7. A.B. Atkinson and J.E. Stiglitz. Lectures on Public Economics. McGraw-Hill, New York, 1980.
8. T.J. Wales and A.D. Woodland. Estimation of Consumer Demand Systems with Binding Non-Negativity Constraints. Journal of Econometrics, Vol. 21, 1983, pp. 263-85.
9. L.F. Lee and M.M. Pitt. Specification and Estimation of Demand Systems with Limited Dependent Variables. Department of Economics, University of Minnesota, Minneapolis, 1983.
10. A.S. Deaton. Demand Analysis. In The Handbook of Econometrics (Z. Griliches and M.D. Intriligator, eds.), North-Holland, Amsterdam, in preparation.
11. J.J. Heckman. Dummy Endogenous Variables in a Simultaneous Equations System. Econometrica, Vol. 46, 1978, pp. 931-61.

The views expressed are those of the author and should not be attributed to the World Bank, to its affiliated organizations, or to any individual acting on their behalf.

Publication of this paper sponsored by Committee on Passenger Travel Demand Forecasting.

Analysis of Automobile Ownership by Using a Divisive Hierarchical Technique

JOKE M. JAGER and WIM H. SCHELTES

ABSTRACT

A method of analysis of personal automobile ownership is presented that differs from the well-known aggregate and disaggregate methods. The analysis consists of two steps. First, a cluster-segmentation method is applied to data from the Dutch National Travel Survey. The results show that personal automobile ownership is mainly determined by personal net income, age, and sex. Second, a model has been specified that includes these factors. When income and age are accounted for, a structural difference in automobile ownership is shown between men and women. Furthermore, for the period studied (1979-1982) the results indicate that for the age group 65 and older, automobile ownership increased significantly, whereas for those 25 years and younger, it decreased. Advantages of the method are (a) the relative stability of the homogeneous population groups independent of accidental changes in the survey population and (b) insight into the relationship of automobile ownership with the most essential determining factors. Because of these advantages, the method presented can be used to improve both analysis and forecast of automobile ownership.

It has been widely recognized that travel behavior is strongly influenced by automobile ownership. This applies to mode choice as well as to trip frequency and daily mileage (1-3). Therefore numerous models of automobile ownership have been developed, both at aggregate and disaggregate levels (4). Some models are based on time-series data under the assumption that a certain saturation level exists, whereas others are disaggregate at the household level and are based on cross-sectional data (5,6). Because both aggregate and disaggregate methods suffer from a number of disadvantages (7-9), another method is applied in this paper.

Phase 1 of this work aims at finding those demographic and socioeconomic factors that influence personal automobile ownership most. The survey population is split into homogeneous population groups according to these most important factors. Personal automobile ownership is preferred here to household automobile ownership because models of travel behavior are usually specified at a personal level and because this analysis is part of a comprehensive transportation study. Phase 2 of this work focuses on the level of influence of these factors on automobile ownership and investigates trends in the development of automobile ownership in the homogeneous population groups.

DATA BASE

The data base used was Onderzoek Verplaatsingsgedrag (OVG), the Dutch National Travel Survey (10). It contains extensive information, both demographic and

socioeconomic data on the interviewees and data on the trips they made. Since the survey started in 1978, about 25,000 persons have been interviewed annually. In contrast to most other surveys, the interviews were held throughout the year. Because of the large number of questions, the home-interview technique was used. A trained interviewer registered the information concerning the household. The data on trips were recorded by the respondents (12 years old or older) themselves, usually during a 2- or 3-day period.

VARIABLES OF POTENTIAL IMPORTANCE

A whole range of demographic and socioeconomic variables correlate with automobile ownership. A portion of this range formed by the variables that were collected in the Dutch National Travel Survey has been analyzed for the relation with automobile ownership.

These characteristics of interviewees and their environment are listed alphabetically (the number of distinct classes is given in parentheses after each variable):

- Age (5)
- City size (3)
- Educational level (7)
- Employment status (5)
- Household income (6)
- Household size (6)
- Marital status (4)
- Personal income (6)
- Position in household (7)
- Province (13)
- Pattern of visit to office or school (6)
- Sex (2)

Note that a number of other possibly relevant variables (such as population density and public transport quality) were not included.

METHOD OF ANALYSIS

It is obvious that the method of analysis should belong to the group of multivariate techniques. From a mathematical point of view, the interviewees are observations in a multidimensional space formed by the selected demographic and socioeconomic variables and their distinct classes.

Clustered observations in this hyperspace indicate dominating variables or dependencies or both. A well-known technique to find these data concentrations is the cluster-segmentation analysis. Its mathematical robustness makes it easy to use. Several algorithms exist [see, e.g., studies by Everitt (11) and Späth (12)], most of them combining nearby observations to form clusters. However, dividing the data into separate groups is (in the case of many observations) preferable, as it often leads to better group distinction. These algorithms are called segmentation or divisive hierarchical techniques.

The likelihood-segmentation method is one of these. Its algorithm was developed by Hamerslag (1) by using the likelihood-estimation theory. When the number of clusters is equal to the number of observations (so each cluster contains only one element), the likelihood value is maximal. Each data combination causes the likelihood value to decrease, so the lowest value will be found when all observations are merged into one single cluster. Thus the value of the likelihood can be used to measure loss of information during the clustering process. All variables have discrete classes; for each variable, classes are combined and the information loss is calculated. The variable that shows the largest drop is (normally) the most significant.

The opposite is also valid: Dividing a data cluster will increase the likelihood value and the information level. Thus the data base will be split into several segments (subspaces), every segment being a class from the most discriminating variable (dimension). The obtained data groups can be analyzed in the same manner. Theoretically the process ends when all clusters contain only one observation. In practice, however, segmentation is stopped in an earlier phase.

The formula for calculating the decrease or increase of information (using an amount k of simultaneous criteria factors) is

$$d(X,Y) = \sum_k [(N_x\bar{X}_k)\ln(\bar{X}_k + eps) + (N_y\bar{Y}_k)\ln(\bar{Y}_k + eps) - (N_x\bar{X}_k + N_y\bar{Y}_k)\ln(\bar{U}_k)]$$

$$\bar{U}_k = (N_x\bar{X}_k + N_y\bar{Y}_k)/(N_x + N_y) \quad (1)$$

where

$d(X,Y)$ = information loss when group X and group Y are combined;
N_x, N_y = number of elements in groups X and Y, respectively;
$\bar{X}_k, \bar{Y}_k$ = average value of the observations in groups X and Y, respectively;
eps = small number to prevent ln(0) from occurring; and
k = number of criteria factors, used to measure the data distance (normally k = 1).

For more details, see work by Hamerslag (1).

SEGMENTATION RESULTS

In the Netherlands the minimum age for drivers is 18. To make certain that statistics are useful and understandable, all younger interviewees (and those with unstated answers) are eliminated from the 1979 data base, leaving 16,777 persons to analyze.

The first computer run showed that the variables personal income, position in household, and sex had a significant influence on automobile ownership. Although the information loss among these variables did not differ much, personal income was pointed out as most important, for the following reasons:

- It is an important indicator for potential expenditure, in contrast to both other variables mentioned. Fairly static characteristics such as sex and household position cannot easily be used in research applications like forecasting. By using income, continuous annual research is also possible.
- Research on travel behavior and in particular travel performance also showed personal income to be of great significance (1). Linking up to this might bear advantages in interpreting the results.

To maximize the information level, the data base was split up into its distinct income classes. As seen in Figure 1, low income and high income form quite homogeneous groups (small data dispersion). This indicates that there are no other important influences, so further analysis of these groups is not necessary.

In the second computer run, all income classes were analyzed separately. Age, position in household, and sex turned out to be the most discriminating factors. After all information loss was totaled for each class (this is allowed because of interdependencies), age was found to be the most important

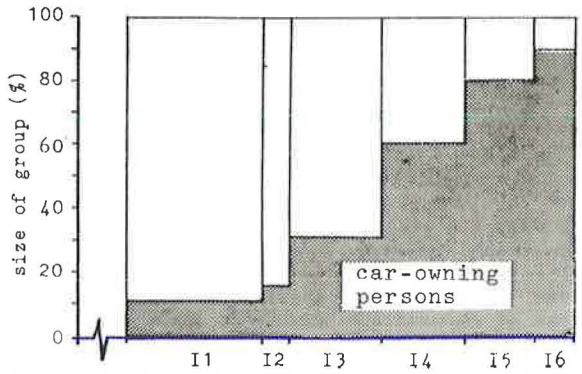

FIGURE 1 Relationship between automobile ownership and personal net income (source: Dutch National Travel Survey).

variable. Thus each income group was split into age classes. The effect of personal income and age on automobile ownership is shown in Figure 2. In most cases the peak level is located around the ages of 36 to 45.

In the third computer run the population groups formed were analyzed further. Depending on the group

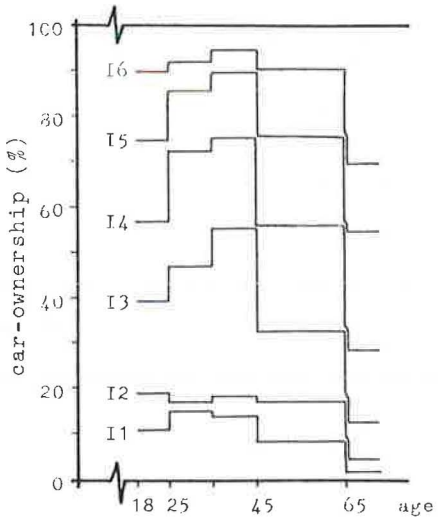

FIGURE 2 Automobile ownership in relation to personal net income and age (source: Dutch National Travel Survey).

characteristics, sex and position in household turned out to be the most important factors. New groups were created based on this knowledge. At this stage some groups did not contain enough interviewees to be analyzed in further detail, whereas the remaining ones had small data dispersion (homogeneous data). Therefore the analysis stopped.

In Figure 3 the results of each computer run are shown. Note that the population groups given are quite homogeneous, or too small to analyze.

CLASSIFICATION OF POPULATION GROUPS

As was demonstrated in Figure 3, the two most relevant factors to personal automobile ownership are personal net income and age. Another important factor is sex, which sometimes appears in a different form, such as position in the household or employment status.

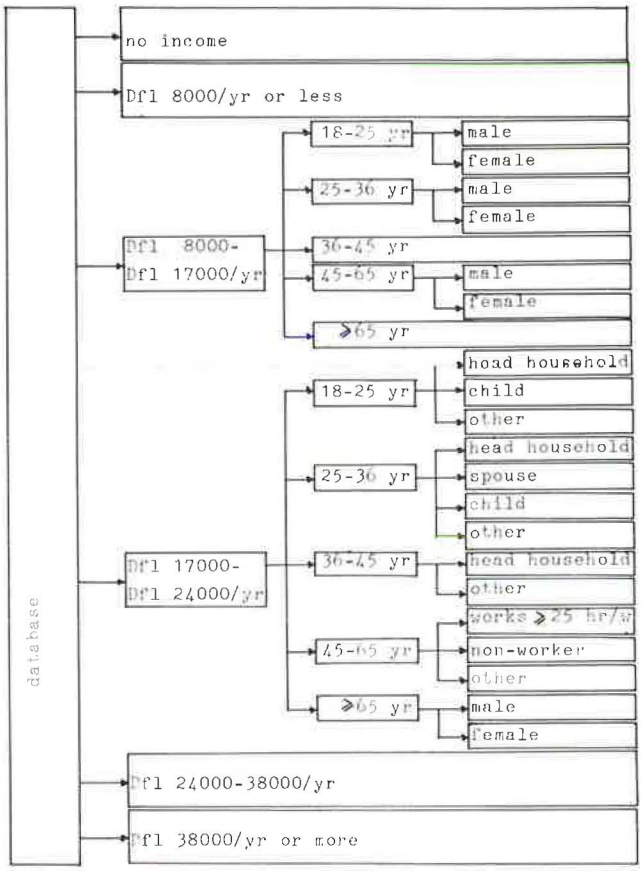

FIGURE 3 Segmentation results visually summarized.

It was therefore decided to classify the survey population according to personal net income, age, and sex. For each group the percentage of personal automobile ownership was calculated. This was done for the survey years 1979-1982, which span the turbulent period of the second oil crisis and the subsequent income stagnation in the Netherlands. The classification is given in Table 1.

TABLE 1 Classification of the Relevant Factors for Automobile Ownership

Personal Income (Dfl x 1,000)	Age (yr)	Sex	Year
No income (I1)	18-25 (A1)	Male (S1)	1979 (Y1)
0-8 (I2)	25-36 (A2)	Female (S2)	1980 (Y2)
8-17 (I3)	36-45 (A3)		1981 (Y3)
17-24 (I4)	45-65 (A4)		1982 (Y4)
24-38 (I5)	≥65 (A5)		
≥38 (I6)			

Source: Dutch National Travel Survey.

COMPOSITION OF SURVEY POPULATION

The composition of the survey population in 1982 according to age and income is given in Figures 4 and 5. As expected, there is a somewhat higher percentage of women than men in the group over 65 years.

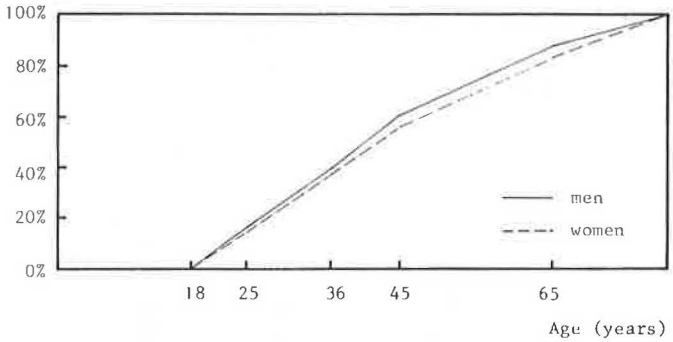

FIGURE 4 Cumulative age distribution for 1982 (source: Dutch National Travel Survey).

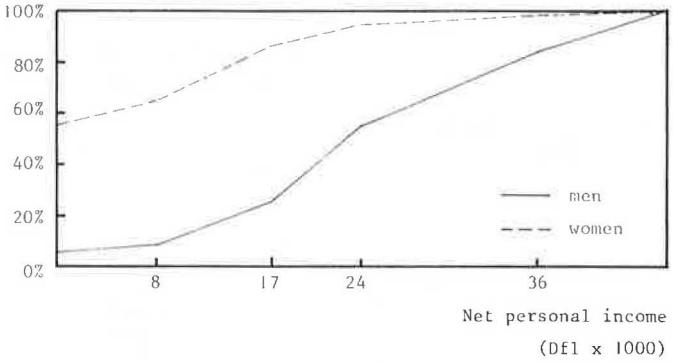

FIGURE 5 Cumulative income distribution for 1982 (source: Dutch National Travel Survey).

The distribution of income by sex shows great differences between men and women: about 65 percent of the women have a personal net income of less than Dfl 8,000 per year versus 7 percent of the men. The percentage of men and women in the annual survey population is found to be stable, ranging from 50.16 to 49.25 percent men. The composition of the survey sample can have important consequences for aggregate data such as personal automobile ownership.

Table 2 gives the percentage of personal automobile ownership for 1979-1982 for both the total population and men and women separately. It can be seen that a 1 percent increase in the proportion of women in the survey may cause a drop of more than 0.5 percent in automobile ownership for the total group.

TABLE 2 Percentage of Automobile Ownership

Year	Total Group	Men	Women
1979	41.9	68.3	15.4
1980	43.2	69.9	16.6
1981	43.4	70.7	16.5
1982	42.2	69.5	15.8

Source: Dutch National Travel Survey.

MODEL FORMULATION

To analyze the influence of the most relevant factors (income, age, and sex) on automobile ownership and also the development of automobile ownership over the years 1979-1982 for the distinct groups, a multiplicative model was used:

$$NUMCO(i,j,k) = C * NPERS(i,j,k) * I(i) * A(j) * S(k) \quad (2)$$

where

NUMCO(i,j,k) = number of automobile-owning persons in groups i, j, and k;
C = constant;
NPERS(i,j,k) = total number of persons in groups i, j, and k;
I(i) = coefficient of income class i;
A(j) = coefficient of age class j; and
S(k) = coefficient of sex class k.

The estimation was carried out by using a weighted multiproportional Poisson estimation method (13,14). This was done both for marginal factors and for the simultaneous inclusion of all factors.

The results of the general estimations are shown in Tables 3 and 4. Table 3 indicates a strong increase in automobile ownership with increasing income as well as a relation to age (lowest in the group of 65 and older, highest in the group 36 to 45). As shown, women reach only about 22 percent of the automobile-ownership level of men.

Table 4 indicates, however, that the simultaneous inclusion of all relevant factors leads to somewhat less marked differences per factor, which is caused by correlation between the factors. The difference between men and women is sharply reduced with the inclusion of the income factor. This is because of the previously mentioned wide difference in personal income distribution between men and women.

TABLE 3 Results of the General Estimation of Automobile Ownership with Marginal Factors

Factor	Class	Weight	Automobile Ownership (%)	Coefficient
Income (Dfl)	Constant	–	–	C(I) = 0.11
	No income	16,513	10.9	I1 = 1.00
	0-8,000	4,471	16.9	I2 = 1.55
	8,000-17,000	10,844	30.8	I3 = 2.82
	17,000-24,000	10,320	62.1	I4 = 5.69
	24,000-38,000	8,563	80.2	I5 = 7.35
	≥38,000	4,402	89.0	I6 = 8.20
Age (yr)	Constant	–	–	C(A) = 0.20
	18-25	8,219	32.4	A1 = 1.60
	25-36	13,959	50.9	A2 = 2.51
	36-45	8,930	54.9	A3 = 2.70
	45-65	16,290	42.2	A4 = 2.08
	≥65	7,715	20.3	A5 = 1.00
Sex	Constant	–	–	C(S) = 0.16
	Men	26,561	70.1	S1 = 4.44
	Women	28,552	15.8	S2 = 1.00

Source: Dutch National Travel Survey.

TABLE 4 Results of the General Estimation of Automobile Ownership with Simultaneous Inclusion of All Factors

Factor	Class	Weight	Automobile Ownership (%)	Coefficient
Constant	–	–	–	C = 0.052
Income (Dfl)	No income	16,513	10.9	I1 = 1.00
	0-8,000	4,471	16.9	I2 = 1.42
	8,000-17,000	10,844	30.8	I3 = 2.26
	17,000-24,000	10,320	62.1	I4 = 3.25
	24,000-38,000	8,563	80.2	I5 = 3.75
	≥38,000	4,402	89.0	I6 = 4.17
Age (yr)	18-25	8,219	32.4	A1 = 1.89
	25-36	13,959	50.9	A2 = 2.17
	36-45	8,930	54.9	A3 = 2.21
	45-65	16,290	42.2	A4 = 1.88
	≥65	7,715	20.3	A5 = 1.00
Sex	Men	26,561	70.1	S1 = 2.18
	Women	28,552	15.8	S2 = 1.00

Source: Dutch National Travel Survey.

It should be noted, however, that when the effects of income and age have been taken into account, the percentage of automobile ownership by women is less than half that of men. The difference in automobile ownership between men and women is least for middle-aged persons and the highest income group.

Similar estimations were performed with the model including a factor for the distinct survey years. No overall trends for automobile ownership were found.

SEPARATE ESTIMATION BY SEX

In Tables 5 and 6 separate estimations are presented for men and women. These results are shown in Figure 6. It is found that the relative effect of the distinct factor classes is much greater for women than for men (although absolute automobile ownership for women is much lower in any case).

The age effect for women is especially strong; there is a sharp drop in automobile ownership for the group over 65 years. Similar estimations have been made with the inclusion of a year factor for the total group and for separate income and age groups.

It was found that there are significant changes in the automobile ownership of the oldest age group; automobile ownership increased in 1982 compared with previous years. This is sometimes referred to as a "generation effect." In the youngest age groups, automobile ownership decreased. There are no significant changes in automobile ownership by income group.

TABLE 5 Results of the Simultaneous Estimation of Automobile Ownership for Men

Factor	Class	Weight	Automobile Ownership (%)	Coefficient
Constant	–	–	–	C = 0.134
Income (Dfl)	No income	1,254	22.1	I1 = 1.00
	0-8,000	677	32.6	I2 = 1.50
	8,000-17,000	4,950	47.6	I3 = 2.37
	17,000-24,000	8,047	69.9	I4 = 3.16
	24,000-38,000	7,532	84.5	I5 = 3.65
	≥38,000	4,101	91.9	I6 = 4.06
Age (yr)	18-25	4,122	47.1	A1 = 1.62
	25-36	6,959	81.9	A2 = 1.83
	36-45	4,569	86.8	A3 = 1.83
	45-65	7,455	75.6	A4 = 1.64
	≥65	3,456	39.7	A5 = 1.00

Source: Dutch National Travel Survey.

TABLE 6 Results of the Simultaneous Estimation of Automobile Ownership for Women

Factor	Class	Weight	Automobile Ownership (%)	Coefficient
Constant	–	–	–	C = 0.024
Income (Dfl)	No income	15,259	9.9	I1 = 1.00
	0-8,000	3,794	14.1	I2 = 1.37
	8,000-17,000	5,894	16.7	I3 = 2.05
	17,000-24,000	2,273	34.5	I4 = 3.62
	24,000-38,000	1,031	48.8	I5 = 4.85
	≥38,000	301	57.5	I6 = 5.98
Age (yr)	18-25	4,097	17.7	A1 = 3.75
	25-36	7,000	20.0	A2 = 4.85
	36-45	4,361	21.5	A3 = 5.63
	45-65	8,835	14.1	A4 = 3.71
	≥65	4,259	4.5	A5 = 1.00

Source: Dutch National Travel Survey.

CONCLUSIONS

The method applied in this paper is a useful alternative to the generally used aggregate or disaggregate methods. It leads to distinguishable population groups, the behavior of which with respect to automobile ownership is different. The groups themselves are homogeneous because they are differentiated on the basis of the most influential factors that were documented in the data base.

With data from the Dutch National Travel Survey, it was found that automobile ownership is mainly determined by personal net income, age, and sex. It is essential to distinguish between the right influential factors, because changes in the composition of the (survey) population in relation to these factors will cause significant changes in the automobile ownership of the aggregate group, whereas other factors will have much less effect.

Analysis of a multiplicative model of automobile ownership versus income, age, and sex for four consecutive years has clarified these relationships and their interdependencies. Because there is a strong relation between automobile ownership and income, it can be expected that future income development will influence automobile ownership as well.

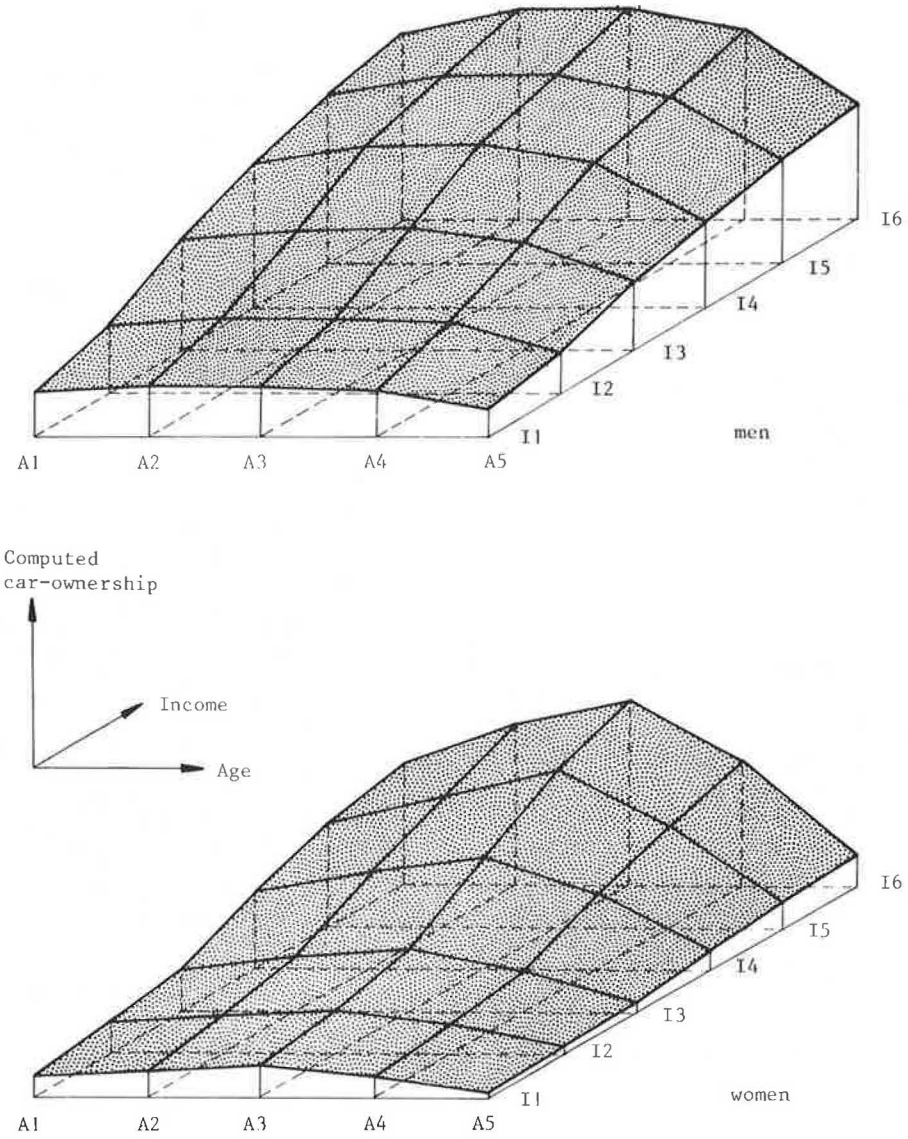

FIGURE 6 Results of estimated automobile ownership coefficients for income and age by sex [CAROW% = C x I(i) x A(j)].

Furthermore, continuing generation effects are expected, which means that the automobile ownership in the group older than 65 will increase. Although the relative effect is especially strong for women, this will only have a limited effect on total automobile ownership, because of the small absolute values. The same applies to developments such as increasing participation of women in the labor force, which would influence their personal income.

Automobile ownership in the age group of 18-25 has decreased slightly. This may have been caused by unemployment, but further analysis will be necessary to confirm this.

It is remarkable that there is a structural difference in automobile ownership between men and women, even in the middle-aged and high-income categories. This can most probably be explained by the position in the household and employment status, which are not very likely to change rapidly in the near future.

The consequences of the established relationships for forecasting purposes are interesting and a number of exercises for possible developments can be performed, assuming that the established relationships remain valid. Because the population group that will be 18-25 years old in the year 2000 has already been born, the group size is known. The same applies to the other age groups, depending on their life expectations. By using the relative group sizes of the survey sample, the automobile ownership in, for instance, the year 2000 (ignoring any income effects) can be calculated. For the total population that is 18 years or older the result of such a calculation shows 48 percent automobile ownership (76 percent for men and 20 percent for women).

A similar exercise is to calculate automobile ownership for several scenarios of future economy (without age effects). An increase of 1 percent per year in personal net income for all income groups may lead to 46 percent automobile ownership (75 percent for men and 18 percent for women) in the year 2000.

If women obtained exactly the same income distribution as men have at present, total automobile ownership would rise to about 50 percent (women would reach 32 percent automobile ownership). This, how-

ever, assumes that there still is a structural difference in automobile ownership between men and women. (This could be caused by factors that are assumed to remain unchanged, e.g., position in the household.) Other exercises such as calculation of combined effects could of course be performed as well.

A time-trend calculation based on historic data or a backward calculation in time using the survey data could be applied to confirm the validity of these coefficients for forecasting purposes.

ACKNOWLEDGMENTS

The authors wish to express their thanks to Rudy Hamerslag of Delft University of Technology for his assistance with the application of the segmentation method and to Velibor Vidakovic of the Municipality of Amsterdam, who provided many useful comments on earlier versions of the paper.

REFERENCES

1. R. Hamerslag. Population Groups with Homogeneous Travel Performance by Car, Public Transport and Bicycle. Department of Civil Engineering, Delft University of Technology, Netherlands, 1983.
2. C.G.B. Mitchell. The Influence of the Car on Personal Travel. TRRL Report SR 681. U.K. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, 1981.
3. J. Supernak. Car Availability Versus Car Ownership Modelling: Theoretical Discussion and Empirical Findings. Presented at World Conference on Transport Research, Hamburg, West Germany, 1983.
4. Car Ownership and Use. Organization for Economic Cooperation and Development, Paris, France, 1982.
5. Models for Car Ownership and Licence Holding. Zuidvleugel Report 9. Rijkswaterstaat and Cambridge Systematics Europe, The Hague, Netherlands, 1981.
6. J.C. Tanner. Saturation Levels in Car Ownership Models: Some Recent Data. TRRL Report SR 669. U.K. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, 1981.
7. C.J. Ruijgrok. Disaggregate Choice Models: An Evaluation of New Developments in Modelling Travel-Demand and Urban Systems. Presented at Transportation Planning Colloquium, The Hague, Netherlands, 1979.
8. J. Supernak. Toward a Better Understanding of Transportation Related Choices. Presented at Transportation Planning Colloquium, Zandvoort, Netherlands, 1983.
9. J. Supernak. Transportation Modeling: Lessons from the Past and Tasks for the Future. Transportation, Vol. 12, No. 1, 1983.
10. F.W.M. Hendrikx. Framework and Execution of the Dutch National Travel Survey (Opzet en Uitvoering van het Onderzoek Verplaatsingsgedrag). Stichting Postacademiale Vorming Verkeerskunde, Delft, Netherlands, 1982 (in Dutch).
11. B. Everitt. Cluster Analysis. Heinemann Educational Books, Inc., Portsmouth, N.H., 1980.
12. H. Späth. Cluster Analysis Algorithms (English translation). Queen Elizabeth College, University of London, United Kingdom, 1980.
13. H. Derriks. Computer Documentation of the Weighted Multiproportional Poisson Estimation Program. Department of Civil Engineering, Delft University of Technology, Netherlands, 1984.
14. R. Hamerslag, J.P. Roos, and M. Kwakernaak. Analysis of Accidents in Traffic Situations by Means of the Weighted Multiproportional Poisson Model. In Transportation Research Record 847, TRB, National Research Council, Washington, D.C., 1982, pp. 29-36.

Publication of this paper sponsored by Committee on Traveler Behavior and Values.