A Comparison of Alternative Approaches to Modelling Car Ownership

Gerard Whelan[1]
MVA Consultancy, London.

1 Introduction

Models to predict changes in the level of car ownership have been under development since the 1930s. They are essential to the transport planning process and are of interest to government, vehicle manufacturers, environmental protection groups, public transport authorities and public transport operators. It is a complex area of research and one in which many theories and methodologies have been advanced.

Within this long history, the earlier methodologies tended to be aggregate in nature, dealing with ownership at a national, regional or local level. They were calibrated to time-series, cross-sectional or pooled data and forecast future levels of car ownership by extrapolating from a trend equation. More recently, the majority of research has centred on the development of disaggregate models of car ownership, in which the household or individual's propensity to own vehicles is related to their socio-economic and locational characteristics, the cost of ownership and the availability of other means of transport.

This paper compares the pros and cons of different data and modelling methodologies in the production of car ownership forecasts.

A range of alternative model forms are estimated to aggregate time-series data. The range of models comprise of two distinct groups. The first group, referred to as aggregate ownership models, includes a series of multivariate regressions showing the total number of vehicles as a function of GDP, motoring costs and household size. The second group of models, referred to as aggregate share models, shows the proportion of households with zero, one, two and three or more vehicles as a function of the same set of explanatory variables.

Next a range of disaggregate choice models are developed to explain car ownership at the household level. The models are calibrated to data from the National Travel Survey over the years 1985 to 2006 and include the ownership decisions of around 50,000 households. The range of models developed includes: multinomial logit, dogit and mixed logit models. All of the models show ownership of zero, one, two and three or more cars as a function of household income, household employment, household location, company car ownership, ownership costs, and running costs. Due to problems in estimating robust coefficients to motoring costs, each of the choice models is constrained to generate cost elasticities equal to those identified from models based on aggregate data.

This work adds to the understanding of the factors that influence household car ownership by drawing on the strengths of a range of complementary data sets. It provides evidence on market saturation and sensitivities to purchase and running cost and develops new ways of applying the models to generate forecasts. The credibility of alternative model forecasts are compared.

[1] This work was undertaken as part of PhD research at the Institute for Transport Studies, University of Leeds.

2 Aggregate Car Ownership Models

This section of the paper describes the estimation of a set of car ownership models based on aggregate time-series data for Great Britain from 1951 to 2000. The work builds on previous aggregate car ownership analysis such as that reported by Wolff (1938), Smeed (1951), Tanner (1958), Kain and Beesley (1965), Tanner (1978) and Button *et al* (1982).

The advantages of developing models on aggregate time-series data include:

- an examination of the impact of changes in income, purchase and running costs on car ownership, including an analysis of gains, losses and lagged effects which is more difficult in discrete choice models;
- the estimation of elasticities of demand to GDP, purchase and running costs. It is here where the compleme ntary nature of the aggregate and disaggregate methodologies is greatest as it has not been possible to estimate cost elasticties using disaggregate data;
- the identification of a time trend;
- the availability of sophisticated estimation procedures involving non-linear least squares and maximum likelihood techniques allow for the simultaneous estimation of all model parameters when previously model calibration was undertaken as a multi-stage process; and
- the estimation of models which are relatively simple to interpret and extrapolate to generate forecasts.

2.1 Aggregate Data

The data used to calibrate the aggregate ownership models spans the years from 1951 to 2000 and includes:

- The total number of private and light goods vehicles currently licensed - taken from Transport Statistics Great Britain (TSGB) 2002;
- The proportion of households with regular use of zero, one, two, and three or more cars - taken from TSGB 2010;
- The total number of households - taken from TSGB 2002;
- A real purchase cost index constructed from two separate series. Data for 1951-1974 was taken from Tanner (1977) and data for 1974-2002 was supplied by the Department for Transport.
- A real running costs index defined to include maintenance, fuel, tax and insurance. As with the purchase cost index, the running cost index was constructed from two separate series: 1951-1974 was taken from Tanner (1977) and 1974-2000 was supplied by the Department for Transport.
- A GDP index at constant factor prices - taken from the Blue Book (2002). This index includes data for Northern Ireland and it has been assumed that the equivalent index for Great Britain is not significantly different.

Table 1: Aggregate Data Summary

Year	Households (Millions)	Population (Millions)	Cars (Millions)	GDP Index (1998=100)	Purchase Price Index (1998=100)	Running Cost Index (1998=100)
1951	14.5	50.3	2.6	31.4	223.5	76.5
1956	15.3	51.2	4.1	35.5	223.0	81.2
1961	16.3	52.8	6.2	40.9	199.8	79.0
1966	17.4	54.6	9.5	47.9	143.6	77.8
1971	18.6	55.9	11.9	54.3	140.9	77.3
1976	19.3	56.2	14.8	60.7	135.4	83.4
1981	20.2	56.4	16.5	63.5	130.6	89.7
1986	21.3	56.9	18.9	74.1	115.9	83.6
1991	22.4	57.8	22.0	82.7	107.4	84.2
1996	23.5	58.8	23.7	93.9	105.3	94.1
2001	24.6	59.8	27.0	108.0	86.3	115.0

Table 1 shows a steady increase in the population, number of households and number of cars since 1951. Differences in the rates of growth between population and number of household indicate a strong trend towards smaller household size. With regard to purchase price, there was a steady decline from the early 1950s to the early 1970s. This decline possibly reflects changes arising from increased market penetration and associated economies of scale in production. Following a brief increase in the price of vehicles in 1978 and 1979 the downward trend in real purchase price continued, picking up speed towards the end of the 1990s, perhaps due to increased competition brought about by the direct importation of cars which are comparatively less expensive in markets in other European Union states. Real running costs have remained remarkably stable over the period although there are signs of rapid increases following the oil shocks of 1973 and 1978 and the introduction of the 'fuel duty escalator' in 1993. Gross Domestic Product shows a general increase over time with periods of recession in the early 1980s and early 1990s.

2.2 Aggregate Ownership Models

The aggregate ownership models explain the number of cars per household as a function of time, GDP per household, real purchase costs, real running costs and household size (average number of people per household). Two categories of model are developed:

- those which are based on product life cycle theories and follow an S-shaped growth curve until saturation is reached; and
- those which specify a diminishing rate of growth to saturation but which are not S-shaped.

Given the importance of saturation in the specification of aggregate ownership models it is useful to provide a discussion of its meaning and role in the forecasting process. In attempting to define what saturation is, Button *et al* (1982) note three possibilities:

- Saturation is simply a statistical parameter defining the upper asymptote during the period under review. This saturation level in this context is simply a parameter to be estimated and is of no consequence in its own right.
- Saturation may be regarded as a ceiling for average car ownership level amongst a group of households independent of income and costs. This definition for saturation has been described by Kirby (1976) as 'money no object saturation'.
- Saturation may be defined as the average long term level of car ownership allowing for the possibility that subgroups in the population may have different 'ceilings' and the point at which all groups reach their ceilings is the long run average.

The third definition of saturation presented above provides a consistent interpretation of saturation for both aggregate and disaggregate models and it is this definition that is adopted here. By specifying car ownership models with an S-shaped functional form and a saturation level, forecasts of vehicle ownership will be curtailed as saturation is approached. This feature is likely to be of little significance to forecasts for emerging markets but is highly significant in more mature markets such as Great Britain. Conversely, product take-up is not of great relevance to mature markets and therefore an S-shaped function may be inappropriate and a better fit to the data could be achieved using non-sigmoid functions.

S-shaped Functions

Tanner (1958) was among the first in the UK to develop trend extrapolation techniques to be used for long term forecasting. Although other research methodologies were explored, there was an initial preference for logistical time-trend extrapolation. This was because it avoids the need to forecast the future levels of explanatory variables contained in causal models and Tanner believed that the rate of growth in ownership in the forecast time period was closely related to the rate of growth in preceding periods. He assumed that a saturation point exists where car ownership rates stabilise and believed the logistic curve to be compatible with this theory of ownership. In this analysis, the logistic model relates the number of cars per household in a particular time period (C_t) to a time trend t:

$$C_t = \frac{S}{1 + b\exp(-aSt)} \tag{1}$$

Where S is the saturation rate, b is a constant that performs a similar role to the intercept term in a linear model, and a is a constant scaling the proportionate change in the car ownership rate per year (which decreases as saturation is approached). By allowing a to depend upon income, motoring costs and household size, a more general model is obtained. This model has the same form as that proposed by Tanner (1975) and is termed the 'modified logistic' model:

$$C_t = \frac{S}{1 + bY_t^{-fS}O_t^{-gS}R_t^{-hS}I_t^{-mS}\exp(-aSt)} \tag{2}$$

where *Y* is a measure of GDP per household, *O* an index of real car purchase costs, *R* an index of real car running costs and *I* the average number of individuals per household[2].

The logistic function is however symmetrical around its mid point, whereas empirical evidence suggests that ownership reaches half of the saturation level in a lesser time than it takes to complete the remaining growth to saturation (Tanner, 1977; Bates *et al*, 1978). To accommodate this characteristic Tanner (1977) specified a 'power growth' model similar to equation 3 below:

$$C_t = \frac{S}{1 + (b + at + f\ln(Y_t) + g\ln(O_t) + h\ln(R_t) + m\ln(I_t))^{-n}} \tag{3}$$

Althernatively, the Gompertz function, used by Button *et al (*1982), Mogridge (1983) and Dargay & Gately (1999), generates a 'non-symetrical', sigmoid function as:

$$C_t = S\exp(b\exp(at + fY_t + gO_t + hR_t + mI_t)) \tag{4}$$

where *b* governs the slope of the lower portion of the s-curve and the saturation level *S* is approached more slowly than for the causal logistic model.

Finally, Tanner (1983) modified the power growth model to allow for inertia effects by the inclusion of lagged car ownership, income and motoring cost variables. This model is based on a 'stock adjustment' hypothesis, whereby demand for new cars in a given year represents an adjustment to the desired stock of cars:

$$C_t = \lambda \frac{S}{1 + (b + f\ln(Y_t) + g\ln(O_t) + h\ln(R_t) + m\ln(I_t))^{-n}} + (1-\lambda)C_{t-1} \tag{5}$$

where λ is the adjustment parameter.

Non-Sigmoidal Functions

An interesting alternative to the S-shaped functional form is to specify a model that shows a decreasing rate of growth to saturation but does not involve calibration of the lower portion of the s-curve. Although moving away from product life cycle theories, this form of model is justified on the grounds that the lower portion of the S-shaped curve is not relevant for relatively mature car markets such as that in Great Britain. One such form that is characterised by these properties is the 'constrained exponential' given in equation 6.

$$C_t = S - b\exp(at + fY_t + gO_t + hR_t + mI_t) \tag{6}$$

2.3 Aggregate Share Models

The models outlined so far all explain the average number of cars per household as a function of time, income, the costs of ownership and household size. An alternative set of data is also available which shows the proportion of households owning zero, one, two and three or more vehicles. This data has been used to develop models which explain the proportion of households in each market segment (0, 1,

[2] It is important to note that Tanner did not distinguish between ownership and running costs, and because his models were specified in terms of cars per capita they did not include household size.

2, 3+) as a function of time, income, motoring costs and household size. These models are calibrated to annual time series data from 1951 to 2000 and are referred to as aggregate share models. The most straightforward share model is the multinomial logit model:

$$P_i = \exp(V_i) \Big/ \sum_j \exp(V_j) \tag{7}$$

where

P_i is the aggregate share of the *i*th segment

V_i is a function of time, GDP per household, motoring costs and household size

$V_0 = 0$ and $V_i = b_i + a_i T + f_i Y + g_i O + h_i R + m_i I$

It is likely that there will always be a proportion of the population who for one reason or another will never own a car. This feature can be taken into account by estimating an aggregate version of the Dogit model (Gaudry and Dagenais, 1979) and specifying a minimum share for households without access to vehicles. In this model, the share of market segment *i* is given by:

$$P_i = \left(\exp(V_i) + \theta_i \sum_j \exp(V_j)\right) \Big/ \left(\left(1 + \sum_j \theta_j\right) \sum_j \exp(V_j)\right) \tag{8}$$

The θ parameters define the proportion of the population that is captive to that segment.In this application of the dogit model, a minimum share fraction is specified for zero car households only with other market segments constrained to have a zero minimum.

2.4 Aggregate Model Estimation

Table 2 shows the estimated coefficients for the aggregate ownership models in which the dependent variable is cars per household and the independent variables include: a time trend, GDP per household, a real purchase cost index, a real running cost index and average number of people per household. All models are estimated using non-linear least squares regression in LIMDEP (Econometic Software, 2003).

The models show a good fit to the data. When used to generate 'forecasts' for 1951-2000, the lagged power growth model shows the best fit to the data, followed by the modified logistic, constrained exponential, power growth, logistic and Gompertz. The differences between the models are, however, marginal.

Saturation levels are estimated to a high degree of precision and range from 1.07 (±2.4%) cars per household in the logistic model to 1.57 (±13.0%) in the constrained exponential model. Considering that the actual number of cars per household in 2001 was 1.08 and that there has been a steady rate of increase in the number of cars per household over recent years, it would appear that the saturation levels estimated in the logistic and modified logistic models are too low and the models are therefore likely to be inappropriate for forecasting.

Table 2: Aggregate Ownership Models

	Logistic	Modified Logistic	Power-Growth	Gompertz	Lagged Power-	Constrained Exponential

					Growth	
Saturation	1.07 (83.8)	1.14 (89.3)	1.25 (25.9)	1.19 (63.5)	1.54 (6.6)	1.57 (15.2)
b (constant)	5.4629 (25.1)	55369.7 (0.4)	-0.5811 (1.0)	-69.2 (0.5)	-0.6972 (0.4)	1.6572 (8.8)
a (trend)	0.0882 (23.8)	0.1219 (6.7)	0.0194 (4.1)	-0.0638 (4.1)	n.a.	-0.0164 (8.2)
f (GDPHH)	n.a.	0.5348 (1.3)	0.1679 (1.7)	-0.01073 (2.9)	0.9312 (4.1)	-0.0046 (2.1)
g (Purchase)	n.a.	-0.3770 (3.7)	-0.0588 (2.5)	0.0011 (2.6)	-0.0962 (1.5)	0.0003 (1.5)
h (Running)	n.a.	-0.5571 (2.9)	-0.0756 (2.1)	0.0020 (1.3)	-0.0751 (0.7)	0.0005 (0.6)
m (HH size)	n.a.	8.3992 (4.4)	0.9741 (2.6)	-0.9587 (1.7)	-1.4160 (3.6)	n.a.
n (power)	n.a.	n.a.	4.6723 (3.0)	n.a.	1.9247 (2.6)	n.a.
λ (lag)	n.a.	n.a.	n.a.	n.a.	0.3861 (5.2)	n.a.
R-squared	0.99311	0.99831	0.99985	0.99854	0.99908	0.99653
Observations	50	50	50	50	49	50

Note: t-statistics with respect to zero are shown in brackets

With the exception of the lagged power growth model, all models show a positive relationship between ownership rates per household and time, GDP per household and average household size, and a negative relationship between ownership rates and purchase and running costs. The household size coefficient was the wrong sign and insignificant in the constrained exponential model and because it is correlated with other variables had a significant detrimental effect to the income and cost coefficients. This variable was subsequently dropped. The same coefficient was wrong signed and significant in the lagged power growth model but in this instance, its removal resulted in a very poor model.

The elasticities to GDP, purchase and running costs implied by the model decrease in absolute size as the market approaches saturation. There is significant variation between models which can be attributed to a combination of inappropriate saturation levels and relatively imprecise coefficient estimates due to the limited size of the data set and correlations between coefficients.

The variation in purchase and running cost elasticities is of concern since they are required to constrain the cost coefficients in the disaggregate models. The modified logistic and power growth models show statistically significant coefficients on purchase and running costs, and generate similar elasticity values for the year 2000. Because the saturation level estimated in these models provides some cause for concern, an alternative set of models was calibrated with saturation constrained to equal 1.1, 1.3, 1.5, 1.7 and 1.9 cars per household in order to test for the influence of saturation on cost elasticities. In the power growth model, changing the saturation level has relatively little impact on the implied elasticities, with estimated values the GDP elasticity ranging from 0.13 to 0.14, the ownership cost elasticity ranging from −0.036 to −0.043 and the running cost elasticity ranging from −0.019 to -0.045. The variation in elasticities within the modified logistic model is, however, much greater and calls into question their validity for transfer to the disaggregate models. All things considered, the elasticities derived from the power growth model will be used in the development of disaggregate choice models.

Table 3: Aggregate Share Models

	Aggregate MNL			Aggregate Dogit		
	1 car	2 cars	3+ cars	1 car	2 cars	3+ cars
b (constant)	-2.642 (8.5)	-14.50 (27.5)	17.09 (11.0)	-17.064 (13.3)	-29.4035 (20.9)	-32.0293 (5.3)
f (GDPHH)	1.088 (15.7)	3.612 (32.4)	4.769 (15.9)	4.1352 (22.6)	6.7606 (30.9)	7.9579 (21.3)
g (Purchase)	-0.0111 (38.3)	-0.0147 (25.4)	-0.0354 (16.1)	-0.0067 (17.1)	-0.0102 (21.4)	-0.0314 (18.2)
h (Running)	-0.0075 (10.7)	-0.0078 (8.0)	-0.0297 (13.5)	-0.0123 (6.4)	0.00122 (4.7)	-0.0104 (4.5)
θ		n.a.			0.3434 (56.6)	
Rho Squared (0)		0.3206			0.3215	
Rho Squared (C)		0.0779			0.0791	
Observations		50			50	

Note: t-statistics with respect to zero are shown in brackets. The models are estimated using maximum likelihood and are weighted to take account of the fact that each observation is derived from Family Expenditure Survey data.

Table 3 shows the estimated coefficients for the aggregate share models. These models show a high correlation between GDP per household, household size and the time trend to the extent that it was not possible to estimate a reasonable coefficient for each. Because it is useful to compare the GDP elasticities implied by the share models with those derive from other models, GDP per household was retained and the time trend and household size omitted. Further, to avoid forecasts of multiple vehicle households increasing too rapidly over time, GDP per household was specified in logarithmic form.

The same specification and transformation of the explanatory variables was adopted for the estimation of the aggregate dogit model. This model improves the overall level of fit to the data and implies that 25.6% of households will never have access to at least one vehicle. Given that the share of households with zero cars was 27% in 2000 this figure is arguably too high and casts doubt on the usefulness of the dogit model for estimating aggregate share models.

The models developed in this section imply generate elasticities of demand with respect to indicies of GDP per household, purchase costs and running costs – see Table 4.

Table 4: Aggregate Model Elasticities

Variable	Year	Modified	Power-	Gompertz	Const.	Lagged	MNL	Dogit

		Logistic	Growth		Exp.	Power Growth		
	1960	0.417	0.679	0.757	0.900	1.215	1.161	2.924
	1970	0.275	0.377	0.472	0.484	0.688	1.128	1.888
GDP	1980	0.172	0.247	0.320	0.337	0.492	1.176	1.435
	1990	0.078	0.143	0.204	0.257	0.294	1.190	1.095
	2000	0.040	0.084	0.130	0.221	0.209	1.139	0.780
	1960	-0.294	-0.238	-0.278	-0.203	-0.126	-1.793	-0.973
	1970	-0.194	-0.132	-0.101	-0.064	-0.071	-0.986	-0.493
Purchase Costs	1980	-0.121	-0.086	-0.060	-0.039	-0.051	-0.946	-0.391
	1990	-0.055	-0.050	-0.026	-0.020	-0.030	-0.706	-0.366
	2000	-0.028	-0.029	-0.011	-0.012	-0.022	-0.514	-0.248
	1960	-0.435	-0.306	-0.191	-0.141	-0.098	-0.443	0.594
	1970	-0.287	-0.170	-0.101	-0.064	-0.055	-0.359	0.265
Running Costs	1980	-0.179	-0.111	-0.068	-0.044	-0.040	-0.384	0.158
	1990	-0.081	-0.064	-0.035	-0.027	-0.024	-0.376	-0.077
	2000	-0.042	-0.038	-0.027	-0.028	-0.017	-0.477	-0.194

Of the ownership models, the GDP elasticities are generally highest for the lagged power growth model (these are long run values) and are lowest for the modified logistic. The constrained exponential, power growth, Gompertz and short run lagged power growth models show similar patterns. For the share models, elasticities are derived by converting ownership shares to cars per household, assuming 3.2 cars per household in the three plus category (this is the average ownership rate in the National Travel Survey 1985-1998). The GDP elasticities shown by these models are considerably higher and less plausible that those shown by the ownership models.

With the exception of the aggregate share models, the modified logistic and power growth models generally show the largest absolute values for purchase cost elasticities. The constrained exponential and Gompertz functions show a rapid decline from 1960 to 1970 the a steady decline similar to the long run lagged power growth thereafter. As with the GDP elasticities, the purchase cost elasticities fall with the purchase cost and as saturation begins to bind. The aggregate share models show substantially higher elasticities, especially the unconstrained multinomial logit.

Following patterns in the ownership elasticities, the modified logistic and power growth show the greatest sensitivity to running cost changes with the constrained exponential and lagged power growth the least sensitive. The share models again show less plausible results, especially the dogit model which has an intuitively incorrect sign on the running cost coefficient for two car households.

2.5 Aggregate Model Summary and Discussion

Using published aggregate time series data it has been possible to estimate a series of models that explain ownership as a function of time, GDP, motoring costs and average household size. This series of models comprise two distinct groups. The first group, known as aggregate ownership models include a series of functional forms previously employed by the UK government as well as the Gompertz function and the newly developed constrained exponential model. This group of models was calibrated using non-linear least squares procedures. The second group of models explain the proportion of households with zero, one, two and three or more vehicles as a function of GDP, motoring costs and household size. This group of models, referred to as aggregate share models, were estimated using maximum likelihood procedures and include the commonly applied multinomial logit and the first application of the dogit model to aggregate car ownership data.

The aggregate ownership models include the direct estimation of saturation in terms of cars per household. These estimates range from 1.07 in the simple logistic model to 1.57 in the constrained exponential. A direct comparison with other research is difficult since saturation is often expressed in terms of cars per capita rather than cars per household. Nevertheless, assuming that the average household size will stabilise at 2.2 people per household, crude estimates of cars per capita from the logistic and constrained exponential models are 0.49 and 0.71 respectively. In comparison, Tanner's initial estimate of saturation was 0.4 (Tanner, 1958) increasing to 0.45 using the modified logistic specification, 0.5 in the power growth model and 0.6 in the lagged model. Higher estimates of saturation, equal to 0.65, were assumed by Mogridge (1967) and DoT (1989). This figure equates to 90% of the driving age population owning cars. Analysis of data from the US shows that in 2000 there were 0.68 cars per capita and 1.78 cars per household. Given that car ownership in the US increased by 14.5% between 1990 and 2000 it is likely that this market has still to reach saturation. When applying the models to generate long-term forecasts, it is prudent to bear in mind the soundness of the assumed saturation when assessing the credibility of the forecasts.

A comparison of the reported elasticities to GDP and motoring costs with other evidence is difficult since they are strongly related to the functional form of the model and, where saturation is included, the year in which the elasticities are applicable. For 1990, the elasticities implied by the ownership models reported here range from 0.08 to 0.29 for GDP, -0.02 to -0.06 for purchase costs and -0.02 to -0.08 for running costs. Evidence from elsewhere suggests a wide range of possible values. Dargay and Vythoulkas (1999) report short run elasticities of demand to income, purchase and running costs of 0.27, -0.44 and -0.28 respectively. Hanly and Dargay (2000) report income elasticities of between 0.15 and 0.2. Dargay (2000) reports a range of income and ownership elasticities at different levels of cars per household. Assuming one car per household (approximately equal to the average for 1990), Dargay reports a short run (rising) income elasticity of 0.42 and a short run car purchase cost elasticity of -0.07. Finally Romily *et al* (1998) report short run elasticities to GDP of 0.34 and to ownership costs of -0.29.

On the basis of goodness of fit to the data, precision and credibility of parameter estimates and implied elasticities, the power growth is the preferred function for the ownership models and the MNL is the preferred function for the share models. There are however inherent problems with each approach. Firstly, is difficult to use aggregate data to estimate a saturation point when actual ownership rates are some way from saturation and secondly, aggregate share models are unable to take account of the limited range of households who would, money no object, own multiple vehicles.

3 Disaggregate Models

During the 1970s and early 1980s research into discrete choice models made considerable progress, as did the modelling and forecasting of car ownership. As modelling techniques improved, researchers pitched their analysis at increasingly finer levels of detail incorporating many different aspects of car ownership. The first batch of models, developed before 1980, focused on the choice of how many vehicles to own (Burns and Golob, 1975; Lerman and Ben-Akiva, 1976; Kain and Fauth, 1977; Mogridge, 1978; Train, 1980). Later, around 1980, new models were developed to examine the choice of car type (Lave and Train, 1979; Manski and Sherman, 1980; Beggs and Cardell, 1980; Lave and Bradley, 1980). However, it was not until the mid-1980s when analysts explicitly began to recognise the importance of the interrelationships between the choice of the number of cars to own and their type (Lave and Train, 1979; Manski and Sherman, 1980; Hensher and Le Plastrier, 1983; Mannering and Winston, 1985; Train 1986).

The advantages of developing disaggregate models of household car ownership include:

- A readily available data set provided by the National Travel Survey from 1985 to 1998 containing the characteristics and ownership details of 45,692 households;
- The calibration of ownership models at the household level will enable forecasting at a local level;
- The choice models used have a strong behavioural foundation seated in the theory of utility maximisation; and
- The models developed can account for taste variation across the population, correlations between choice alternatives and captivity to one particular alternative.

Disaggregate models of household car ownership are, however, more difficult to develop and apply to forecasting than their aggregate counter parts.

3.1 National Travel Survey Data

The National Travel Survey is a household survey of travel covering residents in Great Britain. The survey is conducted by the Office for National Statistics on behalf of the Department for Transport. It was first conducted in 1965 and then at periodic intervals until 1988 from which time it has been continuous. The survey is based on a stratified random sample of households, taken from the Postcode Address File, and is conducted throughout the year. Each household member is asked to keep a seven-day travel diary, detailing all travel taken in Great Britain. The NTS includes questions for many key variables including: age, sex, occupations, socio-economic status, driving licence holding, car ownership or availability to house.

Table 5: NTS Data Summary

NTS Year	Number of Households	Cars per Household	Proportion of Households with			
			No Car	One Car	Two Cars	Three Plus Cars

1985	5061	0.840	38%	43%	16%	3%
1986	5205	0.818	38%	45%	14%	3%
1988	1754	0.857	35%	46%	16%	3%
1989	3675	0.911	35%	43%	18%	3%
1990	3535	0.934	33%	45%	19%	3%
1991	3542	0.972	31%	45%	20%	4%
1992	3453	0.962	32%	44%	20%	4%
1993	3418	0.944	33%	45%	19%	4%
1994	3425	0.944	33%	44%	20%	3%
1995	3339	0.975	31%	44%	20%	4%
1996	3210	0.988	31%	44%	21%	4%
1997	3139	0.993	30%	45%	22%	3%
1998	2936	1.029	29%	45%	21%	5%

3.2 Disaggregate Model Estimation

The object is to develop discrete choice models that explain the household's ownership decision (0, 1, 2 or 3+ vehicles) as a function of: household income, household size and structure, household employment, motoring costs, and accessibility (proxied by area type and population density).

The first model to be estimated is the multinomial logit model (MNL). This model is by far and away the most popular form of choice model and acts as the benchmark on which to judge other models. There are, however, well-documented problems associated with the error structure of the MNL model that result in undesirable common cross-elasticities between options at the household level. This problem arises due to the Independence of Irrelevant Alternatives (IIA) property of the model. To reduce the impact of the IIA property, more flexible model forms are estimated including the dogit and mixed logit models.

It is quite conceivable that there will be some households who for one reason or another will never own a car. These households are captive to one choice alternative (zero cars) and failure to take account of this during model calibration can lead to biased coefficients (Swait and Ben-Akiva, 1986). In car ownership modelling, this captivity is often couched in terms of market saturation and the dogit model (Gaudry and Dagenais, 1979) can be specified to include captivity.

Finally, a mixed logit model is specified to allow for more complex patterns of correlations between choice alternatives and to allow for random taste variation across decision-makers. Specifically, the mixed logit model is calibrated to allow for a cross nested structure in which contiguous alternatives are correlated.

Multinomial Logit Model Calibration

Table 6 shows the estimated coefficients for the multinomial logit model with the following utility expressions:

$$V_0 = 0 \tag{9}$$

$$V_1 = \alpha_1 + \left(\beta_1 + \sum_{a=1}^{A-1}\beta_{a1}D_a + \sum_{h=1}^{H-1}\beta_{h1}D_h\right)Y^{\lambda_1} + \tau_1 T + \delta_1 PD \\ + \omega_{11}E1 + \omega_{12}E2 + \omega_{13}E3 + \mu_1 O + \varphi_1 R \tag{10}$$

$$V_2 = \alpha_2 + \left(\beta_2 + \sum_{a=1}^{A-1}\beta_{a2}D_a + \sum_{h=1}^{H-1}\beta_{h2}D_h\right)Y^{\lambda_2} + \tau_2 T + \delta_2 PD \\ + \omega_{21}E1 + \omega_{22}E2 + \omega_{23}E3 + \mu_2 O + \varphi_2 R + \gamma_2 C1 \tag{11}$$

$$V_3 = \alpha_3 + \left(\beta_3 + \sum_{a=1}^{A-1}\beta_{a3}D_a + \sum_{h=1}^{H-1}\beta_{h3}D_h\right)Y^{\lambda_3} + \tau_3 T + \delta_3 PD \\ + \omega_{31}E1 + \omega_{32}E2 + \omega_{33}E3 + \mu_3 O + \varphi_3 R + \gamma_3 C1 + \psi_3 C2 \tag{12}$$

Where:

Y	is gross household income in 1998 prices
T	is time trend
PD	is population density
E1	is equal to one if the household has one adult in employment, else zero
E2	is equal to one if the household has two adults in employment, else zero
E3	is equal to one if the household has more than two adults in employment, else zero
O	is a purchase cost index
R	is a running cost index
C1	is equal to one if the household owns one company car, else zero
C2	is equal to one if the household owns two company cars, else zero
D_a	is a vector of dummy variables for each area type
D_h	is a vector of dummy variables for each household type
$\alpha, \lambda, \tau, \beta, \omega$ $\delta, \varphi, \mu, \gamma, \psi$	are parameter vectors to be estimated

Equations 9-12 show the utility of ownership as a function of household income, household structure, area type and population density, purchase costs, running costs, company car acquisition and a time trend.

Table 6: Multinomial Logit Model of Car Ownership

	1 Vehicle	2 Vehicles	3+ Vehicles
ASC	-11.2826 (-4.9)	-8.2556 (-24.4)	-9.9922 (-18.9)
Income	7.4058 (3.3)	0.5647 (4.9)	0.0781 (0.8)
Income (Outer London)	0.4801 (7.7)	0.2501 (7.6)	0.1677 (4.3)
Income (Metropolitan Areas)	0.2948 (5.6)	0.2030 (6.9)	0.0893 (2.7)
Income (Urban)	0.4977 (9.0)	0.2898 (9.1)	0.1661 (4.6)
Income (Rural)	0.7301 (10.6)	0.4352 (10.6)	0.2801 (5.6)
Income (One Adult, Retired)	-0.6389 (-11.4)	-0.7836 (-6.9)	0
Income (One Adult, with Children)	-0.2347 (-4.4)	-0.4064 (-3.9)	0
Income (Two Adults, Retired)	0.3900 (8.9)	0.6054 (11.8)	0.4495 (5.6)
Income (Two Adults, no Children)	0.2541 (6.3)	0.5928 (11.6)	0.4342 (5.7)
Income (Two Adults, with Children)	0.3485 (8.1)	0.6627 (12.1)	0.4489 (5.8)
Income (Three plus Adults, no Children)	-0.1298 (-2.6)	0.6142 (11.1)	0.6679 (6.5)
Income (Three plus Adults, with Children)	-0.0412 (-0.7)	0.5795 (10.7)	0.5624 (6.1)
Income Power Term	0.1167 (4.4)	0.4292 (19.0)	0.5589 (14.3)
Time trend	0.0392 (12.3)	0.0802 (17.8)	0.0832 (10.2)
Population density	-0.0100 (-14.5)	-0.0186 (-18.6)	-0.0269 (-13.6)
1 Worker	0.4047 (10.5)	1.1085 (15.1)	1.7247 (7.3)
2 Workers	0.4885 (9.5)	1.3995 (17.3)	2.0992 (8.7)
3 or more Workers	0.7134 (7.4)	1.7209 (14.4)	3.1536 (12.2)
One Company Car	n.a.	1.3334 (27.6)	1.5830 (18.7)
Two Company Cars	n.a.	n.a.	2.0188 (15.7)
Purchase Costs	-0.0025 (fixed)	-0.0035 (fixed)	-0.0037 (fixed)
Running Costs	-0.0018 (fixed)	-0.0024 (fixed)	-0.0026 (fixed)
Final Likelihood		-36837.1	
Observations		45692	

Note: t-statistics w.r.t. zero shown in brackets

The MNL model shows a very good level of fit with a ρ^2 statistic with respect to constants of 0.3027. The model has intuitively signed coefficients estimated to a high degree of precision. The coefficients imply the following.

The specification of the household income needs some explanation. Firstly, to allow for the possibility that different household and area types with the same income may have different disposable income, a separate income coefficient is specified for each household and area type. Secondly, within a linear utility expression, a £1 increase in income has the same impact on utility whether the household has an income of £10,000 or £100,000. This restriction can be relaxed either by specifying income bands and estimating separate coefficients for each band or by specifying a non-linear utility expression with regard to income. Although there are many types of non-linear function (e.g. logarithmic, Box-Cox, quadratic, piecewise), the power function has some desirable properties. If λ is equal to one, the utility function is linear in response to changes in income, if λ is greater than one, the increase in utility following an increase in income is greater than proportional and, if λ is less than one the impact of a change income is less than proportional. Substantial improvements to model fit were achieved by transforming household income using a power term and the flexibility of the GAUSS software used to calibrate this model allows these parameters and their standard errors to be estimated directly.

Ownership propensity increases with income to the extent that in 1998 the market an elasticity of demand for cars with respect to household income was 0.33. This value is higher than equivalent estimates generated by the aggregate time series models but consistent with the findings of Dargay and Vythoulkas (1999), Dargay (2000) and Romily *et al* (1998). In general, income elasticities fall as income increase, are higher in Inner London, and are higher in households with three or more adults and where the head of the household is retried. The power terms on income are all significantly different from one and imply that the income elasticity of demand falls as income increases but that this reduction is less marked for two and three plus car households. The inclusion of income power terms improves the log likelihood from -37461.9 to –36837.1.

Single adult households retired or with children have a lower ownership propensity than other single-adult households. Two adult households have a higher ownership propensity than single adult households and as with single adult households retired status reduces ownership propensity but unlike single adult households, the addition of children increases ownership propensity. Three adult households have higher ownership propensity than two adult households but the addition of children reduces ownership propensity.

Ownership propensity decreases as population density increases. For example, for a single working adult living in an urban area earning £20K per year, the forecast number of cars per household ranges from 0.90 to 0.83 at population densities of 20 and 50 people per hectare respectively.

The influence of the number of workers was found to be non-linear. To take this into account, three dummy variables were specified to include: single worker households, two worker households and three plus worker households. A three person-household living in an urban area and earning £30K per year has an ownership propensity of 1.63 cars per household with one worker, 1.72 with two workers and 1.95 with three workers.

Ownership propensity for households with company vehicles is substantially higher than for households without company vehicles. To take this into account, additional dummy variables were added to the utility expressions for two and three plus cars. For two car households an additional dummy was set to one if one of the household’s cars is company owned and zero otherwise, and for three car households two dummy variables were specified: the first taking a value of one if one of the household’s cars is company owned and zero otherwise and a second dummy variable taking a value of one if two or more of the households vehicles are company owned and zero otherwise. For a three worker household living in an urban area and earning £30K per year, the ownership propensity is 1.95 cars per household if no cars are company owned, 2.34 if one car is company owned and 2.72 if two cars are company owned.

The NTS data does not contain suitable information on ownership and running costs for inclusion within the choice model, and therefore an alternate approach to incorporate costs is needed. One approach is to use published cost information to engineer ownership and running cost data for each vehicle in the household, however, without reference to the type and quality of vehicle, the cost information is pretty meaningless to the ownership decision. Another approach is to use the time-series cost indices used in the construction of aggregated models and estimate sensitivity to aggregate measures of ownership and use costs. Because there are only thirteen data points for ownership and running costs, and the fact that these are correlated with GDP and time, it was not possible to freely estimate cost coefficients. The costs coefficients were constrained to generate the same cost elasticites as shown by the power growth model developed in Section 2. In the absence of information on the strength of response for one, two and three plus car households, the relativities between coefficients are assumed to be equal to the relativities between the income coefficients.

Dogit Model Estimation

The dogit model (equation 8) supports an investigation into market captivity. The starting point for estimation was to adopt the same utility function specifications as in the multinomial logit and define a set of minimum non-car-ownership thresholds for each household and area type. Where minimum thresholds were not significantly different from zero, they were dropped and where they were not significantly different from each other, they were combined. This 'general to specific' approach led to the identification of minimum thresholds for single adult and retired couple households in London.

There is a strong positive relationship between income and car ownership but the relationship is dampened as income increases (evident from the relatively low income-power terms calibrated for the multinomial logit). This dampening of income's effect on ownership goes someway towards explaining why, 'money no object', not every household will acquire a car and because there is an overlap in the role played by the income-power term and the coefficient defining ownership thresholds it is difficult to estimate significant coefficients when both are included within a single model. Where the income-power coefficients are constrained to unity, bigger and more significant thresholds are estimated but the overall fit of the model is reduced when compared to the preferred model presented in Table 7.

Table 7: Dogit Model of Car Ownership

	1 Vehicle	2 Vehicles	3+ Vehicles
ASC	-9.2031 (-6.5)	-8.1998 (-24.6)	-9.9843 (-18.9)
Income	5.7555 (4.3)	0.6819 (5.6)	0.1707 (1.6)
Income (Outer London)	0.2328 (2.4)	0.1711 (4.2)	0.1238 (3.2)
Income (Metroploitan Areas)	-0.0780 (-0.9)	0.0702 (1.9)	0.0131 (0.4)
Income (Urban)	0.1031 (1.2)	0.1526 (4.1)	0.0874 (2.6)
Income (Rural)	0.3095 (3.3)	0.2906 (6.8)	0.1977 (4.6)
Income (One Adult, Retired)	-0.6036 (-11.4)	-0.7643 (-7.0)	0
Income (One Adult, with Children)	-0.2286 (-4.5)	-0.3981 (-4.0)	0
Income (Two Adults, Retired)	0.3605 (8.5)	0.5838 (11.8)	0.4401 (5.6)
Income (Two Adults, no Children)	0.2108 (5.4)	0.5586 (11.4)	0.4168 (5.7)
Income (Two Adults, with Children)	0.2855 (7.0)	0.6216 (12.0)	0.4289 (5.7)
Income (Three plus Adults, no Children)	-0.1528 (-3.2)	0.5703 (10.9)	0.6366 (6.5)
Income (Three plus Adults, with Children)	-0.0943 (-1.6)	0.5296 (10.4)	0.5297 (6.1)
Income Power Term	0.1519 (5.8)	0.4439 (19.7)	0.5688 (14.7)
time trend	0.0407 (12.5)	0.0819 (17.9)	0.0850 (10.4)
Population density	-0.0106 (-15.0)	-0.0193 (-18.9)	-0.0275 (-13.9)
1 Worker	0.3996 (10.2)	1.1078 (15.0)	1.7236 (7.4)
2 Workers	0.4739 (9.0)	1.3903 (17.0)	2.0907 (8.7)
3 or more Workers	0.7176 (7.2)	1.7316 (14.2)	3.1666 (12.2)
One Company Car	n.a.	1.3228 (27.4)	1.5717 (18.6)
Two Company Cars	n.a.	n.a.	2.0129 (15.7)
Purchase Costs	-0.0030 (fixed)	-0.0041 (fixed)	-0.0045 (fixed)
Running Costs	-0.0020 (fixed)	-0.0028 (fixed)	-0.0030 (fixed)
Theta Coefficients			
Inner London – Single Adult	0.2582 (3.6)		
Inner London – Two Adults, Retired	0.2674 (2.3)		
Inner London – Other	0.1105 (3.9)		
Outer London - Single Adult	0.0939 (2.4)		
Outer London - Two Adults, Retired	0.0928 (2.6)		
Outer London – Other	0.0166 (3.0)		
Final Likelihood	-36807.3		
Observations	45692		

Note: t-statistics w.r.t. zero shown in brackets

The dogit model shows similar properties to the multinomial logit with respect to the behavioural coefficients but gives additional insight into captivity. The estimation procedure generates values for θ (together with their standard errors) which can be used to obtain estimates of market captivity (Table 8).

For a given household type, households in inner London have lower maximum ownership propensity than households in outer London which in turn are lower than the rest of the country. Estimates from the 2001 Census confirm this finding showing the City of London with the highest proportion of non-car owning households (62%), followed by Islington (58%), Tower Hamlets (57%) and Westminster (56%). Areas with the lowest proportion of households without access to a car or van are Hart (8.8%,) Wokingham (9.2%) and Surrey Heath (10.5%). As would be expected single adult households in London have a lower maximum propensity to own vehicles than two-adult-retired households in London which in turn have a lower maximum propensity than 'other' households. It is important to note, however, that captivity is only present in approximately 10% of the NTS data.

Table 8: Minimum proportion of non-car owning households

Category	Mean Estimate	10th Percentile Estimate	90th Percentile Estimate	Percentage of Total Market
Inner London - Single Adult	20.5%	14.8%	25.8%	1.84%
Inner London – Two Adults, Retired	21.1%	11.4%	28.9%	0.27%
Inner London – Other	10.0%	7.2%	12.6%	1.67%
Outer London - Single Adult	8.6%	4.6%	12.3%	2.55%
Outer London - Two Adults, Retired	8.5%	4.8%	12.0%	0.58%
Outer London – Other	1.6%	1.0%	2.3%	3.64%

Mixed Logit Model

The final disaggregate ownership model to be estimated is the mixed logit model. This model has a similar specification to the multinomial logit but three notable differences.

Firstly, household income is included in absolute terms and not raised to a power. This is justified on the grounds that raising income to a power takes account of differences in behaviour across the sample with higher income households having lower income elasticities of demand. These differences in 'taste' can be specifically accommodated within the mixed logit models through additional error components without the need to transform the income variable.

Secondly, normally distributed coefficients were specified for population density, employment and company vehicle ownership to explain additional variation in the data. A random component was specified for income but this proved insignificant and was subsequently dropped.

Finally, error components were specified to take account of differences in the variance and covariance of choice options. Table 9 shows a covariance matrix for the choice set. In the first instance error components were specified for elements of the first off-diagonal of the covariance matrix allowing for correlation between contiguous alternatives (in a cross nested structure), but this proved unsuccessful as the model failed to converge. Next, error components were added to represent differences in the variance of choice options in a similar structure to the Heteroscedastic Extreme Value (HEV) model.

Table 9: Mixed Logit Model Covariance matrix

	No Car	One Car	Two Cars	Three plus Cars
No Car	0	ε_{01}	0	0
One Car	ε_{10}	ε_{11}	ε_{12}	0
Two Car	0	ε_{21}	ε_{22}	ε_{23}
Three plus Cars	0	0	ε_{32}	ε_{33}

The utility functions for the mixed logit model are shown below:

$$V_0' = V_0 \tag{13}$$

$$V_1' = V_1 + \sigma_{11}\xi_1 PD + \sigma_{12}\xi_2 E1 + \sigma_{13}\xi_2 E2 + \sigma_{14}\xi_4 E3 + \sigma_{17}\xi_{11} \tag{14}$$

$$V_2' = V_2 + \sigma_{21}\xi_1 PD + \sigma_{22}\xi_2 E1 + \sigma_{23}\xi_2 E2 + \sigma_{24}\xi_4 E3 + \sigma_{25}\xi_5 C1 + \sigma_{28}\xi_{22} \tag{15}$$

$$V_3' = V_3 + \sigma_{31}\xi_1 PD + \sigma_{32}\xi_2 E1 + \sigma_{33}\xi_2 E2 + \sigma_{34}\xi_4 E3 + \sigma_{35}\xi_5 C1 + \sigma_{36}\xi_6 C2 + \sigma_{39}\xi_{33} \tag{16}$$

Where:

V_0 is utility as expressed in equation 9

V_1 is utility as expressed in equation 10

V_2 is utility as expressed in equation 11

V_3 is utility as expressed in equation 12

E1 is equal to one if the household has one adult in employment, else zero

E2 is equal to one if the household has two adults in employment, else zero

E3 is equal to one if the household more than two adults in employment, else zero

C1 is equal to one if the household owns one company car, else zero

C2 is equal to one if the household owns two company cars, else zero

ξ is a vector or normally distributed error components

σ is a parameter vector to be estimated

The model was estimated in using 1,000 Halton Draws and the estimated coefficients are shown in Table 10.

Table 10: Mixed Logit Model of Car Ownership

	1 Vehicle	2 Vehicles	3+ Vehicles
ASC	-2.4101 (13.8)	-6.9209 (27.5)	-10.9282 (11.7)
Income	0.0704 (12.0)	0.0072 (1.0)	-0.0206 (1.1)
Income (Outer London)	0.0640 (10.1)	0.0785 (11.5)	0.0915 (7.9)
Income (Metroploitan Areas)	0.0650 (10.5)	0.0843 (12.5)	0.0810 (6.9)
Income (Urban)	0.0903 (15.4)	0.1098 (17.2)	0.1154 (10.4)
Income (Rural)	0.1187 (17.2)	0.1458 (19.7)	0.1567 (12.9)
Income (One Adult, Retired)	-0.0715 (11.2)	-0.2255 (6.9)	0
Income (One Adult, with Children)	-0.0134 (1.5)	-0.0913 (3.7)	0
Income (Two Adults, Retired)	0.0490 (8.5)	0.1306 (17.6)	0.1453 (7.7)
Income (Two Adults, no Children)	0.0234 (4.5)	0.1082 (16.7)	0.1280 (7.5)
Income (Two Adults, with Children)	0.0418 (7.6)	0.1319 (19.7)	0.1399 (8.3)
Income (Three plus Adults, no Children)	-0.0066 (1.0)	0.1002 (13.5)	0.1878 (8.6)
Income (Three plus Adults, with Children)	-0.0017 (0.2)	0.0962 (10.6)	0.1544 (7.6)
time trend	0.0351 (8.7)	0.0854 (15.3)	0.0929 (7.2)
Population density	-0.0149 (12.9)	-0.0341 (13.1)	-0.0703 (9.4)
1 Worker	0.6899 (11.1)	1.5147 (16.8)	0.7778 (0.9)
2 Workers	2.8934 (9.0)	4.0074 (12.1)	3.0659 (4.5)
3 or more Workers	4.1675 (6.4)	5.4016 (8.3)	6.1383 (6.9)
One Company Car	n.a.	1.8021 (17.9)	2.2845 (10.5)
Two Company Cars	n.a.	n.a.	3.3790 (5.0)
Purchase Costs	-0.0028 (fixed)	-0.0038 (fixed)	-0.0041 (fixed)
Running Costs	-0.0020 (fixed)	-0.0028 (fixed)	-0.0030 (fixed)
Standard deviation - Population density	0.0145 (6.9)	0.0155 (5.4)	0.0530 (9.1)
Standard deviation – 1 Worker	2.0467 (10.6)	1.4768 (7.5)	0.4660 (0.6)
Standard deviation – 2 Workers	4.7827 (10.3)	3.8541 (14.1)	6.6440 (9.2)
Standard deviation – 3 or more Workers	4.5531 (8.9)	5.4491 (9.9)	7.0011 (10.9)
Standard deviation – 1 Company Car	n.a.	1.9844 (8.4)	1.8030 (3.9)
Standard deviation – 2 Company Cars	n.a.	n.a.	2.0654 (2.3)
Variance Error Component	-0.0403 (0.4)	-0.0593 (0.2)	1.1483 (2.2)
Final Likelihood		-37206.0930	
Observations		45692	

Note: t-statistics w.r.t. zero shown in brackets

The inclusion of the error component terms improve the fit to the data when compared to the multinomial logit model with a linear income specification (this model is not reported here but had a log likelihood of -37461.9) but shows a worse fit to the MNL in which income is transformed via a power term (log likelihood -36837.1). The error component terms show significant variation across area and household type and show the three plus alternative to have a greater variance than the other choice alternatives. This confirms the findings from the nested logit model calibrations and is to some extent expected given the fact that three plus car owning households are not precisely defined (i.e. the include households with 3, 4, 5 or 6 vehicles).

Table 11 shows the elasticities of household car ownership with regard to household income, ownership costs and running costs for each of the estimated choice models.

Table 11: Disaggregate Model Elasticities

Model	Year	MNL	Dogit	Mixed Logit
Household Income	1985	0.445	0.453	0.441
	1990	0.387	0.396	0.415
	1995	0.372	0.378	0.392
	1998	0.329	0.329	0.349
Ownership Costs	1985	-0.083	-0.085	-0.083
	1990	-0.063	-0.065	-0.064
	1995	-0.057	-0.059	-0.058
	1998	-0.050	-0.053	-0.050
Running Costs	1985	-0.045	-0.045	-0.047
	1990	-0.039	-0.036	-0.035
	1995	-0.036	-0.036	-0.037
	1998	-0.035	-0.035	-0.036

The cost coefficients were all set to generate elasticities in line with those of the aggregate power growth model for 1985 and they show similar variation across each of the sampled years. What are more interesting is the similarity of the income elasticities across models. This is understandable for the MNL and dogit given that only 10% of the market is affected by captivity however there are more substantial differences in the specification of income in the mixed logit model, yet this does not translate to markedly different income elasticities.

3.3 Disaggregate Model Summary and Discussion

This chapter has described the development of a range of disaggregate choice models to explain and predict household car ownership as a function of a range of explanatory variables. The models are calibrated to data from the National Travel Survey over the years 1985 to 1998 which includes the ownership decisions of 45,692 households. The range of models developed includes a multinomial logit, a nested logit, a dogit model to account for captivity to 'non car ownership' and a mixed logit model to account for correlations between choice alternatives and a wider distribution of tastes across the sample. All of the calibrated models explain ownership of zero, one, two and three or more cars as a function of household income, household employment, household location, company car ownership, ownership costs, and running costs. To take account of differences in real purchasing power different income coefficients were specified by area and household type and in some specifications the impact of increases in income was dampened via an income power term. Due to problems in estimating sensible coefficients for motoring costs, each of the choice models is constrained to generate cost elasticities equal to those shown by the aggregate power growth function. The models show elasticities of demand with respect to income and motoring costs which are comparable with evidence found elsewhere and generate very good forecasts of market share when compared with actual data.

4 Model Application and Forecasts

The models developed in this paper are applied to generate forecasts under the following assumptions:

- GDP growth of 2.25% per annum;
- Population growth of 3.3% from 1996 to 2031;
- Household size falling by 17% from 1996 to 2031;
- Purchase costs reduction of 0.37% per annum; and
- Running costs remaining constant.

Table 12 and Table 13 provide a full set of car ownership forecasts shown as indices with the base set equal to 100. Although all sets of forecasts have been made using identical forecasting assumptions there is considerable differences between models. Even if we discount the forecasts from the rather unsatisfactory aggregate MNL and aggregate Dogit models, the forecast show increases in the number of cars ranging from 23.6% to 63.5%. In terms of growth in the number of cars per household, a number of models show limited growth over the forecast period. These models include the logistic (5%), modified logistic (7%), Gompertz (8%), NRTF1997 (12%), power growth (13%) and National Model (13%). Although household size is forecast to reduce from 2.46 people per household to 2.19 people per household, the relatively modest increases in ownership forecast by these models appears too low relative to the latest data. Of the remaining aggregate models, the constrained exponential model forecasts a 24% growth in the number of cars per household, and the lagged power growth forecasts an 18% increase. The three new disaggregate models forecast increases in the number of cars per household of 38%, 37% and 35% for the MNL, dogit and mixed logit model, which when translated to number of vehicles, show increases in the vehicle stock by 64%, 63% and 60%. Given that the models do not take account of increased road congestion, it seems likely that these increases are too high.

Table 12: Forecast Indices of Cars Owned– Aggregate Models

Year	Logistic	Modified Logistic	Power Growth	Gompertz	Const. Exp.	Lagged Power Growth	Aggregate MNL	Aggregate Dogit
2001	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
2006	105.5	106.4	107.6	106.7	109.1	107.9	114.2	111.5
2011	110.9	112.6	115.2	113.3	118.6	116.0	130.9	124.1
2016	115.6	117.8	121.9	119.0	127.6	123.7	148.3	136.8
2021	119.2	121.8	127.4	123.5	135.6	130.2	165.2	148.9
2026	122.1	124.7	131.7	127.0	142.5	135.6	181.4	160.1
2031	123.6	126.5	134.6	129.1	148.0	139.6	195.7	169.7

Table 13: Forecast Indices of Cars Owned – Disaggregate Models

Year	NRTF 1997	National Model	MNL	Dogit	Mixed Logit
2001	100.0	100.0	100.0	100.0	100.0
2006	108.9	107.7	111.2	111.1	110.3
2011	116.4	116.2	123.3	123.3	121.8
2016	122.7	122.6	135.1	134.8	132.8
2021	127.2	128.3	145.5	145.2	142.6
2026	129.9	133.6	155.3	154.8	151.9
2031	132.5	136.5	163.5	162.9	159.7

5 Summary and Conclusions

Given the importance of car ownership to transport and land-use planning and its relationship with energy consumption, the environment and health, growth in car ownership has been one of the most intensely researched transport topics over many years. It is the primary objective of this research to build on previous research to:

- Develop aggregate models of household car ownership;
- Develop disaggregate models of household car ownership; and
- Apply the models to generate forecasts of car ownership over the years 2001 to 2031.

Using published aggregate time-series data it has been possible to estimate a series of models that explain ownership as a function of time, GDP, motoring costs and average household size. This series of models comprise two distinct groups. The first group, known as aggregate ownership models include a series of functional forms previously employed by the UK government as well as the Gompertz function and the newly developed constrained exponential model. This group of models were calibrated using non-linear least squares procedures. The second group of models explain the proportion of households with zero, one, two and three or more vehicles as a function of GDP, motoring costs and household size. This group of models, referred to as aggregate share models, were estimated using maximum likelihood procedures and include the multinomial logit and the first application of the dogit model to aggregate car ownership data.

Significant findings of the aggregate models indicate:

- Saturation levels between 1.07 and 1.57 car per household (0.49 and 0.71 cars per capita);
- Elasticities of demand ranging from 0.08 to 0.29 for GDP, -0.02 to -0.06 for purchase costs and -0.02 to -0.08 for running costs, for 1990; and
- Ownership forecasts ranging from 1.18 to 1.35 cars per household by 2031.

The research also shows the development of a range of disaggregate choice models to explain and predict ownership at the household level. The models are estimated to data from the National Travel Survey over the years 1985 to 1998 and include the ownership decisions of 45,692 households.

All of the calibrated models explain ownership of zero, one, two and three or more cars as a function of household income, household employment, household location, company car ownership, ownership costs, and running costs. To take account of differences in real purchasing power, different income coefficients were specified by area and household type and in some specifications the impact of increases in income was reduced via an income power term. Due to problems in estimating sensible coefficients for motoring costs, each of the choice models is constrained to generate cost elasticities equal to those shown by the aggregate power growth function.

Significant findings of the disaggregate models show:

- Captivity to non-car ownership influences less than 10% of households;
- Income elasticity of demand for the market is 0.33 but there is considerable variation across household and area types;
- Population density has a significant impact on ownership, with ownership levels typically being higher in sparsely populated areas;
- An increase in the number of workers in a household increased the ownership propensity over and above the income related impact; and
- The acquisition of company cars increases the household's propensity to own additional cars.

When applied to generate forecasts, the findings show:

- The average number of cars per household is forecast to increase from 1.08 in 2001 to 1.49 in 2031;
- The total number of vehicles is forecast to increase by 64% to 43.5 million vehicles;
- There will be significant increases in the proportion of households with two cars and a modest increase in the proportion of households with three or more cars;
- In 2031 only 9% of households are forecast not to have access to a car; and

- There is a forecast increase in the proportion of new cars (<3 years old) in the vehicle stock and a forecast increase in average engine size.

References

Bates, J.J., Gunn, H., and Roberts, M. (1978) 'A Disaggregate Model of Car Ownership', DOE/DTp Research Report Number 20 London, HMSO.

Beggs, S. and Cardell, N.S. (1980) 'Choice of Smallest Car by Multi-vehicle Households and the Demand for Electric Vehicles', Transportation Research - Part B, Volume 14, Numbers 5/6, pp389-404.

Burns, L and Golob, T. (1975) 'An Investigation of the Role of Accessibility in Basic Transportation Choice Behavior', Research publication GMR-1900, Research Laboratories, General Motors Corp., Warren, MI.

Button, K.J., Pearman, A.D. and Fowkes, A.S (1982) 'Car Ownership Modelling and Forecasting', Gower, Aldershot, England.

Dargay, J. (1997) 'Modelling Car Ownership in France and the UK: a Pseudo-panel Approach', ESRC Transport Studies Unit, University College London.

Dargay, J and Gately. D (1999) 'Income's Effect on Car and Vehicle Ownership, Worldwide: 1960-2015', Transportation Research - Part A, No 33 pp101-138.

Dargay, J. and Vythoulkas, D. (1999) 'Estimation of a Dynamic Car Ownership Model: A pseudo Panel Approach', Journal of Transport Economics and Policy, Volume 33, Number 3, pp238-302.

Department of Transport (1989) 'National Road Traffic Forecasts (Great Britain) 1989', HMSO, London.

Gaudry, M. and Dagenais, M. (1979) 'The Dogit Model', Transportation Research – Part B, 13 (2), pp105-112.

Hanly, M. and Dargay, J.M. (2000) 'Car Ownership in Great Britain – A Panel Data Analysis', ESRC Transport Studies Unit, University Collage London.

Hensher, D.A and Le Plastrier, V. (1983) 'A Dynamic Discrete Choice Model of Household Automobile Fleet Size and Composition', Working Paper Number 6, Dimensions of Automobile Demand Project, Macquarie University, North Ryde, Australia.

Kain, J.F. and Fauth, G (1977) 'Forecasting Auto Ownership and Mode Choice for US Metropolitan Areas', Department of City and Regional Planning, Harvard University, Cambridge, MA.

Kirby, H.R. (1976) 'The Saturation Level of Car Ownership: Estimation Problems and a Regional Time-Series Analysis', PTRC Seminar Proceedings, Urban Traffic Models, PTRC, London.

Lave, C. and Bradley, J. (1980) 'Market Share of Imported Cars: a Model of Geographic and Demographic Determinants', Transportation Research - Part A Volume 14, Number 5/6, pp379-388.

Lave, C. and Train, K (1979) 'A Disaggregate Model of Auto-type Choice', Transportation Research - Part A, Volume 13, Number 1, pp1-9.

Lerman, S.R. and Ben-Akiva, M.E. (1976) 'A Behavioural Analysis of Automobile Ownership and Modes of Travel', Report Number DOT-05-3005603 prepared by Cambridge Systematics Inc, Cambridge MA, for the US Department of Transportation Office of the Secretary and Federal Highway Administration.

Mannering, F. and Winston, C. (1985) 'Dynamic Models of Household Vehicle Ownership and Utilization: an Empirical Analysis', Rand Journal of Economics.

Manski, C. and Sherman, L. (1980) 'An Empirical Analysis of Household Choice Among Motor Vehicles', Transportation Research - part A, Volume 14, Number 5/6, pp349-366.

Mogridge, M.J.H. (1978) 'The Effect of the Oil Crisis on the Ownership and Use of Cars', Transportation, Volume 7, pp45-68.

Mogridge, M.J.H. (1983) 'The Car Market', Pion, London.

Romilly, P., Song, H., and Liu, X. (1998) 'Modelling and Forecasting Car Ownership in Britain – A Cointegration ans General to Specific Approach', Journal of Transport Econoomics and Policy, Volume 32, Part 2 pp165-185

Smeed, R.J. (1951) 'Likely Increases of Road Traffic in Great Britain' Research Note RN/1518, Road Research Laboratory, Harmondsworth.

Tanner, J.C. (1958) 'An Analysis of Increases in Motor Vehicles in Great Britain', Research Note RN/1631, Road Research Laboratory, Harmondsworth.

Talvitie, A.D. (1973) 'Aggregate Travel demand Analysis with Disaggregate Travel Demand Models', Proceedings of the Transportation Research Forum, Volume 14, pp583-603.

Tanner, J.C. (1962) 'Forecasts of Future Numbers of Vehicle in Great Britain', Roads and Construction, Volume XL, pp263-274.

Tanner, J.C. (1974) 'Forecasts of Vehicles and Traffic in Great Britain', Transport and Road Research Laboratory Report LR 650, Transport and Road Research Laboratory, Crowthorne.

Tanner, J.C. (1975) 'Forecasts of Vehicles and Traffic in Great Britain: 1974 Revision', Transport and Road Research Laboratory Report LR 650, Transport and Road Research Laboratory, Crowthorne.

Tanner, J.C. (1977) 'Car Ownership Trends and Forecasts', Transport and Road Research Laboratory Report LR 799, Transport and Road Research Laboratory, Crowthorne.

Tanner, J.C. (1978) 'Long Term Forecasting of Vehicle Ownership and Road Traffic (with discussion)', Journal of the Royal Statistical Society, Series A, Volume 14, pp14-63.

Tanner, J.C. (1979) ‘Choice of Model Structure for Car Ownership Forecasting’, Transport and Road Research Laboratory Report SR523, Transport and Road Research Laboratory, Crowthorne.

Tanner, J.C. (1983) ‘A Lagged Model for Car Ownership Forecasting’, Transport and Road Research Laboratory. LR1072 Crowthorne.

Train, K. (1980) ‘A Structured Logit Model of Auto Ownership and Mode Choice’, Review of Economic Studies, Volume 47, pp357-370.

Train, K. (1986) ‘Qualitative Choice Analysis (Theory, Econometrics and an Application to Automobile Demand)’, MIT press, Cambridge MA.

Transport Statistics Great Britain (2002) Department for Transport.

Wolff, P. de (1938) ‘The Demand for Passenger Cars in the United States’, Econometrica, pp113-129.

Wooton, H.J. and Pick, G.W. (1967) ‘A Model for Trips Generated by Households’, Journal of Transport Economics and Policy, Volume I pp137-153.