Q4

POLICY ANALYSIS USING SAMPLE ENUMERATION - an application to car ownership forecasting

A Daly,
Cambridge Systematics Europe

PTRC Summer Annual Meeting 1982
University of Warwick

POLICY ANALYSIS USING SAMPLE ENUMERATION- AN APPLICATION TO CAR OWNERSHIP FORECASTING

A.J. Daly

Cambridge Systematics Europe b.v., The Hague

1. INTRODUCTION

The intention of this paper is to describe the sample enumeration method of forecasting with disaggregate travel demand models and to illustrate its use in an application in a study of car ownership.

Sample enumeration has been in use in transportation planning for a number of years (Atherton and Ben-Akiva, 1976). The technique has, however, not been widely used despite its technical attractions and practical advantages. Because of this unfamiliarity a description is given of the circumstances in which the technique can be used, the reasons why alternative techniques were felt to be inadequate, and the mechanics of sample enumeration itself.

Sample enumeration is a technique for forecasting with travel demand models. The range of demand models that can be treated in this way is immense. Applications have been made in studies of demand for modes, for predicting travellers' destinations, forecasting the frequency of travel and, as in this paper, to analyse car ownership and driving licence holding. Many other applications can readily be imagined both within and outside the transport sector. Nor is the type of model restricted to formal statistical models. Many laboratory or panel analysis are effectively sample enumerations in the sense that the behaviour of a comparatively small sample of individuals is predicted and expanded to predict the behaviour of a larger population.

The paper is written in two parts, the first (Sections 2-4) giving an outline of the situations in which classical zonal models are often inappropriate, a general discussion of approaches to aggregation, and the mechanics of the sample enumeration process. The second part describes the specific example of an application in the West Netherlands to car ownership forecasting. The models developed are described briefly, and the sample enumeration evaluations of two policies are given as illustrative examples.

2. DISAGGREGATE MODELLING AND ZONAL FORECASTING

A large literature now exists describing the method and adventages of disaggregate modelling. For example, the conference report (edited by Stopher, Meyburg and Brög) shows a wide divergence of approach to the fundamental problem of investigating travel behaviour; nearly all modern research, however, is based on studies of individual people or households - the disaggregate approach. The strengths of the disaggregate approach have been well documented. They lie primarily in the greater insight obtained from representing all the known detail about the travellers, together with the improved scientific techniques (statistical, econometric or psychological) that have been devised for disaggregate analysis.

Disaggregate models have also been used frequently as a basis for population demand forecasting. Here the objective is to predict totals, whether the total flow on a link in a network, total car ownership for a population group, or total vehicle-kilometres in a region. The improved insight given by disaggregate modelling makes this approach to forecasting population demand rather attractive in principle.

The method that most often been used for deriving population demand models from a desaggregate basis is to replicate the zonal structure of the classical transportation models (generation, distribution, etc.). In such an implementation the models, estimated on the basis of individual data, are applied to data of zonal averages. It has been recognised at least since 1975 that care must be taken if the accuracy of the data on which the model is applied is different from the accuracy of the data on which the model was estimated. A number of procedures have been proposed and implemented for correcting the "aggregation bias" that can arise in such cases.

A more fundamental problem, however, is the structure of the classical models using zones as the basic unit of analysis. The concentration on using large matrices to represent interactions between all possible pairs of zones, and the computational burden that this has imposed, leaves little room for the introduction of other variables. A limited segmentation is often introduced, in the form of car ownership or availability variables, but the full wealth of socio-economic variation cannot be represented in such a model.

The omission of socio-economic variation from zonal models is not only a result of the computational burden that would be imposed. It is also very difficult to obtain data on a number of key variables (e.g. licence holding) without undertaking expensive specific surveys.

It is worth asking whether this concentration on zonal modelling is appropriate , and whether some other approach to population demand forecasting might be better for some applications. This issue turns on the accuracy and sensitivity with which particular variables can be represented and the importance of those variables for the application in question.

Locations are of course represented rather well at a coarse level by zoning. Zone sizes can usually be chosen so that assignment will put trips to and from a zone on to the correct major roads or public transport routes. Journey lengths can also be reasonably well represented, except for travel within a zone. A difficult problem, however, is access to public transport which generally varies rather widely between locations in the same zone.

The socio-economic variables are, as mentioned, the great weakness of zonal modelling. These variables are of particular importance in determining travel frequency (generation) and the availability of cars. Thus policies expected to affect frequency or car availability (such as petrol rationing or park-and-ride promotion) are particularly unsuited to investigation by zonal models.

In summary, zonal modelling imposes a substantial computer burden and gives a poor representation of a number of important variables. The strength of zonal models is in the strong (if coarse) locational representation. A particular application for which zonal modelling is suitable is in preparing assignments.

For very many policy investigations, however, assignments are not necessary. In these cases, it is worth considering whether a different approach to population demand forecasting would be more appropriate.

3. APPROACHES TO AGGREGATION

The general aggregation problem may be stated in the following way:

> Given a means of predicting demand for individuals, determine a corresponding means of predicting total demand for a population.

Usually, the means used for predicting individual demand will be a disaggregate statistical model, but the aggregation problem exists equally for experimental techniques such as in-depth interviewing.

The most frequent approach to problems of this type is to divide the population into more of less homogeneous sub-groups or segments. A population total can then be determined by predicting the demand for a typical member of each segments, multiplying by the size of the segment and summing the segment totals. This is a simple, straightforward and easily explained method.

The three important problems with segmentation are the following:

- the assumption that behaviour within the segments is homogeneous may not be good;
- the need to increase the number of segments to improve the homogeneity assumption imposes a computational burden;
- the need to collect average values of the independent variables for the segments may restrict the variables that can be included in the model.

All of these problems have already been discussed in connection with zonal models. It is interesting to note that a zonal model is simply a form of segmentation as outlined above, the segments being defined primarily on a geographical basis.

The question of whether a zonal model is an appropriate form of segmentation can be formulated in terms of the homogeneity of behaviour. For assignment applications, the assumption of homogeneity (use of the same roads) is clearly better for a zonal segmentation than for segmentations with weaker locational components. For other applications, the weakness of the homogeneity assumption is more of a problem.

It is, however, difficult to find a more appropriate segmentation that does not introduce an unacceptable computational burden. Socio-economic segmentation based on household structure, income, the rôles of household members, licence holding, etc., can obviously improve the homogeneity assumption with respect to these variables. To avoid excessive computation, geographical segmentation must be reduced, thus losing much of the relationship with the transport networks. A segmentation approach clearly has structural limitations.

What is needed is a totally different approach that will avoid the homogeneity assumption and allow the full variety of locational and socio-economic variables to appear in the model. A standard approach to mathematical problems of this type, where many variables are important, is to use a random sampling approach. Such techniques in mathematics are called "Monte Carlo" techniques; in transportation planning it is called sample enumeration or micro-simulation.

The essence of this technique is to draw a random sample of individuals (or households), to calculate the demand for each member of the sample, then to expand to a population total. The sample of individuals may be drawn uniformly or using some stratification, depending on the policies to be investigated. The method of sampling will obviously affect both the expansion factors and the sampling error introduced in the expansion. The demand for each member of the sample may be calculated either as an expectation (when the disaggregate model is a probability model) or as an actualised value of the demand (e.g. from laboratory testing).

The sample enumeration technique has a number of very attractive characteristics for transportation demand forecasting. It is also important, however, to be aware of its limitations.

4. ADVANTAGES AND LIMITATIONS OF SAMPLE ENUMERATION

One of the major advantages of sample enumeration is that it is a direct exploitation of a disaggregate model. Not only does this mean that complicated adjustment procedures are avoided, but more importantly that the best available model can be used. This advantage is particularly important when compared with segmentation approaches which often require very restricted models. Further, the data requirements for an application of the model are exactly those required for estimation: thus if the same sample is used for estimation and for application (as is often done), no further data collection is required.

The advantage of sample enumeration most immediately apparent to the policy analyst is the great reduction in the time and cost required to evaluate a policy, relative to a zonal model. Technically, this advantage arises because the model is exercised once for each individual in the sample, whereas a zonal model has to be exercised once for each combination of origin, destination and segment. Since a reasonable level of accuracy can be obtained from quite small samples (a few hundred), this saving can be very large.

The accuracy and cost of a sample enumeration can be adjusted to suit the particular circumstances of the applications being made, by adjusting the size of the sample. The error introduced by sampling can easily be calculated,and reduced to any desired level by increasing the sample. Note that errors in the parameters of the model (caused by estimating on a sample survey) will remain, as will errors in the specification of the model. Such errors are common to all types of model,although it could be argued that the use of disaggregate models can increase estimation accuracy and reduce specification error.

A further large saving can be made when the policies being studied can be represented in simple terms - for example, a uniform percentage change in a variable. Such cases arise most often when the policies under study concern tax changes or other price changes. Given a simple statement of the policy to be tested, the computer run to evaluate it can be set up and executed in a few minutes.

The limitations of sample enumeration is that it can obviously be applied only in cases when a sample is available that is sufficiently large to estimate the demands of interest with sufficient accuracy. When an assignment must be made to a whole network , the necessary sample size would usually be so great as to make the method impractical, so that zonal models remain the "state of the art" for assignments.

A further difficulty arises when forecasts must be made for future years. Clearly it is not possible to draw a representative sample from the population in the usual way. Methods are available, however, for adjusting samples to take account of expected changes in the composition of the population. These methods are the subject of current research, from which it is hoped to extend the applicability of sample enumeration beyond the short-term policies which are its present domain.

5. CAR OWNERSHIP AND LICENCE MODEL

During 1981 a car ownership and driving licence model was developed on data collected in the Zuidvleugel area of the West Netherlands. This model has been described in a previous paper (in Dutch) by Pol, Geinzer and Daly. The objective of the model was to investigate policies and scenarios relevant to car ownership, and to provide a car ownership forecast for input to the Zuidvleugel model system. Both zonal and sample enumeration implementations were made.

Structuring the licence and car ownership models proved to be an interesting and illuminating problem, because of the unusually high interdependence found in the Netherlands. After an investigation, the structure shown in Figure 1 was eventually adopted, a sequence of logit models representing the choices of :

- driving licence, for each of the first two adults in the household;
- conditional on single licence holding, the choice of ownership of the first car;
- conditional on double licence holding, the choice of ownership of the second car.

This choice of structure reflects the particular characteristics of the data that households with more than two adults are somewhat rare, and that almost all households with two licence own at least one car. Ownership of more cars than licences is of course rare.

The models estimated were of the linear logit form. The coefficients estimated for them are shown in Tables 1-3, together with some summary statistics.

The driving licence model shows an interesting range of socio-economic variables, including income, employment, and household structure. No accessibility variables were found to be significant in influencing licence holding, although the density variable does to some extent represent accessibility. The model gives a good explanation of the overall variation in the data.

The first-car model also shows a very weak influence of accessibility, socio-economic effects being much stronger. The very strong connection between licence and car ownership can be seen in the fact that 89% of these households (even with only one licence) have a car. That the constant gives most of the explanation and that the effective quantity of data is small is seen in the low explanation and the small number of variables found.

The second- car model shows a much stronger accessibility effect, but socio-economic influences also remain strong. The impression is given that the second-car decision is much more "explicable" (in terms of the purchase of accessibility) than the first-car decision. This interpretation is of course intuitively appealing.

Further details of the modelling work are given in the referenced paper and in the study Report.

6. ILLUSTRATIVE APPLICATIONS

The car ownership and licence models are incorporated into a forecasting program which uses a sample enumeration algorithm for a sample data file. This program may be used alone or in conjunction with the Zuidvleugel travel-demand forecasting program (see van Zwam and Daly). When used together these programs are capable of evaluating the impacts of policies and/or scenarios on licence holding, car ownership, and travel demand patterns.

Two applications are presented here as examples of the uses of the program. The first application is an example of scenario analysis, in which it is used to evaluate changes in car ownership and licence holdings resulting from a severe economic downturn. The second application is a example of policy analysis in which the program is used together with the Zuidvleugel sample enumeration program to investigate a shift in the tax burden from car ownership to car use.

6.1. THE ECONOMIC DEPRESSION SCENARIO

This sombre view of the future is represented by specifying that total employment will decline by 10%, and that this decline will be from unemployment in the building and industrial sectors. Within households it is assumed that 40% of multiple worker households will lose 1 worker and 2% of single worker households will lose that worker. It is assumed that income will decrease by 10% and that car operating cost will increase by 15%.

Implementation of this scenario requires that zonal and household level data be altered. Total zonal employment was decreased by 10% in those zones with sufficient building and industrial employment which was building and industrial, and was decreased by the fraction of total employment which was building and industrial otherwise. A 40% random sample of multi-worker households had the number of workers decreased by 1 and a 2% random sample of single-workers households had the number of workers set to 0. Household income was reduced by 10% and car operating costs were increased by 15%.

The results of this scenario, shown in Table 4 show a decrease of 6.3% in the total number of cars and a decrease of 3.3% in the total number of drivers. When examined further it can be seen that just over half of the decline in total cars is from a shift from two-car to one-car households, the remainder being a shift from one-car to no-car households. A similar trend can be seen for drivers.

6.2. THE TAXATION SHIFT POLICY

To illustrate the combined use of the car ownership and Zuidvleugel sample enumeration programs it was decided to test the effect of the elimination of tax on car ownership and it replacement in government revenue by increased taxes on petrol. Calculation of the appropriate petrol tax increase involved some complication because there will be a reduction in the number ot kilometres covered by average car together with a possible increase in the number of

To deal with this complication it was necessary to use both programs iterativel to converge on an appropriate value for the petrol tax increase. Fortunately with a sample enumeration program this approach is feasible, and in two iterations a reasonably close approximation was found. Additional it rations could be made to achieve a better solution, but for this illustrative evaluation this was not considered necessary.

The ownership tax for the cheapest common car (1977 basis) was taken as ƒ235,- per year, representing 8.3% of the total ownership cost (A.N.W.B. data for a Citroën 2 CV). The price of petrol was ƒ1,08 per litre, which was increased to ƒ1,43 for the policy evaluation. The predicted car ownership and travel consequences of this policy are shown in Table 5.

This table shows the main consequences of the policy as a 1.7% increase in the number of cars, but decreases of 2.5% and 4.4% respectively in the number and average length of car driver tours. Negligible change in the number of licences is predicted for this policy. More detailed analysis of the changes in car ownership shows that two thirds of the increase comes from the purchase of second cars.

Calculating the revenue effects requires the additional information that the ownership tax averaged ƒ345,- per year over all cars in 1977. Elimination of this tax implies a revenue loss of ƒ272.8 million for the Zuidvleugel area. The additional tax collected on fuel is collected over a total kilometrage reduced by 6.3% thus implying a revenue gain of ƒ271.5 million.

The figures given here rely on a number of approximations and simplifications, and are to be seen only as broadly indicative of the potential of the methods.

ACKNOWLEDGMENTS

I am grateful to my colleagues on the study described for their permission to use their work in this paper. Responsibility for errors, omissions or interpretations is that of the author.

REFERENCES

T.J. Atherton and M. Ben-Akiva "Methodology for Short Range Travel Demand Predictions: Analysis of Carpooling Incentives", Journal of Transport Economics and Policy, 1977.

P.J. Stopher, A.H. Meyburg, and W. Brög (eds.) "New Horizons in Travel-Behaviour Research" Lexington Books, 1981.

H. Pol, J.C. Geinzer and A.J. Daly "Ontwikkeling en toepassing van een gecombineerd rijbewijs- en autobezitsmodel" (Development and application of a combined driving licence and car ownership model) Colloquium Vervoerplanologisch Speurwerk, The Hague, 1981.

A.J. Daly and H.H.P. van Zwam "Development of Travel Demand Models for the Zuidvleugel Study" P.T.R.C. Summer Annual Meeting, 1981.

A.N.W.B. "Kosten Voorcalculatie in de prijsklasse van ƒ6.750,- tot ƒ7.500,- voor personenauto's met bezine motor" (Cost estimates for petrol-driven pricate cars costing between ..), 1977.

Zuidvleugel Study, Report 9 "Car Ownership Models" to be published by Ministry of Transport and Public Works, The Netherlands.

TABLE 1 : DRIVING LICENCE MODEL

VARIABLE NAME	VARIABLE DEFINITION	ESTIMATED COEFFICIENTS	ACCURACY OF ESTIMATION *
C24	1 for alternatives 2 and 4, (where the first adult has a licence, 0 otherwise).	-.710	4.14
C34	1 for alternatives 3 and 4, (where the second adult has a licence), 0 otherwise.	-2.57	10.3
C4	1 for alternative 4, (where both adults have a licence), 0 otherwise.	1.647	4.61
LIN1	income in thousands of guilders per month, for alternative 1 (where neither adult has a licence), 0 otherwise.	-.532	4.94
LIN4	income in thousands of guilders per month, for alternative 4 (where both adults have a licence), 0 otherwise.	.407	5.74
HED234	1 for alternatives 2 and 4 if the first adult has a high education, 1 for alternatives 3 and 4 if the second adult has a high education, and 2 for alternative 4 if both adults have a high education, 0 otherwise.	.758	5.84
MWK234	1 for alternatives 2 and 4 if the first adult is a worker, 1 for alternatives 3 and 4 if the second adult is a worker, and 2 for alternative 4 if both adults are workers, 0 otherwise.	.727	6.12
1FM24	1 if the first adult is a female ("female head of household") for alternatives 2 and 4, 0 otherwise.	-1.66	7.82
16524	(age-65) of first adult if the first adult is over 65, for alternatives 2 and 4, 0 otherwise.	-.128	5.59

TABLE 1 : Driving Licence Model (continued)

VARIABLE NAME	VARIABLE DEFINITION	ESTIMATED COEFFICIENTS	ACCURACY OF ESTIMATION
24534	(age-45) of second adult if the second adult is between 45 and 55, 10 if the second adult is over 55, for alternatives 3 and 4, 0 otherwise.	-.165	7.89
NCH24	number of children (0-17) for alternatives 2 and 4, 0 otherwise.	.417	4.68
PDN24	population density (per hectare) of home-based zone for alternatives 2 and 4, 0 otherwise.	-.0331	2.75

SUMMARY STATISTICS:

ALTERNATIVE	NUMBER AVAILABLE	NUMBER CHOSEN
1. neither adult licenced	1612	612
2. 1st adult licenced	1612	563
3. 2nd adult licenced	1268	32
4. both adults licenced	1268	405
Total Observations		1612

$L^*(0) = -1996.26^*$ $\rho^2 = 0.371^*$

$L^*(C) = -1764.94$ $\rho^2_c = 0.288$

$L^*(\beta) = -1256.60$

The L function gives the log likelihood respectively with zero coefficient values, with constants only, and with the estimated coefficients. The p^2 function is Tardiff's statistic.

TABLE 2 0/1 Car Ownership Model for Households with 1 Licenced Driver

VARIABLE NAME	VARIABLE DEFINITION	ESTIMATED COEFFICIENTS	ACCURACY OF ESTIMATION
C2	1 for alternative 2 (household owns a car) 0 otherwise.	2.49	3.87
PW12	the sum of the personal business destination logsums for all working adults for both alternatives	.154	.694
PK12	the sum of the personal business destination logsums for all household members of age 4-17 for both alternatives.	.0996	.291
RIN12	the log of monthly income remaining after rent, daily needs expenditures, and fixed car costs, for both alternatives	8.8	1.97
LFM2	1 if the licenced driver is female, for alternative 2, 0 otherwise.	-1.36	3.65
L452	(age-45) if the licenced driver is between 45 and 55, 10 if the licenced driver is over 55 for alternative 2, 0 otherwise.	.136	3.17

SUMMARY STATISTICS

ALTERNATIVE	NUMBER AVAILABLE	NUMBER CHOSEN
1. no cars owned	486	54
2. one car owned	486	432
Total Observations		486

$L*(0) = -336.87$* $\rho^2 = 0.547$*

$L*(C) = -169.53$ $\rho_c^2 = 0.100$

$L*(\beta) = -152.53$

* See Table 1 for definitions.

TABLE 3 1/2 Car Ownership Model for Households with two Licenced Drivers

VARIABLE NAME	VARIABLE DEFINITION	ESTIMATED COEFFICIENT	ACCURACY OF ESTIMATION
C2	1 for alternative 2 (household owns 2 cars), 0 otherwise.	-.485	2.53
PNW12	The sum of the personal business destination logsums for all non-working adults, for alternatives 1 and 2.	1.47	2.03
RIN12	The log of monthly income remaining after rent, daily needs expenditures, and fixed car costs, for alternatives 1 and 2.	16.0	4.06
MWK2	1 if the household has more than 1 worker, 0 otherwise, for alternative 2.	1.82	2.68
NKD2	the number of household members of ages 0-11, for alternative 2.	.545	2.57
NTN2	the number of household members of ages 12-17 for alternative 2.	.791	3.76

SUMMARY STATISTICS

ALTERNATIVE	NUMBER AVAILABLE	NUMBER CHOSEN
1. one car owned	348	296
2. two cars owned	348	52
Total Observations:		348

$L^*(0) = -241.21^*$ $\rho^2 = 0.471^*$

$L^*(C) = -146.76$ $\rho_c^2 = 0.130$

$L^*(\beta) = -127.65$

* see Table 1 for definitions.

Table 4 IMPACT OF ECONOMIC DEPRESSION SCENARIO

Total Cars and Licences in Study Area

	Cars	Licences
Base	790600	1104800
Scenario	740600	1068700
Change	-50000	-36100
% Change	-6,3	-3,3

Households owning given number of Cars

	None	One	Two
Base	490000	574300	108200
Scenario	513400	577600	81500
Change	23400	3300	-26700
% Change	+4,8	+0,6	-24,7

Households holding given number of licences

	None	One	Two
Base	403300	435700	334500
Scenario	418600	439100	313800
Change	16400	3400	-19700
% Change	+4,1	+0,8	-5,9

Table 5 TRAVEL CONSEQUENCES OF TAXATION SHIFT POLICY

Car Driver Tours	Number	Average tour length
Base	930100	15,05
Policy	906900	14,39
Change	-23200	-0,66
% Change	-2,5	-4,4

Total Cars

Base	790600
Policy	804100
Change	13500
% Change	+1,7

Total Licences

Base	1104800
Policy	1104800
Change	0
% Change	0