INSTITUTE OF ACTUARIAL SCIENCE & ECONOMETRICS

REPORT AE 9/87

A CENSORED REGRESSION MODEL OF PRIVATE CAR USE

G.C. de Jong & J.S. Cramer

University of Amsterdam
Faculty of actuarial Science & econometrics

Title: **A CENSORED REGRESSION MODEL OF PRIVATE CAR USE**

Author: G.C. de Jong & J.S. Cramer

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

Date: 1–9–1987

Series and number: Report AE 9/87
(Revision and extension of Report AE 4/87)

Pages: 26

Price: no charge

JEL Subject Classification: 211/921/933

Keywords: Censored regression model/car ownership/car use

Abstract. We develop a model for the joint determination of private car ownership and private car use by households. To own a private car is not worth the cost unless permanent annual mileage exceeds a certain minimum; observed annual mileage however differs from permanent mileage because of transitory factors. This model calls for a nonzero Tobin threshold and for an unusual distribution of the disturbances. We estimate its parameters from the 1980 Dutch household budget survey and obtain satisfactory results. The threshold is about 7500 kilometers per year, and the overall income elasticity of mileage 0.5. We also estimate a variable threshold model. In this case the overall income elasticity is slightly higher. The estimates of the variable threshold model confirm the old idea which says that the higher incomes will have a lower threshold.

The research reported here was supported by the Netherlands Organization for the advancement of pure research (ZWO); file number 46.128. We express our thanks to Geert Ridder for some very enlightening remarks and to Hermance Mettrop who did the typewriting.

CONTENTS

1.	Introduction	1
2.	The data	2
3.	Preliminary exercises	4
4.	A proper model	7
5.	Extension to business cars	9
6.	The income elasticity of private car use	13
7.	A variable threshold model	15
	References	20
appendix A.	Definitions of some variables	22
appendix B.	Another route to the loglikelihood function of section 4	25

1. Introduction

While there are many models of car ownership – see Mogridge (1983) or Tanner (1984) – analyses of car use are rare. With the exception of Train (1986) these car use models moreover explain or predict the number of trips, not the annual mileage or distance driven (see Daly and van Zwam, 1981). Yet mileage rather than the number of trips is the decisive variable when it comes to pollution, congestion, or the number of road traffic casualties. In this paper we present a model for private car use as measured by annual mileage. The use variable is determined jointly with private car ownership in a censored regression model. This approach is quite different from that of Train (1986), who derives the joint determination of car ownership and mileages from the maximization of a single indirect utility function. We learned of his work only after our own analyses had been completed.

The present type of model is due to Tobin (1958). Our model differs from the original Tobit in the treatment of the threshold value, which is nonzero, and in the specification of the disturbances. The principal regressand is y_i, the logarithm of the annual number of kilometers (x 100) driven by household i for private purposes. This includes travel to and from work, but not business or professional use. If household i does not dispose of a car, y_i is not observed; it is a censored or limited dependent variable.

A household may have a car at its disposal on either of two counts. It may have a business car, which is necessary for professional or business purposes but also available for private use, or it may own a private car. In the former case the household has no choice in the matter of the car's presence. Note that the definition of a business car is independent of the financial and fiscal arrangements in respect of costs, which merely affect disposable income. If the household has no business car, the option to acquire a private car represents a genuine choice. We assume that this choice depends on the permanent mileage y_i^* exceeding a certain threshold value. In view of its large fixed costs, car ownership is not worthwhile at a lower annual mileage, and the threshold must be substantial. This is the first difference with the Tobit model, which has a zero threshold. The second difference is that permanent mileage y_i^* is not itself observed. It is a long–term variable that enters into the consumer's decisions, exactly like Friedman's permanent income (Friedman, 1957) from which it takes its name. Actual observed mileage y_i that is in fact realized once a car has been acquired may differ from y_i^* by all sorts of transitory effects, including errors of observation.

After a brief description of the data in section 2 we present two models of private car use and ownership in sections 3 and 4. The model is extended to business cars in section 5, and we discuss the income elasticity of car use in section 6. In section 7 we generalize the model of section 4 by treating the threshold as a dependent variable.

2. The data

We estimate all models from the same data set by means of Ridder's general Maximum Likelihood programme GRMAX (Ridder, 1982; Ridder and Bekkering, 1986). This data set has been taken from the Dutch national budget survey of 1980 (CBS, 1982, 1985).

The survey covers 2859 households and provides a wealth of information about background variables, consumption patterns, and about car ownership and private car use. A combination of several variables is used to identify business cars, as is explained in appendix A. Annual mileage is recorded for one car only, and in the case of multiple car ownership this can not be identified; households with both business and private cars had therefore to be discarded. Households with two or more private cars were also disregarded because other research (Dix c.s., 1983) has shown that their behaviour is different from households with only one private car. Some households were omitted because of insufficient or inconsistent information on other variables. Table 1 shows the composition of the total sample by car ownership categories, and indicates the subsamples that have been used in estimation. We shall label the observations in index sets as follows:

o no car (n = 763)
p private cars (n = 1545)
b business cars (n = 245)

– see here Table 1 –

The car use variable y_i is the logarithm of the number of kilometers (x 100), driven annually for private purposes. This includes travel between home and work. Table 2 shows the mean and variance of mileage (not logarithm of mileage) for private and business cars. The mean is much higher for private cars, confirming that there is endogenous selection with a substantial threshold at work.

– see here Table 2 –

Table 1. Composition of the sample
(numbers of households)

	total	omitted	used
Car ownership category:			
(o) no car	763	-	763
(p) one private car	1564	19	1545
Subtotal used in § 3, 4 and 7			2308
(b) one or more business car(s)	249	4	245
Subtotal used in § 5			2553
- two or more private cars	165		
- both business and private cars	106		
- no information on car ownership	12		
Total sample	2859		

Table 2. Annual mileage for private purposes (in kilometers)

	mean	s.d.
household with one private car only (n = 1564)	13860	7042
household with business cars only (n = 249)	7928	7385

Permanent car use y_i^* is commensurate with y_i, but it is a latent, not an observed variable.

Throughout the analysis we use the same set of six regressor variables x_i, selected by an exploratory OLS regression analysis of car use y_i among the 1545 households with only one private car (de Jong, 1987). These six variables are as follows:

CONS	unit constant, with intercept coefficient;
LNINCH	log of annual net household income per equivalent adult;
LNUMB	log of household size expressed in number of adult equivalents;
AGEH	age of head of household, by classes, measured in five-year intervals;
DA	(0,1) dummy for farmer head of household;
DF	(0,1) dummy for woman driver;

The exact definitions of these variables are given in appendix A.

3. Preliminary exercises

The simplest way of allowing for the fixed costs of private car ownership and the ensuing lower bound for private mileage is to modify the Tobit model by the introduction of a nonzero threshold, as in

$$y_i = x_i'\beta + u_i \quad \text{if} \quad y_i > \gamma, \tag{1a}$$

$$y_i \quad \text{not observed otherwise,} \tag{1b}$$

$$\gamma > 0, \tag{1c}$$

$$u_i \quad \text{censored i.i.d. } N(0,\sigma^2_u). \tag{1d}$$

The trouble with this model is that with straightforward Maximum Likelihood estimation the estimate of γ tends to infinity, as inspection of the likelihood function will show. The value of γ must therefore be fixed a priori, but if we do so we find that any sensible value is contradicted by the presence of a handfull of very low observed mileages. To fix γ, and then to delete the offending observations with $y_i \leqslant \gamma$, would constitute too much data trimming.

We account for these low observed mileages by the distinction of permanent and transitory components of car use. Permanent private car use y_i^* must exceed γ to induce private car ownership, but observed private car use y_i may fall short of the expected value of y_i^* (and of γ!) for a variety of transitory reasons, including reporting errors. We write this model as

$$y_i^* = x_i'\beta + v_i, \tag{2a}$$

$$y_i = x_i'\beta + u_i \quad \text{if} \quad y_i^* > \gamma, \tag{2b}$$

$$y_i \quad \text{not observed otherwise,} \tag{2c}$$

$$\gamma > 0, \tag{2d}$$

$$v_i \quad \text{i.i.d. } N(0,\sigma^2_v), \tag{2e}$$

$$u_i \quad \text{i.i.d. } N(0,\sigma^2_u). \tag{2f}$$

Up to a constant, the loglikelihood function of this model is

$$\log L = \sum_0 \log (1-\Phi_i^*) + \sum_p \log \Phi_i^*$$

$$-n_p \log \sigma_u - \frac{1}{2\sigma^2_u} \sum_p (y_i-x_i'\beta)^2, \tag{3a}$$

with

$$\Phi_i^* = \Phi \left(\frac{x_i'\beta-\gamma}{\sigma_v}\right), \tag{3b}$$

where $\Phi(.)$ denotes the standard Normal distribution function. We recall than o denotes the set of households without a car and p the households owning only one private car; n_p is the number in the latter subsample.

— see here Table 3 —

Table 3. Estimates of model (2)
(n = 2308; t-ratios in brackets)

elements of β, by variable:			
	CONS	1.34	(4.51)
	LNINCH	0.36	(11.45)
	LNUMB	0.33	(11.10)
	AGEH	-0.030	(-8.89)
	DA	-0.057	(-1.53)
	DF	-0.162	(-4.51)
γ		4.59	(189.82)
σ_v		0.28	
σ_u		0.50	

Maximization of this loglikelihood readily yields the parameter estimates of Table 3. The results are quite satisfactory, with high t−ratios and plausible values for most parameters. The threshold $\hat{\gamma}$ of 4.59 corresponds to 9850 kilometers per year, 70% of the mean of Table 2, and an altogether acceptable value *). The income elasticity is .36 and the household size elasticity .33. With increasing age, mileage declines at the rate of .6% for each year; farmers drive less (presumably because they live near their work), and so do woman drivers.

It is also gratifying to find that σ^2_u exceeds σ^2_v, as we must assume that the transitory disturbances are in some sense added to the variation of permanent mileage. But it is a weak point that this relation between the two disturbances is not elaborated. In so far as v_i enters into u_i, the latter is truncated, but this effect is in no way reflected in the distribution of u_i as specified above. We shall remedy this defect in the next section. As it stands, model (2) is weak, possibly even self−contradictory, and it should be regarded as an approximation to a better specification.

4. A proper model

We retain the main systematic features of model (2) in

$$y_i^* = x_i'\beta + v_i, \tag{4a}$$

$$y_i = x_i'\beta + u_i \quad \text{if} \quad y_i^* > \gamma, \tag{4b}$$

$$y_i \quad \text{not observed otherwise,} \tag{4c}$$

$$\gamma > 0. \tag{4d}$$

As before, permanent and actual private car use are governed by regression equations with identical systematic parts but different disturbances. The disturbance term of permanent car use is specified as

$$v_i \quad \text{i.i.d. } N(0,\sigma^2_v). \tag{5}$$

---

*) The t−ratio of $\hat{\gamma}$ of nearly 190 may seem absurdly large. It should however be noted that the corresponding t−ratio of the 9850 kilometers, not in logs, is only 41 − still large, but not at all inconceivable.

This disturbance of permanent mileage presumably reflects the effect of unobserved factors affecting household behaviour, including tastes, that have a fundamental, long–term character. This effect is likely to persist in actual car use, even though additional transitory disturbances occur here. We therefore write

$$y_i = y_i^* + w_i, \tag{6}$$

with

$$w_i \quad \text{i.i.d. } N(0,\sigma^2_w), \tag{7}$$

and w_i independent of y_i^* and hence of v_i. It follows from (4a), (4b) and (6) that

$$u_i = v_i' + w_i. \tag{8}$$

Because of (4c), v_i' is a truncated version of v_i of (5), or

$$v_i' \quad \text{i.i.d. } N(0,\sigma^2_v) \text{ if } v_i > \gamma - x_i'\beta, \tag{9a}$$

$$u_i \quad \text{not observed otherwise.} \tag{9b}$$

Note that, by (8), u_i is <u>not</u> independent of v_i.

By this specification the distribution of u_i is the convolution of a truncated Normal and a complete Normal distribution. This distribution has earlier turned up in the context of production theory (see Aigner, Lovell, and Schmidt, 1977), and its analytical form has been provided by Weinstein (1964). With a nonzero threshold value, as is here the case, we must use the density given by Stevenson (1980), that is

$$h(u_i) = \frac{1}{\sigma_u} \; Z\left(\frac{u_i}{\sigma_u}\right) \left[1-\Phi\left(\frac{\gamma^*}{\sigma_u\lambda} - \frac{(u_i+x_i'\beta-\gamma)\lambda}{\sigma_u}\right)\right] \left[1-\Phi\left(\frac{\gamma^*}{\sigma_v}\right)\right]^{-1}, \tag{10a}$$

with

$$\gamma^* = \gamma - x_i'\beta, \tag{10b}$$

$$\lambda = \sigma_v/\sigma_w, \tag{10c}$$

$Z(.)$ = the standard Normal density function,

and

$$\sigma^2_u = \sigma^2_v + \sigma^2_w. \tag{10d}$$

Upon making use of this expression, and some rearrangement of terms, the loglikelihood function is given by

$$\log L = \sum_0 \log (1-\Phi_i^*) - n_p \log \sigma_u$$

$$- \frac{1}{2\sigma^2_u} \sum_p (y_i - x_i'\beta)^2 + \sum_p \log (1-\Phi_i^{**}), \qquad (11a)$$

with

$$\Phi_i^{**} = \Phi \left\{ \frac{\gamma - x_i'\beta}{\sigma_u \lambda} - \frac{(y_i-\gamma)\lambda}{\sigma_u} \right\}, \qquad (11b)$$

A simpler and more straightforward representation is to treat the model as a case of sample selection of the type studied by Heckman (1979). We then start from a simple regression model for y_i, and treat y_i^* as the related variable that governs the selection of observed values. With the same substantive assumptions about the structure of the disturbances this leads of course to identically the same loglikelihood function, as is shown in appendix B *).

— see here Table 4 —

Maximum Likelihood estimation of the parameters of this model yields the results of Table 4. The values, and their interpretation, are not widely different from the values obtained in Table 3 by an approximate model, although all elasticities are reduced in absolute size. We postpone the discussion of the results until the end of the next section.

5. Extension to business cars

The present model is easily extended to the private use of business cars. For households owning a business car we specify private mileage y_i

$$y_i = x_i'\beta + u_i, \qquad (12a)$$

$$u_i \quad \text{i.i.d. } N(0,\sigma^2_u). \qquad (12b)$$

---

*) We owe this point to Geert Ridder.

Table 4. Estimates of proper model
(n = 2308; t-ratios in brackets)

elements of β, by variable:			
	CONS	2.25	(5.51)
	LNINCH	0.25	(6.19)
	LNUMB	0.25	(6.24)
	AGEH	-0.019	(-5.11)
	DA	-0.013	(-0.44)
	DF	-0.11	(-4.18)
γ		4.578	(115.99)
σ_v		0.19	
σ_u		0.53	

We thus impose the same car use regression with the same coefficients as for owners of private cars. Since there is no car ownership decision for this group of households, we only have to add

$$-n_b \log \sigma_u - \frac{1}{2\sigma^2_u} \sum_b (y_i - x_i'\beta)^2 \qquad (12)$$

to the loglikelihood function (11). We recall that the index set b refers exclusively to owners of a business car. Estimation now involves 2553 observations (see Table 1).

— see here Table 5 —

The results are shown in Table 5. All estimated coefficients have larger t—values than in Table 4, which is in keeping with the increase of some 10% in sample size. Except for the constant, they are also greater in absolute value. The same variables that influence the private mileage of private cars apparently affect the private mileage of business cars even more strongly. But the standard deviation of the disturbance increases substantially too, as the overall standard deviation σ_u goes up from 0.53 to 0.67 *). This is in keeping with the earlier observation of Table 2 that private mileage of business cars is widely dispersed. We suspect that this variable is not very accurately recorded in the survey.

The estimates of Table 5 are based on all the observations that we can use, and these are the estimates of the present model we accept. Their values are quite plausible. The short—term income elasticity of private car use is 0.32, and the elasticity in respect of family size is similar at 0.31. These values may seem rather low, but they reflect the behaviour of car owners only and do not take into account the income (or family size) effects on private car ownership itself. We remedy this defect in section 6.

Car use declines with age at 0.7 percent per year, farmers drive 10% less, and woman drivers 25% less. The threshold value for private car ownership is 7500 kilometers per year, with a t—ratio of 29.

---

*) The standard deviations of the constituent parts of u increase as well: σ_v of the permanent component from 0.19 to 0.37, and σ_w of the transitory component from 0.49 to 0.56. But the additional business car data contain no information about these separate terms, only about σ_u.

Table 5. Estimates of proper model, including business cars (n = 2553; t-ratios in brackets)

elements of β, by variable:			
	CONS	1.51	(5.62)
	LNINCH	0.32	(6.95)
	LNUMB	0.31	(7.01)
	AGEH	-0.036	(-6.24)
	DA	-0.10	(-2.09)
	DF	-0.25	(-4.78)
γ		4.316	(125.41)
σ_v		.37	
σ_u		.67	

6. The income elasticity of private car use

The effect of an overall increase of disposable income by, say, 10% consists of two parts, namely i) the response of households already using a car, and ii) the effect of increased car ownership. We shall assess both terms on the basis of the sample of households of 1980, that is on the assumption that this sample is a fair representation of Dutch households.

To begin with we revert to Table 1 and reproduce the composition of the sample, adding the average and total number of kilometers of private car use recorded.

— see here Table 6 —

In Table 5, the income elasticity of private car use for car users has been estimated at 0.32. We assume that this value holds for all car–owning households, including the categories that have been excluded from the analysis. The response of car users to an income increase of 10% is then

$$0.32 \times 0.10 \times 30126 = 964 \ (\times 10^3 \text{ km}). \qquad (13)$$

The increase of all incomes will also induce a number of households without a car to acquire a private car. By our model, the probability for household i to own a private car is

$$P_i = P(y_i^* > \gamma) = \Phi(\frac{x_i'\beta - \gamma}{\sigma_v}), \qquad (14)$$

and if income has increased by 10% it is

$$P_i^0 = \Phi(\frac{x_i'\beta + 0.032 - \gamma}{\sigma_v}). \qquad (15)$$

We can thus calculate the expected number of private car owners in the subsample of 2308 households owning no car or one private car as

$$m = \sum_{i \epsilon o,p} P_i, \qquad (16)$$

or, after income has gone up,

$$m^0 = \sum_{i \epsilon o,p} P_i^0. \qquad (17)$$

Table 6. Composition of sample and private car use in kilometers per year (main car of the household only)

	number	average*	total*
Car ownership category:			
(o) no car	763	0	0
(p) one private car	1564 1)	13.86	21677
Subtotal	2327		21677
(b) business car(s)	249 2)	7.93	1975
Two or more private cars	165	27.84	4594
Both business and private cars	106	17.74	1880
Subtotal	2847	10.58	30126
No information on car ownership	12		
Total	2859		

*) × 1000 kilometers per year

1) of which 19 deleted from analysis

2) of which 4 " " "

We find

$$m = 1578, \qquad m^0 = 1635.$$

The first value is close to the observed sample frequency of 1545, as it should *). The overall income increase of 10% thus leads to an increase in expected private car ownership by 57. But while we can calculate expected frequencies, we can not identify predicted car owners nor these new entrants. And as we can not identify these households, we can not calculate their annual mileage.

We may assume, however, that the permanent or envisaged car use of new entrants was just under γ before the increase in income, and just over γ afterwards. We therefore attribute the number of kilometers that correspond to γ to these households, and assess their mileage as

$$57 \times 7.500 = 428 \ (\times 10^3 \text{ kilometers}) \qquad (18)$$

We thus equate the transitory disturbance w_i to its expected value of zero, as is usual in projections.

It now follows from Table 6, (13) and (18) that total mileage of main cars for private purposes (in the subsample of 2847 households) goes up by 964 + 428 = 1392 ($\times 10^3$ kilometers), or by 4.62%, as a result of a 10% increase in income. The income elasticity of private car use is therefore approximately 0.46.

7. A variable threshold model

It stands to reason that the minimum mileage which justifies private car ownership will vary with income: only the rich can afford to own durable goods for which they have little use. This argument, which goes back all the way to Roos (1934), suggests that the threshold value will decline as income rises, basically because of the declining marginal

---

*) If we use the households with no or one private car in a model of the probability of ownership alone, the two would be identical, provided the model allows for an intercept constant. But we use a model that allows for information from the mileage data and moreover include business car in the sample used in estimation. For these reasons the observed and predicted frequencies may differ.

utility of income.

This leads to an obvious generalization of the model of section 4 whereby the constant threshold value γ is replaced by an endogenous variable γ_i. When we introduce a straightforward regression equation for γ_i, with income and other household characteristics as the regressors, the model becomes:

$$y_i^* = x'_{1i}\beta_1 + v_i, \tag{19a}$$

$$y_i = x'_{1i}\beta_1 + u_i \text{ if } y_i^* > \gamma_i, \tag{19b}$$

$$y_i \text{ not observed otherwise,} \tag{19c}$$

$$\gamma_i = x'_{2i}\beta_2 + z_i. \tag{19d}$$

The loglikelihood function of this model is identical to that of Nelson (1977) for an unobserved stochastic censoring threshold model. As in Nelson's model, all parameters are identifiable provided at least one element of x_{1i} is not included in x_{2i}.

For x_{1i} we use the same variables as before, but in x_{2i} we omit the dummy variables DA and DF, and add a new variable that is of interest in its own right, viz.

URBA degree of urbanization, measured in 6 categories, ranging from rural (1) to highly urbanized (6).

The stochastic structure of this variable threshold model is:

$$v_i - z_i = s_i \text{ i.i.d. } N(0, \sigma^2_v + \sigma^2_z), \tag{20a}$$

$$u_i = v_i + w_i \text{ if } s_i > -(x'_{1i}\beta_1 - x'_{2i}\beta_2), \tag{20b}$$

$$u_i \text{ not observed otherwise,} \tag{20c}$$

$$v_i \text{ i.i.d. } N(0, \sigma^2_v), \tag{20d}$$

$$w_i \text{ i.i.d. } N(0, \sigma^2_w), \tag{20e}$$

$$z_i \text{ i.i.d. } N(0, \sigma^2_z). \tag{20f}$$

As before v_i and w_i are the permanent disturbance and the transitory disturbance of mileage respectively, and z_i is the disturbance of the threshold value. We assume z_i, v_i and $w_i = u_i - v_i$ to be independent.

The loglikelihood function of this model is:

$$\log L = \sum_{0} \log (1-\Phi_i{}^{c}) - n_p \log \sigma_u - \frac{1}{2\sigma^2{}_u} \sum_{p} (y_i - x'_{1i}\beta_1)^2 + \sum_{p} \log \Phi_i{}^{c*}, \qquad (21a)$$

with

$$\Phi_i{}^{c} = \Phi \left[\frac{x'_{1i}\beta_1 - x'_{2i}\beta_2}{(\sigma^2{}_v + \sigma^2{}_z)^{\frac{1}{2}}}\right], \qquad (21b)$$

$$\Phi_i{}^{c*} = \Phi \left[\frac{x'_{1i}\beta_1 - x'_{2i}\beta_2 + \frac{\sigma^2{}_v}{\sigma^2{}_u}(y_i - x'_{1i}\beta_1)}{[(\sigma^2{}_v + \sigma^2{}_z)\sigma^2{}_u - \sigma^4{}_v]\sigma^{-1}{}_u}\right]. \qquad (21c)$$

Maximizing this loglikelihood function yields the results of table 7.

— see here table 7 —

All parameter estimates have high t–ratios. The coefficients of y_i have the same signs as in Table 4, but, except for the constant, they are all greater in absolute value. The threshold γ_i has opposite signs for the three major regressors (income, family size and age) which is what we would expect; it strengthens the car ownership effects of these variables. Urbanization raises the threshold value, and thus has a negative effect on car ownership.

In this model σ_v and σ_w take the values 0.11 and 0.48. In the model of section 4 σ_v was 0.19 and σ_w was 0.49, which is rather similar. The value of σ_z, the disturbance of the regression equation for the minimum mileage, is very large in comparison. We must have overlooked many variables that affect the threshold value.

We do not have an estimate for the threshold value in this model, but the mean threshold value is calculated at 8000 kilometers, not far from the value for the previous models.

Table 7. Estimates of variable threshold model
(n = 2308; t-ratios in brackets)

Elements of β_1 and β_2 by variable	Dependent variable			
	y_i		γ_i	
CONS	1.39	(3.41)	11.28	(13.01)
LNINCH	0.38	(9.20)	-0.66	(- 8.55)
LNUMB	0.28	(6.21)	-0.81	(-10.12)
AGEH	-0.046	(-7.89)	0.021	(2.87)
DA	-0.17	(-3.34)		
DF	-0.22	(-4.25)		
URBA			0.059	(4.73)
σ_u	0.50			
σ_v	0.11			
σ_z			0.75	

As in section 6 we can assess the income elasticity of private car use by simulating an overall income increase of 10%. The response of households already using a car is

$$0.38 \times 0.10 \times 30126 = 1145 \ (\times 10^3 \text{ km}). \quad (22)$$

m (see equation (16)) is calculated at 1542, which is very close to the observed sample frequency. m^0 (see equation (17)) now takes the value 1627. The expected number of new entrants is therefore 85.

If we combine this number with the mean threshold value in the new situation, we get

$$85 \times 7.500 = 638 \ (\times 10^3 \text{ km}). \quad (23)$$

Total mileage goes up by 1145 + 638 = 1783 (x 10^3 km), which is 5.92%.

About half of the increase of the income elasticity from 0.46 to 0.59 is due to stronger ownerships effects in the variable threshold model, the other half being the result of the higher estimate of the income elasticity for households already using a car. The ownership effects would have been larger if σ_z would have been smaller. Further investigation on the determinants of the threshold is needed.

References

- Aigner, D.J., C.A. Lovell, and P. Schmidt (1977); Formulation and Estimation of Stochastic Frontier Production Function Models; Journal of Econometrics 6: 21–37.
- Centraal Bureau voor de Statistiek (1982); Budgetonderzoek 1978; deel 1: methodologie; enquêtedocumenten; 's–Gravenhage, Staatsuitgeverij.
- Centraal Bureau voor de Statistiek (1985); Budgetonderzoek 1980; 's–Gravenhage, Staatsuitgeverij.
- Daly, A.J., and H.H.P. van Zwam (1981); Travel demand models for the zuidvleugelstudy; Cambridge systematics Europe b.v. and Dienst Verkeerskunde van Rijkswaterstaat.
- Dix, M.C., S.M. Carpenter, M.I. Clarke, H.R.T. Pollard, and M.B. Spencer (1983); Car Use, A Social and Economic Study; Oxford studies in Transport, Aldershot, Gower.
- Friedman, M. (1957); A theory of the consumption function, Princeton, Princeton University Press.
- Heckman, J.J. (1979); Sample selection bias as a specification error; Econometrica 47: 153–161.
- Jong, G.C. de (1987); Naar een model van autobezit en –gebruik, Een verkenning van gegevens uit het budgetonderzoek 1980; A&E note N2/87, Faculty of Actuarial Science & Econometrics, University of Amsterdam.
- Mogridge, M.J.H. (1983); The Car Market; London, Pion.
- Mood, A.M., F.A. Graybill, and D.C. Boes (1974); Introduction to the Theory of Statistics; Tokyo, McGraw–Hill, Kogakusha, Ltd.
- Nelson, F.D. (1977); Censored regression models with unobserved stochastic censoring thresholds; Journal of Econometrics 6: 309–327.
- Ridder, G. (1982); GRMAX: een algemeen Maximum Likelihood programma; Faculty of Actuarial Science & Econometrics, University of Amsterdam.
- Ridder, G. (with J. Bekkering) (1986); Manual GRMAX; Stichting voor Economisch Onderzoek, University of Amsterdam.
- Roos, C.F. (1934); Dynamic Economics, Bloomington Principia Press.
- Stevenson, R.E. (1980); Likelihood functions for generalized stochastic frontier estimation; Journal of Econometrics 13: 57–66.
- Tanner, J.C. (1984), Modelling the Car Market; Transport and Road Research Laboratory Report 1116, Crowthorne, Berkshire, Department of the Environment/of Transport.
- Tobin, J. (1958); Estimation of Relationships for Limited Dependent Variables; Econometrica 26: 24–36.
- Train, K. (1986); Qualitative Choice Analysis, Theory, Econometrics and an Application to Automobile Demand; Cambridge Massachusetts, the MIT Press.

– Weinstein, M.A. (1964); The Sum of Values from a Normal and a Truncated Normal Distribution; Technometrica 6: 104–105.

Appendix A Definitions of some variables

A1 Business car or private car?

The decision to classify cars as a business car or a private car depends on the answers of the respondents to the following two questions.

I) Whose name is mentioned on Part II of the registration certificate (deel II van het kentekenbewijs)? Possible answers are: a) a member of the household, b) a firm or institution other than c), c) a car lease company, d) someone else.

II) Do you need the car for your profession or your firm? Possible answers are: a) Yes, b) No.

If question I) is answered a) or d) and II by a), and if in addition more than half of annual mileage is for business or professional purposes, the car is classified a business car.

If question I) is answered b) or c) the car is always a business car.

All other cases are classified as private cars.

A2 Main car

Only registered are the kilometers of the main car of the household. This is the car that has the highest mileage in private use.

A3 Adult equivalents

The variables LNINCH and LNUMB depend on the number of adult equivalents as calculated from the following weights:

Table A1. Common market adult equivalent scale

age/sex	number of adult equivalents
0 and 1 years	.2
2 and 3 years	.3
4 and 5 years	.4
6 and 7 years	.5
8 and 9 years	.6
10 and 11 years	.7
12 and 13 years	.8
14 - 59 years, male	1
14 - 59 years, female	.8
older than 60	.8

(source: CBS, 1982)

In the data set we used, the total number of adult equivalents in the household was multiplied by 10.

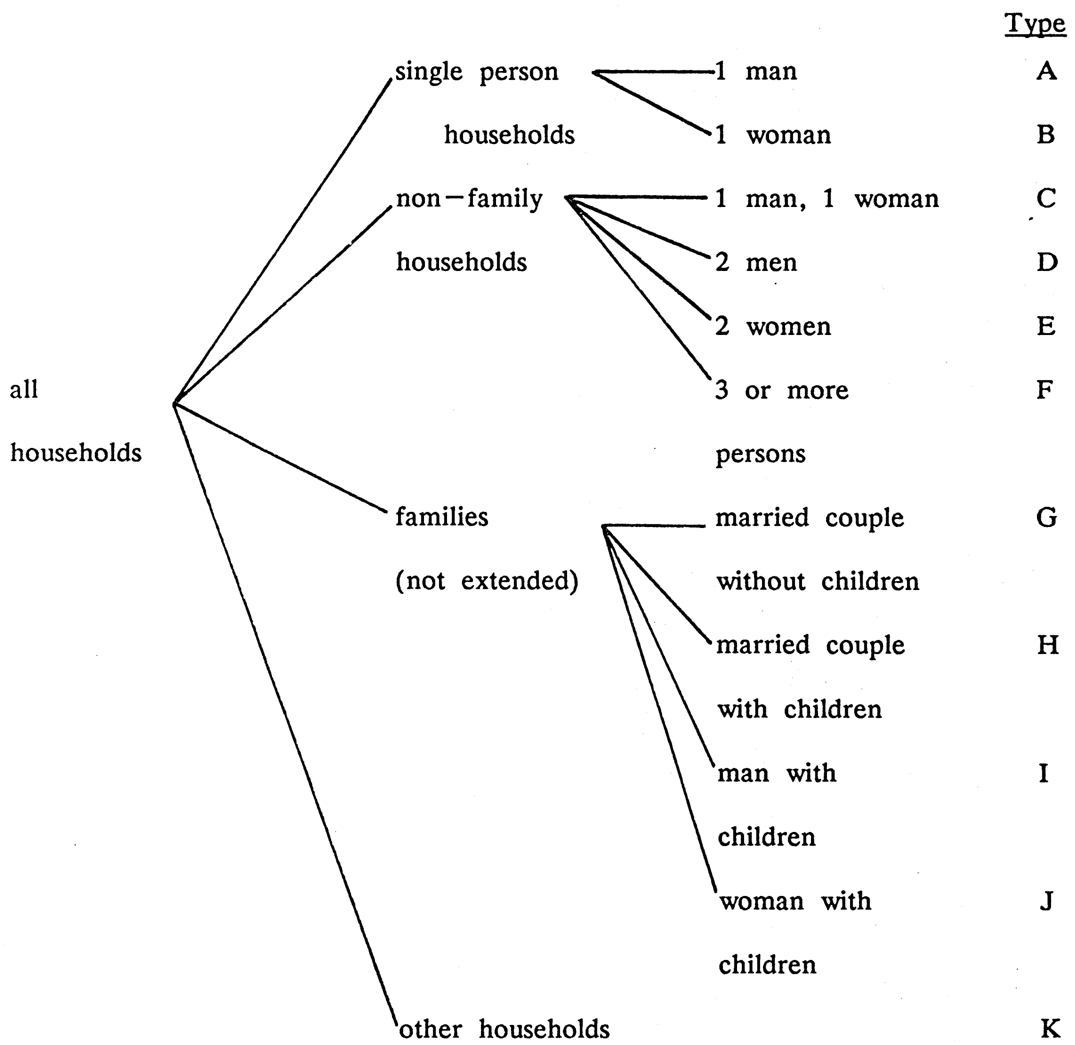
A4 Age classification

The variable AGEH is classified in the following way:
1) <20 years, 2) 20–24, 3) 25–29, 4) 30–34, ..., 12) 70–74, 13) $\geqslant 75$.

A5 Woman driver

Households can be classified as follows:

Figure A1: Types of households



There is a woman driver (DF=1) if the household is of type B, E, or J, and uses a car. In households of type J there is a possibility of children driving the car.

Appendix B Another route to the loglikelihood function of section 4

The contribution of non-owners to the likelihood function is:

$$P(\underline{y}^*_i \leqslant \gamma) = P(\underline{v}_i \leqslant \gamma - x'_i\beta) = 1 - \Phi\left(\frac{x'_i\beta-\gamma}{\sigma_v}\right), \qquad \text{(B1)}$$

as it was in section 4.

The contribution of the car-owners is:

$$P(\underline{y}_i\ ,\ \underline{y}_i^* > \gamma) = P(\underline{u}_i = y_i - x'_i\beta\ ,\ \underline{v}_i > \gamma - x'_i\beta) =$$

$$= P(\underline{v}_i > \gamma - x'_i\beta \mid \underline{u}_i = y_i - x'_i\beta)\ P(\underline{u}_i = y_i - x'_i\beta), \qquad \text{(B2)}$$

while in fact in section 4 we used:

$$P(\underline{y}_i\ ,\ \underline{y}_i^* > \gamma) = P(\underline{u}_i = y_i - x'_i\beta \mid \underline{v}_i > \gamma - x'_i\beta)\ P(\underline{v}_i > \gamma - x'_i\beta).$$

Using the formulas for a conditional bivariate normal distribution (see Mood, Graybill, and Boes, 1974, p. 167, 168) we can write for $P(\underline{v}_i > \gamma - x'_i\beta \mid \underline{u}_i = y_i - x'_i\beta)$:

$$1 - \frac{1}{(2\pi)^{\frac{1}{2}}\sigma_v\left[1-\left[\frac{\sigma_v}{\sigma_u}\right]^2\right]^{\frac{1}{2}}} \int_{-\infty}^{\gamma} \exp\left\{-\tfrac{1}{2}\left[\frac{t - x'_i\beta - \frac{\sigma^2_v}{\sigma^2_u}(y_i - x'_i\beta)}{\sigma_v\left[1-\left[\frac{\sigma_v}{\sigma_u}\right]^2\right]^{\frac{1}{2}}}\right]\right\} dt =$$

$$= 1 - \Phi\left[\frac{\gamma - x'_i\beta - \frac{\sigma^2_v}{\sigma^2_u}(y_i - x'_i\beta)}{\sigma_v\left[1-\left[\frac{\sigma_v}{\sigma_u}\right]^2\right]^{\frac{1}{2}}}\right] = 1 - \Phi\left[\frac{\gamma - x'_i\beta - \frac{\sigma^2_v}{\sigma^2_u}(y_i - x'_i\beta)}{\sigma_v\sigma^{-1}_u(\sigma^2_u - \sigma^2_v)^{\frac{1}{2}}}\right] =$$

$$= 1 - \Phi\left[\frac{(\gamma - x'_i\beta)(\sigma^2_u - \sigma^2_v) - (y_i - \gamma)\sigma^2_v}{\sigma_v\sigma_u(\sigma^2_u - \sigma^2_v)^{\frac{1}{2}}}\right] \qquad \text{(B3)}$$

The loglikelihood function now becomes:

$$\text{Log L} = \sum_{0} \log \left\{1-\Phi\left[\frac{x'_i\beta-\gamma}{\sigma_v}\right]\right\} - n_p \log \sigma_u - \frac{1}{2\sigma^2_u} \sum_{p} (y_i-x'_i\beta)^2$$

$$+ \sum_{p} \log \left\{1-\Phi\left[\frac{(\gamma-x'_i\beta)(\sigma^2_u-\sigma^2_v) - (y_i-\gamma)\sigma^2_v}{\sigma_v\sigma_u(\sigma^2_u-\sigma^2_v)^{\frac{1}{2}}}\right]\right\} . \qquad \text{(B4)}$$

Which is the same as (11a) on p.9.