prediction

For the post sample validation of the model the main issue is whether application of the 1980 estimates to the 1985 sample reproduces the observed levels of ownership and mileage. But the application of a set of estimates to a sample is not without conceptual problems, (i) because of contrasts between the prediction of aggregate values and of individual behaviour, and (ii) because of the joint determination of ownership and mileage. These problems arise equally well in prediction of the 1980 survey from the 1980 estimates as in post—sample prediction of the 1985 survey, and we shall treat both cases.

We consider automobile ownership first, and annual mileage next.

4.1. Aggregate Ownership

The model produces only ownership probabilities (equation (6)). We predict ownership by summing these probabilities using all households (A) and (B).
Define Z_i as a (0,1) ownership variable for household i. By (6)

$$P(Z_i = 1) = P(x_i, \theta) \ , \tag{9}$$

and its estimate is

$$\hat{P}_i = P(x_i, \hat{\theta}) \ . \tag{10}$$

Let m denote the number of private car owners in the sample. Its expected value and variance are given by

$$E(m) = \Sigma P_i \ , \qquad Var(m) = \Sigma \ P_i(1 - P_i) \ . \tag{11}$$

We predict m by the estimate of its expected value, or

$$\hat{m} = \Sigma \ \hat{P}_i \ , \tag{12}$$

and also calculate its estimated variance

$$\mathrm{Var}(\hat{m}) = \Sigma\ \hat{P}_i(1 - \hat{P}_i)\ . \tag{13}$$

The result of this exercise, using the 1980 parameter estimates in calculating the individual probabilities for the 1980 and the 1985 survey households, are shown in Table 3. The predicted overall ownership rate is 3% too high in 1980, 4% too low in 1985; prediction in the alien sample is not much worse than in the estimation sample, and the increase in aggregate ownership is correctly predicted, although its extent is understated. The relevant variables are individual incomes, age, household size; the effect of lower real incomes in the 1985 sample is apparently offset by the younger age and the larger size of households. The concomitant rise in ownership is however underestimated, for the actual ownership rate increases from 66.9% to 76.4%, or by 14%, and the predicted rate only from 68.9% to 73.4%, or by 6.5%.

— see here Table 3 —

4.2. Individual Ownership

Although we realize that the residual variation is high and we are primarily interested in predicting aggregates, we are curious to know how the model performs at the individual level and whether this performance differs much from 1980 to 1985. Methods developed in this and the next section can also be used for predicting for specific subsamples (see section 4.5). The problem here is one of deriving a discrete variable from a probability. While (10) provides an estimate of P_i or $P(Z_i = 1)$, in order to predict the value of Z_i we need a further decision rule. The first thing that comes to mind is to replace the ownership condition from (1) and (2) by its analogue for etimates, that is

$$\hat{Z}_i = 1 \quad \text{iff} \quad x_i'\beta > \hat{\gamma}\ , \tag{14}$$

which is equivalent to the rule

$$\hat{Z}_i = 1 \quad \text{iff} \quad \hat{P}_i > .5\ . \tag{15}$$

At first sight this may appear quite reasonable — $Z_i = 1$ if the probability of $Z_i = 1$ exceeds that of $Z_i = 0$ — but in practice it leads to a far larger predicted overall ownership than we found earlier, nowhere near actual ownership levels. We therefore replace (15) by

$$Z_i = 1 \quad \text{iff} \quad P_i > P_c\ , \tag{16}$$

Table 3. Aggregate private automobile ownership

	1980	1985
	numbers	
sample size 1)	2308	2344
households with one private car		
- actual	1545 (66.9%)	1791 (76.4%)
- predicted 2)	1590 (68.9%) (18.6)	1721 (73.4%) (19.0)

1) Categories A and B of Table 1 only.

2) Standard error in brackets

where the cut−off point P_c is set by predicted number m of (12). Ownership is thus assigned to households with the higher probabilities, starting from the top, until we have m households at P_c = .685. This way the individual ownership predictions are consistent with the aggregate ones.

− see here Table 4 −

Table 4 shows the performance of the model in predicting private automobile ownership of individual households. The overall result is mediocre, as was to be expected in view of the considerable residual variation. Perhaps we must admit, that we can not hope to achieve much better results at the individual level. The remarkable thing is that the result for the alien sample of 1985 is not much worse than for the 1980 sample to which the model has been fitted. To measure performance here we use the classical test statistic for independence in a contingency table, which has a chi−square (1) distribution under the null. Its value is 610 for 1980 and 223 for 1985. Independence of predicted and observed ownership is of course roundly rejected in both cases, but the point is that the difference in the test statistic between the estimation sample and the alien sample is relatively modest.

4.3. Aggregate mileage

Predicted aggregate mileage is built up from predicted values for individual households, just like ownership; but we shall distinguish three diferent methods. All predict the expected value of y_i for household i. As this is the logarithm of annual private mileage, not mileage itself, a correction factor is needed to arrive at expected mileage, (see Aitchison and Brown, 1957). If y_i has expected value $E(y_i)$ and variance σ^2, expected mileage is given by

$$\exp(E(y_i) + \tfrac{1}{2}\sigma^2) \quad . \tag{17}$$

This correction has been applied as a matter of course to all values given below, with σ^2 equal to the estimated value of σ_u^2, in view of (1), (2) and (5). This increases predicted mileage by a factor 1.15.

There are three definitions of $E(y_i)$, the expected value of y_i. The first ignores all information on car ownership , and simply multiplies the expected value of log mileage for household i by the probability that is realized, i.e. by the probability of car

Table 4. Individual private automobile ownership

1980		predicted $\hat{Z}_i = 0$	predicted $\hat{Z}_i = 1$	Σ
actual	$Z_i = 0$	497	266	763
	$Z_i = 1$	221	1324	1545
Σ		718	1590	2308

1985		predicted $\hat{Z}_i = 0$	predicted $\hat{Z}_i = 1$	Σ
actual	$Z_i = 0$	282	271	553
	$Z_i = 1$	339	1452	1791
Σ		621	1723	2344

ownership. This leads to

$$E(y_i)_1 = \hat{P}_i \cdot x_i'\hat{\beta} \;. \tag{18}$$

The other two methods consider expected log mileage conditional upon automobile ownership. This calls for a Mills' ratio term for the expectation of the disturbance of the Engel curve for log mileage, for this disturbance has a truncated distribution among car owners, (see for example Heckman, 1979). We write $\Phi(\sigma_V)$ for this term, which is defined as

$$\Phi(\sigma_V) = \sigma_V \cdot \frac{f[(x_i'\beta - \gamma)/\sigma_V]}{F[(x_i'\beta - \gamma)/\sigma_V]} \;, \tag{19}$$

where f[.] and F[.] are the density and the distribution function of the Standard Normal.

We then have

$$E(y_i)_2 = x_i'\hat{\beta} + \Phi(\hat{\sigma}_V) \qquad \text{if household i owns car} \tag{20}$$

(observed ownership) or, in an obvious variant,

$$E(y_i)_3 = x_i'\hat{\beta} + \Phi(\hat{\sigma}_V) \qquad \text{if it is predicted that household} \tag{21}$$

i owns a car (predicted ownership).

Having obtained these predictions of individual mileage, average sample mileage is obtained by summing over all households. We compare the result in the first column of Table 5, which gives values **per household**; this corrects for sample size, while avoiding the awkward choice of a numerator for expressing the prediction according to (18) in mileage per car.

— see here Table 5 —

For 1980 all three methods perform very well. For 1985 we find that the naive first method does best, coming very close to observed mileage. The difference of just 2% is quite encouraging. The more sophisticated methods overestimate mileage by about 10% in the 1985 sample, method 2 actually doing worse than method 3 even though it makes use of additional factual information about automobile ownership.

In the second column of Table 5 we express these result in mileage per car, using observed numbers for method 1 (for want of anything better) and for method 2; for method 3 the predicted number is in order. The only change is the relative performance of method 3, which improves in 1980 and worsens in 1985.

The naive method may perform well, but it is of course not at all in keeping with the

Table 5. Annual private mileage

		per household (x 1000 km)	per private car (x 1000 km)
1980 survey: observed		9.315	13.860
predicted	I	9.247	13.813
	II	9.409	14.055
	III	9.722	14.112
1985 survey: observed		9.591	12.553
predicted	I	9.779	12.798
	II	10.756	14.077
	III	10.488	14.284

The naive method may perform well, but it is of course not at all in keeping with the spirit of the model. We must conclude that proper application of the model, while doing well in 1980, overestimates the 1985 private mileage.

4.4. Individual mileage

At the individual household level we may compare predicted mileage with actual mileage whenever this has been observed, that is among households with a private automobile. Predicted mileage is then obtained according to (20). Writing M_i for mileage (in thousands of kilometers per year), not log mileage, we calculate a simple linear regression for observed mileage as a function of predicted mileage. The result is

$$\text{1980 sample} \quad M_i = \underset{(11.3)}{1.33} \; M_i^* - \underset{(-2.5)}{3.990} \qquad R^2 = 0.08 \; , \tag{22}$$

$$\text{1985 sample} \quad M_i = \underset{(12.1)}{1.45} \; M_i^* - \underset{(-4.2)}{6.610} \qquad R^2 = 0.08 \; . \tag{23}$$

With 1545 and 1791 observations the relations are definitely significant, but hopes for a constant of nearly zero and a slope coefficient of close to 1 are deceived. The model is just too simple to describe or predict individual behaviour.

4.5. Predicting for subsamples

So far we have obtained predictions which at the individual level are clearly not good enough, but which are not unacceptable at the level of the total sample used. It is still open to question how the model performs at the level of specific subsamples. Another reason for looking at subsamples is to respond to a possible form of criticism at the aggregate validation results, which says that the 1980 and 1985 samples are not fundamentally different, so that the 1985 outcomes will resemble the results for 1980. Although we regard the 1985 sample as really different from the 1980 one, consisting of other individuals at another point in time, we create subsamples, some of which are altogether different from the aggregate samples in the sense that a much lower or higher percentage of the households owns a car.

We divide both the 1980 and 1985 samples in 4 parts, each consisting of the same number of households. The first part (group 1) contains the 25% lowest household incomes (not per capita), the second part (group 2) the next 25% incomes, and so forth. We now expect the first group to have a low car ownership rate and the last group to have a high rate. The subsample ownership predictions, obtained by using (12), are listed in Table 6.

— see here Table 6 —

As we expected the first group has a low ownership rate. There is a striking increase in the rate of this group from 1980 to 1985, and, to a lesser degree, also in the second group. It seems that the increase in private car ownership from 1980 to 1985 is mainly the result of low income households being more prone to car ownership in 1985. The model predicts most of the increase for these two groups, but not all of it. For the two higher income groups, where there is less change, the model performs well in 1980 and 1985.

The predictions of subsample mileage can be found in Table 7.

— see here Table 7 —

The methods used are the same as in (20) and (21). If we use method 3 (equation (21)) we have to predict which household owns a car. This is done by applying the decision rule of section 4.2, where the cut–off point P_c is set by the predicted group car ownership rates of Table 6. This way we do not assign ownership to precisely the same set of households as in predicting mileage in the whole sample by method 3. In case of this method the weighted average of the group mileages therefore does not have to be equal to the mileage reported in Table 5.

For 1980 the mileage of the first two groups is overestimated, and that of the higher income groups slightly underestimated, resulting in a small overall overestimating of the outcome. For 1985 the predicted mileage is too high for all groups, but the overestimation is less in the higher income groups. In 1985 the actual mileage of the first group is slightly higher, that of the other groups lower than in 1980. The model predicts an increase for the first group, but also for the second, and a decline for the third and fourth group. Both in 1980 and 1985 the model reproduces the position of the groups relatively to each other rather well. It does not matter much if we use method 2 or 3; thus using our ownership predictions does not worsen the mileage predictions.

We resume by saying that the quality of the subsample predictions is not much less that that for the whole sample, definitely better than those for individual households.

Table 6: Subsample private automobile ownership

1980	number of households with one private car	
group:	actual	predicted
1	187 (32.4%)	205 (35.5%)
2	388 (67.2%)	401 (81.8%)
3	470 (81.5%)	472 (81.8%)
4	500 (86.7%)	512 (88.7%)
total	1545 (66.9%)	1590 (68.9%)

1985	number of households with one private car	
group:	actual	predicted
1	333 (58.8%)	293 (50.0%)
2	448 (76.5%)	431 (73.5%)
3	484 (82.6%)	470 (80.2%)
4	526 (89.8%)	527 (89.9%)
total	1791 (76.4%)	1721 (73.4%)

Table 7: Subsample annual private mileage

1980	mileage per private car (x 1000 km)		
group:	actual	predicted by method 2	predicted by method 3
1	10.818	12.148	12.106
2	12.975	13.323	13.298
3	14.478	14.268	14.261
4	15.316	15.137	15.097
total	13.860	14.055	14.009

1985	mileage per private car (x 1000 km)		
group:	actual	predicted by method 2	predicted by method 3
1	10.867	13.195	13.302
2	11.788	13.709	13.919
3	12.227	14.052	14.083
4	14.591	14.978	14.980
total	12.552	14.078	14.184

5. Re–estimation

When we apply 1980 estimates to 1985 data, the model correctly predicts the direction of change but understates its extent: ownership rates are underestimated, mileage is overestimated, Re–estimation of the parameters on the new data shows what parameters have changed.

– see here Table 8

The estimates for the two surveys are given in Table 8. The first two columns refer to data sets A and B, households without a car or with one private car; these estimates, in the first columns, have been used in the predictions reported above. The second set of estimates is based on the enlarged sample which includes business cars, category C. The elasticities of mileage in respect of income and household size and the effect of age increase substantially when business cars are added to the sample, and the farmer household and woman driver dummies also become more effective. But the fit becomes very much worse, as the variance of the disturbance of 'permanent' private mileage is increased almost fourfold. This makes us suspicious of the private mileage data of business cars, and has led us to prefer the estimates from data sets A and B for the predictions.

A major conclusion in support of the model is that both sample surveys do yield much the same parameter estimates, and that this also holds for the variances of the disturbances. The main change occurs in the threshold of private mileage. The estimated threshold for private driving has declined from 9.732 to 8.614 thousand kilometers per year. In view of the high precision of these estimates, this is a significant change. We can try to find out what has caused it by re–estimating the model in which the threshold is endogenous.

– see here Table 9 –

In Table 9 we see that the estimates of the y_i–equation are still more or less the same in 1980 and 1985. In the threshold equation the sign of the age coefficient is reversed, which means that older people have a lower instead of a higher car ownership threshold. This slightly reduces the threshold. Other forces working in this direction are household size, where there are both an increase (in absolute terms) in the estimated coefficient and an increase in the average value of the regressor, and the increase in the constant term. Log income and urbanization work in the opposite direction.

Table 8. Parameter estimates

	data sets A & B 1980		data sets A & B 1985		data sets A, B & C 1980		data sets A, B & C 1985	
coefficients β of log mileage	n = 2308		n = 2344		n = 2553		n = 2529	
intercept	2.25	(5.5)	2.29	(4.7)	1.51	(5.6)	.82	(5.2)
log income	.25	(6.2)	.22	(4.6)	.32	(7.0)	.38	(7.9)
log size	.25	(6.2)	.24	(5.0)	.31	(7.0)	.38	(7.3)
age	-.02	(-5.1)	-.01	(-3.2)	-.04	(-6.2)	-.03	(-4.9)
DA	-.01	(-.4)	.06	(1.9)	.10	(2.1)	.04	(1.0)
DF	-.11	(-4.2)	-.06	(-2.0)	-.25	(-4.8)	-.20	(-3.9)
threshold γ	4.578	(116.0)	4.456	(90.8)	4.316	(125.4)	4.187	(99.3)
σ_V	.19		.18		.37		.38	
σ_W	.49		.53		.56		.53	
value of log-likelihood (not including constant)	-880		-930		-1237		-1095	

t–ratios in brackets. Income per equivalent adult, size in number of equivalent adults, age in five–year classes; DA farmer household dummy, DF woman driver dummy.

Table 9. Parameter estimates of variable threshold model: (datasets A & B)

	1980 dependent variable y_i		1980 dependent variable γ_i		1985 dependent variable y_i		1985 dependent variable γ_i	
regressor variables:	n = 2308				n = 2344			
intercept	1.39	(3.41)	11.28	(13.01)	1.50	(3.22)	11.00	(14.34)
log income	.38	(9.20)	-.66	(-8.55)	.38	(9.95)	-.67	(-8.44)
log size	.28	(6.21)	-.81	(-10.12)	.22	(5.87)	-.86	(-10.34)
age	-.05	(-7.89)	-.02	(2.87)	-.04	(-8.41)	-.01	(-2.99)
DA	-.17	(-3.34)			-.16	(-2.19)		
DF	-.22	(-4.25)			.31	(-3.83)		
urbanization			.06	(4.73)			.14	(3.71)
σ_v	.11				.08			
σ_w	.48				.52			
σ_z			.75				.87	
value of log-likelihood (not including constant	-811				-827			

t-ratios in brackets. Income per equivalent adult, size in number of equivalent adults, age in five-year classes; DA farmer household dummy, DF woman driver dummy; urbanization in classes from rural to highly urban; z disturbance term in threshold equation.

6. Conclusions

We find that the overall aggregate levels of 1980 are fairly well described by the model, and those of 1985 fairly well predicted. If we look at the differences between the two samples, the model predicts their direction correctly, but understates their extent. The performance of the model in describing and predicting individual behaviour is however poor. The post sample validation is a success, in the sense that the 1985 sample results are almost as well described as the 1980 sample used in the model's calibration; but we must admit that the model fails to describe or predict individual behaviour at all correctly. It is not clear whether a much better performance at the individual level can be reasonably demanded from the present type of relatively simple economic model. Re–estimating the model for 1985 shows that the estimates are relatively stable. This clears the way for making predictions for future years. The decrease in the estimate of the constant threshold may explain the direction of the prediction errors for 1985. In predicting for 1985 we sticked to the 1980 threshold, and this will result in low values for car ownership and in overestimating mileage (because there are more low mileage households in case of a lower threshold). Re–estimating the variable threshold model showns that the decline of the threshold may be attributed to trends in household size, and to changes in the sensitivity to log size, age and the constant.

References

- Aitchison, J. and J.A.C. Brown (1957). The Lognormal Distribution, Cambridge, Cambridge University Press.

- Heckman, J.J. (1979), Sample selection bias as a specification error, Econometrica 47: 153–161.

- Jong, G.C. de and J.S. Cramer (1987), A censored regression model of private car use, Report AE 9/87, Institute of Actuarial Science and Econometrics, University of Amsterdam.