Research Report

Estimation of the National Car Ownership Model for Great Britain

2011 base

James Fox, Bhanu Patruni, Andrew Daly, Hui Lu

RAND Europe

RR-1494-DfT
October, 2016
Prepared for UK Department for Transport

RAND EUROPE

Preface

This report documents the estimation and associated Quality Assurance of the updated and enhanced national car ownership models for Great Britain. This work was funded by the UK Department for Transport, and RAND Europe's work was undertaken as part of a wider project, led by Atkins, to update the National Trip End Model of which the car ownership model forms part.

This report is the first of four related deliverables that RAND Europe have either produced or contributed to for this study:

Number	Deliverable reference	Report title	Report description
1	D19	*Estimation and Quality Assurance of the National Car Ownership Model for Great Britain: 2001 base*	Technical note describing the re-estimation of the Department for Transport's national car ownership model and evidence of the associated QA
	D20	*Licence Cohort Model – Appendix to Estimation Report*	Description of the formulation, estimation and use of the licence cohort model, including the relevant QA
2	D11	*Software Developer's Note and QA*	Developer's note and QA evidence to accompany updated NATCOP software
3	D12	*The NATCOP3 Programme*	User guide for NATCOP software
4	D21	*NATCOP Outputs QA and High Level Comparison*	Results from the updated NATCOP model including performance comparisons and evidence of QA

This report is intended for a technical audience familiar with transport modelling terminology and approaches. For more information about this report please contact:

Dr James Fox

RAND Europe

Westbrook Centre

Milton Road

Cambridge

CB4 1GN

jfox@rand.org

Table of Contents

Figures

Tables

Summary

Introduction

The UK Department for Transport's (DfT's) national car ownership models (NATCOP) have been updated to reflect a 2011 base year, and enhanced to take account of the DfT's experience in applying the previous version of the models (2001 base).

Modelling framework

A brief review was undertaken to consider different approaches to car ownership model types drawing on a few key sources. This review demonstrated that the NATCOP approach of developing household-level disaggregate models of car ownership has been used in a wide range of national and urban studies since the 1980s, and allows the impact of a range of socio-economic and other variables on car ownership to be incorporated.

The household car ownership decision is modelled as a series of linked choices:

- The choice between owning zero and one-plus cars (P_{1+})
- The choice between one and two-plus cars (P_{2+})
- The choice between two and three-plus cars (P_{3+}).

Each of these linked models incorporates a saturation term that accounts for the fact that a fraction of households will never choose to own cars.

Car ownership data

Choice data

The models were estimated from three sets of choice data:

- Family Expenditure Survey (FES) data at five-year intervals from 1971 to 1996 and in 1997/98, 1998/99, 1999/00 and 2000/01
- Expenditure and Food Surveys (EFS) data from 2001/02, 2002/03, 2003/04 and 2004/05
- National Travel Survey (NTS) data from 1999 to 2014.

Analysis of the evolution of the proportions of households owning zero, one, two and three-plus cars over the 1970–2015 period demonstrated that the fraction of households owning one car has remained remarkably constant at around 45 per cent. However, the proportion of households owning no car has fallen from just under half to just under one-quarter, and correspondingly the proportions of multi-car households have increased considerably.

Purchase and running cost data

Purchase and running cost data from 1970–2015 was also assembled. The general trend over the period has been for purchase costs to decline but for running costs to increase in real terms. Significant changes in running costs were observed between 2001 and 2011 which were explored further in the review of model performance and specification.

Review of model performance and specification

Model validation by area type and population density

A review of the previous 2001 base version of the model was undertaken in the first phase of this project to inform the development of NATCOP during the second phase.

Validation of *total* car ownership predictions for 2011 demonstrated that the model performed well across Great Britain as a whole, and reasonably well for the four non-London area types. However, for London the model over-predicted ownership and further investigations demonstrated that the predictive performance was worst in Inner London.

Analysis of the predictions for zero, one, two and three-plus cars revealed a more complex picture, specifically:

- Consistent under-prediction of zero-car households across all area types, which is important in the context of forecasting public transport demand as members of zero-car households are much more likely to travel by public transport than members of car-owning households;
- Consistent over-prediction of one-car households, when in fact this fraction has remained stable over a long period of time;
- Outside of London a general pattern of under-prediction of multiple-car households, particularly those owning three-plus cars.

In addition to the area type validation, the models were validated by examining how the predicted probabilities of the zero-, one-, two- and three-plus-car alternatives varied by population density. This validation demonstrated that while the over-prediction of one-car households persists across the whole range of observed population densities, the errors in multiple car ownership show a clear relationship with population density with car ownership over-predicted in the densest areas.

To investigate the performance of the model in London further, the relationship between car ownership and population density was explored for each of the individual London boroughs. This demonstrated that there was an 'Inner London' effect in addition to the population density effect, which reduced the likelihood of car ownership, probably reflecting factors such as higher congestion, constraints on parking supply, the impact of the congestion charge and high levels of public transport (PT) accessibility.

Validation of the predictive performance of NATCOP by population density across all area types demonstrated a general tendency to over-predict multiple car ownership in densely populated areas.

As a result of this analysis a key recommendation in the first phase of the study was to test separate area types for Inner and Outer London, as well as population density terms across all area types.

Review of exogenous model inputs

A review was undertaken to compare predicted changes in purchase and running costs over the 2001 to 2011 period to those observed over the same period. For purchase costs, the observed reduction in costs was forecast well. However, while running costs were assumed to remain constant over the forecast period, in fact significant increases in running costs (maintenance, fuel and tax and insurance) were observed over the period.

Company car ownership is represented in the models through terms that reflect the higher probability of households owning multiple cars if they own one or more company cars. When the models were applied from a 2001 base to predict car ownership in 2011 it was assumed that there would be no change in company car ownership over the decade. However, as a result of taxation changes company cars fell from around 10 per cent of total cars to just over 8 per cent of total cars, and this means that in model application the assumed company car ownership level for 2011 was an over-prediction, which in turn contributed to the general pattern of over-prediction of multiple car ownership in 2011.

Review of saturation levels and income

A review of the formulation of saturation in the model concluded that the formulation used in the 2001 base version of NATCOP is sound; specifically, the model formulation directly incorporates saturation and gives the expected result that the marginal impact of income reduces as income increases. The recommendation of distinguishing Inner and Outer London area types ensures that the model specification can represent lower saturation rates in Inner London.

Access to public transport

The impact of access to public transport on car ownership was investigated using NTS choice data. This analysis demonstrated walk access effects for both train and bus, with bus having four times the disutility per minute compared to train, consistent with shorter average access distances for bus. On the basis of these results we recommended that tests be undertaken to assess the impact of these terms in addition to the other enhancements during the model estimation work.

Model development

Phase 2 of the project aimed to update and enhance the NATCOP models building on the Department's experience of applying the models and the Phase 1 review of model performance.

Data availability

As described above, the models were estimated using a combination of FES, EFS and NTS data. Some of the model variables from the previous NATCOP specification could only be defined for some choice datasets, specifically area type information and company car ownership. Furthermore, the additional detailed licence holding variables, the separate Inner and Outer London area types and the population density terms could only be estimated from the 1999–2014 NTS data.

Saturation terms

The saturation terms in the models vary with area and household types. As per the previous version of NATCOP, in the final model specification saturation terms are estimated for each possible combination of area and household type. The appropriate level of aggregation was determined by first estimating terms for each possible combination, and then aggregating the terms across similar areas or household types as appropriate.

For the P_{1+} and P_{2+} models the saturation terms in the new models represent significantly lower saturation levels in Inner London compared to Outer London. For the P_{3+} model, only a single saturation term has been estimated, which is consistent with the previous versions of the NATCOP model.

London area types and population density terms

As described above, the new models capture variation in saturation levels between Inner and Outer London area types. In principle the model specification is able to capture variation in income sensitivity between Inner and Outer London; however, the variation in income sensitivity between Inner and Outer London was not statistically significant.

The population density terms capture variation in car ownership behaviour over and above that represented by the variation in saturation and income sensitivity with area type. In all three models statistically significant terms have been identified that capture that the probability of owning cars decreases as population density increases.

Public transport accessibility and parking terms

Using the NTS data, it was possible to identify significant PT accessibility terms, reflecting lower car ownership levels for households with good public transport accessibility. However, it was decided not to implement these terms on the basis that the improvements in model fit were relatively modest and because it would be difficult and time-consuming to make forecasts of how PT accessibility might evolve in the future.

For parking, while the NTS data collects parking information at the destination, the household data does not record information on parking cost and/or residents' parking schemes. Furthermore, even if such information were to be available it would again be difficult and time-consuming to assemble future forecasts of parking costs. Therefore no (household) parking terms have been included in the final model specifications.

Improved treatment of licence holding

One of the key improvements to the new NATCOP model is an enhanced treatment of licence holding that has been achieved by the development of a licence cohort model. The cohort model was documented in full in D20, *Licence Cohort Model*, which is included as an appendix to this report. In summary, in addition to the Great Britain average licences per adult (LPA) time trend term used in previous versions of NATCOP, cross-sectional variation in licence holding by age band and gender cohort has been incorporated in the model specification. In implementation, the cohort model provides a mechanism for the models to take account of future changes in licence holding such as higher licence-holding rates for older females.

Acknowledgements

We would like to acknowledge the contribution of the Department, which funded this research, and in particular Dharmender Tathgur, Robin Cambery and Pawel Kucharski. We would also like to acknowledge the contributions of the two external peer reviewers, John Bates and Peter Davidson, who provided useful comments on the Phase 1 report that we produced to inform the model development work, and have also provided comments on this report.

We would also like to acknowledge the helpful contributions of the two RAND Europe reviewers, specifically Charlene Rohr in the continuous review role and Greg Erhardt in the output review role.

Finally, we would like to acknowledge the contribution of Atkins, who have assisted us with supply of data and overall project management, and in particular the contributions of Dave Williams and Clare Lindsay.

Abbreviations

AT	Area Type
DfT	UK Department for Transport
GB	Great Britain
ITS	the Institute for Transport Studies, University of Leeds
HH	Household
EFS	Expenditure and Food Survey
FES	Family Expenditure Survey
ITS	Institute for Transport Studies
LPA	Licences Per Adult
NATCOP	National Car Ownership Model
NRTF	National Road Traffic Forecasts
NTEM	National Trip End Model
NTS	National Travel Survey
QA	Quality Assurance

1. Introduction

1.1. Aims of study

The aims of this study were to update and enhance the Department for Transport's NATCOP models that are used to model household car ownership as part of the National Trip End Model (NTEM) suite. The models have been updated to reflect a 2011 base year, rather than the 2001 base year used in the previous model, by incorporating more recent National Travel Survey (NTS) data. A number of enhancements have been made to the model specification in light of the issues with the previous NATCOP model that the Department identified in the brief,[1] specifically:

- The previous model is known to predict higher levels of car ownership than are observed in dense urban areas, particularly London, and will require investigation into improving that capability;
- The previous model may potentially be improved by providing information on how PT and parking space provision may impact on the decision to own and operate a vehicle, particularly in those denser areas;
- The appropriateness of the saturation rates in terms of how they are implemented and the validity of their current values should be explored;
- Recent behavioural trends in car ownership, particularly the decline in young males owning driving licences (and a relative increase in female drivers) are not captured in the previous model's methodology; it should be considered how this may improve the forecasts and if it is warranted to be included in the model – this also suggests that it may be necessary to review the age segmentation within this model, and indeed the Scenario Generator;
- Analysis by the NTM team has shown that although there has been no sudden break between income and car ownership, there has been a long weakening of the relationship – it should be investigated whether or not this effect can be included in the model, or further explanatory variables added; and
- The treatment of company cars in the model should be reviewed.

The age segmentation used to implement the model is not described in this report. The implementation of the NATCOP model in the NTEM suite is documented separately in D21, *Software Developer's Note and QA*, and D12, the *NATCOP3 Programme*.

[1] RM494 SO4717 *National Trip End Model Dataset Update*, DfT, Appendix B – Specification.

1.2. Structure of this report

Chapter 2 outlines the modelling framework used for NATCOP models, which are disaggregate household-level models of car ownership incorporating saturation. It also summarises how the household-level utility functions are defined.

Chapter 3 describes the data used for model estimation, outlining both the choice data capturing household-level car ownership choices, and the supporting car ownership cost data.

Chapter 4 presents a review of the performance of the previous 2001 base version of NATCOP. This is a summary of the Phase 1 report for this study that guided the subsequent Phase 2 work to update and enhance the NATCOP models.

Chapter 5 documents the model development process, data availability issues, the identification of the appropriate saturation terms, treatment of car ownership levels in London and other densely populated areas, public transport accessibility and parking terms, and incorporation of an improved treatment of licence holding.

Chapter 6 provides a summary of the model development process and sets out some recommendations for further work.

Appendix A summarises the methodology used to estimate saturation rates in NATCOP. Appendix B describes the QA procedures followed in this project. Finally, Appendix C documents the new licence cohort model.

2. Modelling framework

2.1. Review of car ownership modelling approaches

A useful overview of car ownership models developed for the public sector is provided by de Jong et al. (2004), who identify ten different types of car ownership models (summarised in Table 1 below).

Table 1: Car ownership model types

Model type	Level of aggregation	Static or dynamic	Long- or short-run forecasts	Car use	Car types	Data requirements
Aggregate time series models	aggregate	dynamic	short, medium and long	not included	not distinguished	light
Aggregate cohort models	aggregate	dynamic	medium and long	not included	none	light
Aggregate car market models	aggregate	dynamic	short, medium and long	not included	limited	light
Heuristic simulation methods	disaggregate	static	medium and long	can be included	limited	moderate
Static disaggregate car ownership models	disaggregate	static	long	included in some models via logsum	very limited	moderate
Indirect utility car ownership and use models	disaggregate	static	long	included	often many (brand–model–age)	heavy
Static disaggregate car-type choice models	disaggregate	static	long	included in some models via logsum	very limited	heavy
Panel models	disaggregate	dynamic	short and long	sometimes included in ad-hoc fashion	very limited	very heavy
Pseudo-panel methods	aggregate	dynamic	short and long	not included, but could be	very limited	moderate
Dynamic transaction models	disaggregate	dynamic	short and medium	sometimes included in ad-hoc fashion	very limited in duration model, many in usage model	very heavy

Source: Adapted from de Jong et al. (2004).

Table 2 describes the following characteristics for the ten model types:

Table 2: Model characteristics

Level of aggregation	Whether the models were developed from aggregate-level data (e.g. total fleet by car type) or disaggregate-level information (e.g. car ownership information at the household level)
Static or dynamic	Dynamic models explicitly predict changes over time, whereas static models usually make predictions for a given point in time typically assuming equilibrium at that point in time
Long- or short-run forecasts	Whether the models can be used to make long term forecast (10–20 years), or to assess shorter term impacts
Car use	Whether car usage (typically kilometres/miles) is modelled
Car types	Whether car type choice, e.g. by fuel type, engine size, etc., is modelled
Data requirements	How much data is required to develop the models, e.g. is detailed vehicle-level information required

According to de Jong et al.'s (2004) classification, NATCOP is classed as a disaggregate static model of car ownership used to provide long-run predictions of the total car fleet.

A key consideration in the choice of modelling approach for car ownership is the intended usage of the model, and in particular whether the model is required to produce forecasts of car type choice and usage. The Department maintains other models that are used to predict car type choice and usage, and so the role of NATCOP is to make long-run predictions of the total car fleet. A disaggregate static approach is therefore appropriate.

De Jong et al. (2004) describe applications of the static disaggregate approach used in NATCOP that date back to work on the Dutch National Model in the early 1980s. A number of similar models were developed in the late 1980s and early 1990s, including models for the Italian, Swedish and Danish national transport model systems, and models for Paris and Stockholm. Subsequent applications include Sydney (Tsang & Daly, 2011) and the PRISM model for the West Midlands (Fox et al., 2014), and of course the original work to develop the disaggregate NATCOP models (Whelan, 2001 & 2007).

The key advantage of disaggregate approaches over aggregate approaches is that they allow household-level socio-economic influences on behaviour to be represented. Particularly important in the context of car ownership is household income, but licence holding, number of workers and other socio-economic factors have also been identified in the models reviewed.

A more recent review of car ownership modelling approaches was presented by Anowar et al. (2014). They classify disaggregate household-level models of car ownership such as the NATCOP models as exogenous static models, as the car ownership decision is considered in isolation of other choices, such as mode or destination choice. As such, in model application car ownership forecasts can be made without linkage to the mode and/or destination choice models.[2] They reference a number of different studies that

[2] However, it should be noted that some studies have identified a significant linkage between commute mode-destination accessibility and household-level car ownership.

have developed car ownership models of this type, and note that as well as socio-economic characteristics of the household and its members, these models have incorporated variables to reflect variation in car ownership according to the built environment and public transport accessibility.

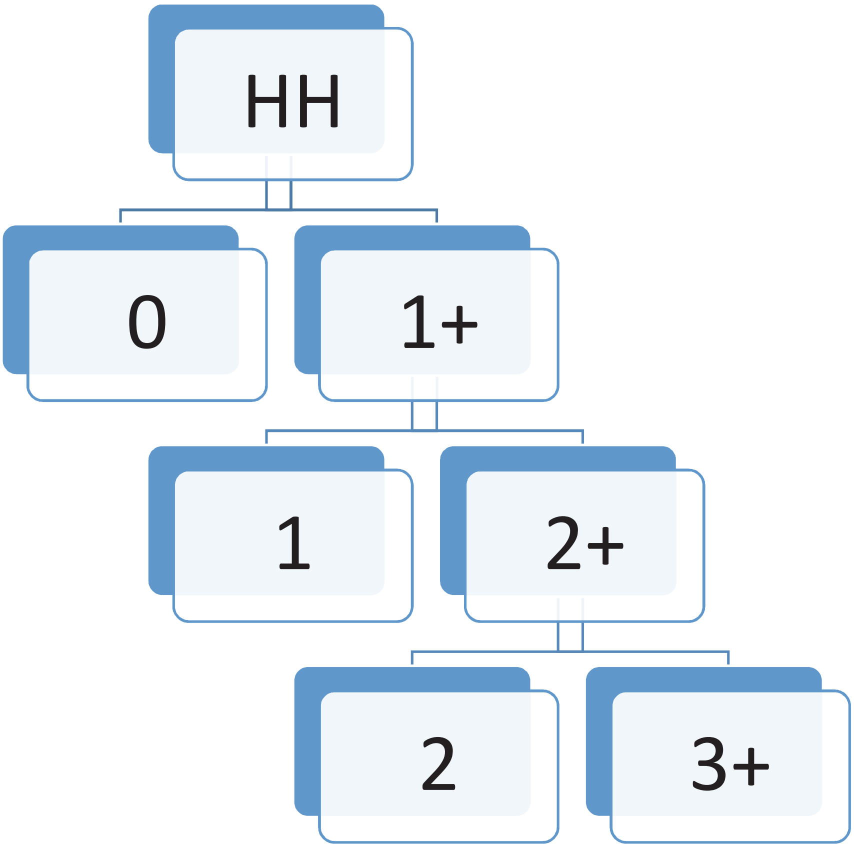
In summary, disaggregate household-level car ownership models have been used in a wide range of national and urban studies since the 1980s, and allow the impact of a range of socio-economic and other variables on car ownership to be represented. Thus the NATCOP modelling approach is consistent with the approach used across a range of national, regional and urban contexts.

2.2. Model structure

The specification used in the previous version of NATCOP (where 2001 was the base year) was originally developed by ITS, University of Leeds (Whelan et al., 2001). An update of the model was subsequently carried out by MVA Consultancy (2007) to use more recent data, but no changes were made to the model specification in that work.

NATCOP represents the household decision as to whether to own zero, one, two or three or more vehicles. It is noted that vehicles include both privately and company-owned vehicles. These household (HH) choices are represented though three linked binary models as illustrated in Figure 1.

Figure 1: NATCOP model structure



Working from the top, the first model predicts the binary choice between owning zero or one-plus vehicles; if one-plus vehicles is chosen, then the second model predicts the binary choice between owning one or two-plus vehicles; if two-plus cars is chosen, then the third model predicts the binary choice between owning two or three-plus vehicles.

The term 'vehicles' is used deliberately in this section because in addition to cars motorcycles/scooters/mopeds and Land Rover/Jeep and light van vehicle types are included. However, for simplicity these groups are collectively referred to as 'cars' in the remainder of this report.

2.3. Model specification

This section summarises the core model specification that was used in both the original 2001 ITS work and the subsequent 2007 MVA-updated work. This core model specification has been retained for the new model with some additional variables added. These are discussed in Chapter 6.

The original 2001 ITS work gave careful consideration to the issue of saturation and the development of the approach used in NATCOP to model saturation is documented in full in the report from that study (Whelan et al., 2001). The rationale behind representing saturation in the context of car ownership is that a fraction of households will never acquire a car for a variety of reasons such as health reasons, individual preferences and so on. Quoting from Whelan (2007):

> The importance of market saturation within car ownership models was highlighted by the Leitch Committee, who noted 'that the accurate determination of the saturation level is of prime importance if the resulting forecasts are to command confidence. If the saturation level cannot be satisfactorily determined then the resulting forecasts are to that extent themselves unsatisfactory' (Department of Transport, 1978).

Daly (1999) showed how it was possible to set up a partially constrained choice model for a binary choice situation to represent a fraction of decisionmakers who are captive to particular alternatives, for example a fraction of households that will never own a car. The equations that underlie this approach are detailed in Appendix A. This novel approach for representing saturation was then incorporated in the NATCOP modelling approach allowing saturation levels – varying by household and area type – to be directly estimated from the data. The approach accounts for variation in saturation by household and area type drawing on evidence from the original ITS work that there are significant differences in saturation across these dimensions (Whelan et al., 2001).

The probabilities associated with the different car ownership probabilities incorporating saturation levels are expressed as follows:

$$P_{1+} = \frac{S_{1,a,\mathrm{h}}}{\left(1 + \exp(-V_{1+})\right)} \tag{2.1}$$

$$P_{2+|1+} = \frac{S_{2,a,\mathrm{h}}}{\left(1 + \exp(-V_{2+|1+})\right)} \tag{2.2}$$

$$P_{3+|2+} = \frac{S_{3,a,\mathrm{h}}}{\left(1 + \exp(-V_{3+|2+})\right)} \tag{2.3}$$

The utility functions used in each of these probability expressions are calculated as follows:

$$V_{1+} = ASC_1 + b_1 LPA + (c_1 + c_{h1} D_h + c_{a1} D_a) Y + d_1 E + e_1 O + f_1 R \tag{2.4}$$

$$V_{2+|1+} = ASC_2 + b_2 LPA + (c_2 + c_{h2} D_h + c_{a2} D_a) Y + d_2 E + e_2 O + f_2 R + g_{21} CC_1 \tag{2.5}$$

$$V_{3+|2+} = ASC_3 + b_3 LPA + (c_3 + c_{h3} D_h + c_{a3} D_a) Y + d_3 E + e_3 O + f_3 R + g_{32} CC_2 \tag{2.6}$$

where: P_{1+}, $P_{2+|1+}$ and $P_{3+|2+}$ are the car ownership probabilities

S is the estimated saturation level by ownership state, area type a and household type h

ASC_1, ASC_2 and ASC_3 are alternative specific constants

LPA is the average driving licences per adult (LPA) for GB as a whole (this varies by year)

Y is gross household income

D_h is a vector of household type constants

D_a is a vector of area type constants

E is the number of adults employed in the household

O is a purchase cost index (this varies by year)

R is a running cost index (this varies by year), which includes fuel, maintenance, tax and insurance costs

CC_1 is a constant if there is one company car in the household

CC_2 is a constant if there are two company cars in the household

b, c, d, e, f, g are parameter vectors that have been estimated.

Eight household types h are distinguished, defined as a function of the number of adults, whether those adults are retired and the presence of children:

1. One adult, not retired
2. One adult, retired
3. One adult, with children
4. Two adults, retired
5. Two adults, no children
6. Two adults, with children
7. Three or more adults, no children
8. Three or more adults, with children.

Five area types a are represented:

1. Greater London
2. Metropolitan districts
3. Non-metropolitan districts, population density greater than 10 pers/ha
4. Non-metropolitan districts, population density 2–10 pers/ha[3]
5. Non-metropolitan districts, population density less than or equal to 2 pers/ha.

It should be noted that while household income, household type, employed adults and company car ownership are household-level variables, in the previous model the LPA measure was a GB-wide average value for adults that varied only by year, thus reflecting changes in aggregate licence holding over time and *not* cross-sectional variation in licence holding between households. In the new model an enhanced treatment of licence holding has been developed that *does* take account of cross-section variation; this is discussed in Section 5.7. It should also be noted that the purchase and running cost indices vary only by year and so will not pick up effects such as higher insurance costs for younger drivers (except in so far as they impact on the overall average insurance cost).

[3] In the original ITS work this band was defined as covering non-metropolitan districts with population densities between 2.22 and 7.9 persons per hectare. The definitions appear to have been revised by MVA in its 2007 work, but that report does not explain why the change was made (MVA, 2007).

3. Car ownership data

3.1. Choice data

The original 2001 ITS work assembled choice data spanning the period 1971 to 1996 for the estimation of car ownership models, specifically:

- Family Expenditure Survey (FES) data at five-year intervals from 1971 to 1996; and
- NTS data from 1991.

When MVA re-estimated the models in 2007 they used the following datasets:

- FES data at five-year intervals from 1971 to 1996 (as per the ITS work) plus 1997/98, 1998/99, 1999/00 and 2000/01;
- Expenditure and Food Surveys (EFS) data from 2001/02, 2002/03, 2003/04 and 2004/05; and
- NTS data from 1999 to 2004 (the 1991 NTS data was dropped).

To estimate the new NATCOP models, the dataset assembled by MVA was supplemented by more recent NTS data. The estimation dataset comprised:

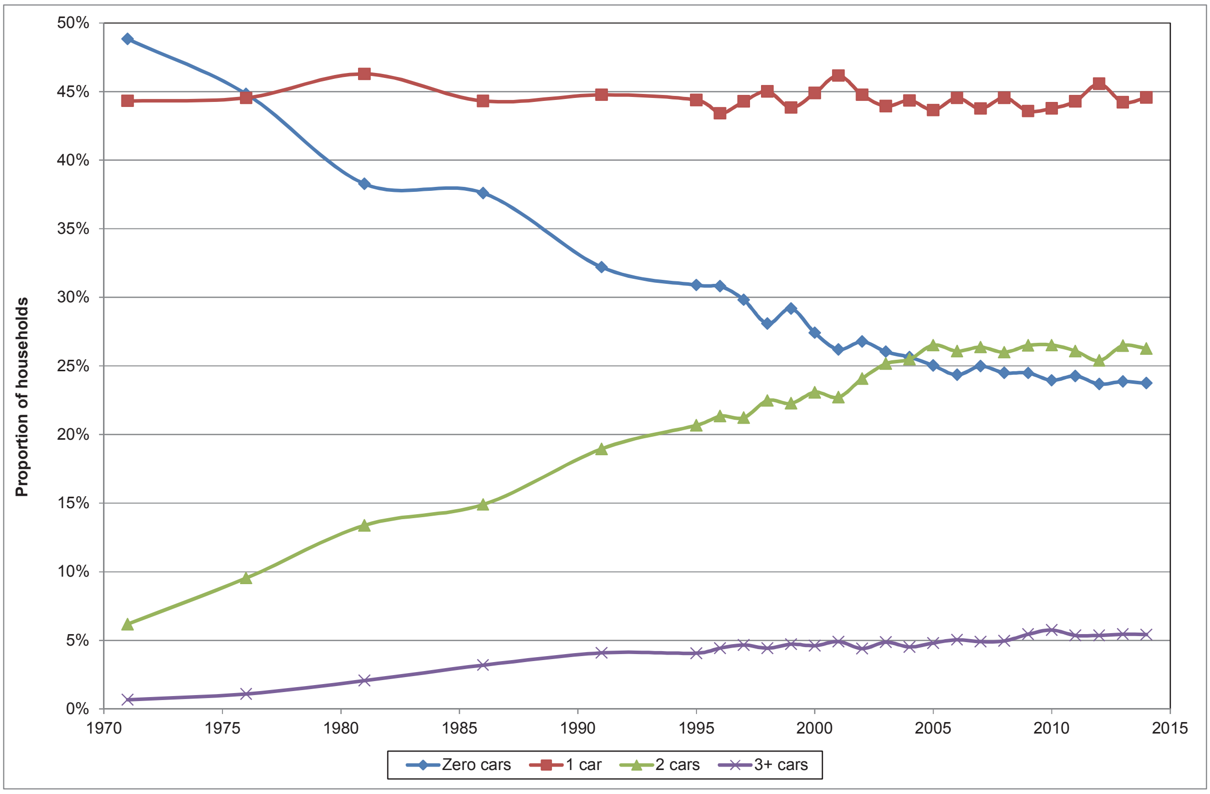
- FES data at five-year intervals from 1971 to 1996 (as per the ITS work) plus 1997/98, 1998/99, 1999/00 and 2000/01 FES data;
- EFS data from 2001/02, 2002/03, 2003/04 and 2004/05; and
- NTS data from 1999 to 2014.

The NTS data covering the 1999–2014 period provided a substantial sample of more recent household data, with observed car ownership information from a total of 126,800 households. Furthermore, the NTS data provided the most comprehensive range of variables to support model enhancement. Therefore it was decided to update the estimation sample relative to the sample used by MVA in 2007 using NTS data alone.

For some years, the FES and EFS data did not provide the household location information required to classify households into area types; this issue is discussed further in Section 5.1.

The trends in the observed proportions of households choosing the zero-, one-, two- and three-plus-car alternatives are plotted in Figure 2. These figures are unweighted, and as such will be impacted by any biases between the sample of households surveyed in the estimation sample for each year surveyed and the actual number of households in the GB population each year.

Figure 2: Proportion of households owning cars by year



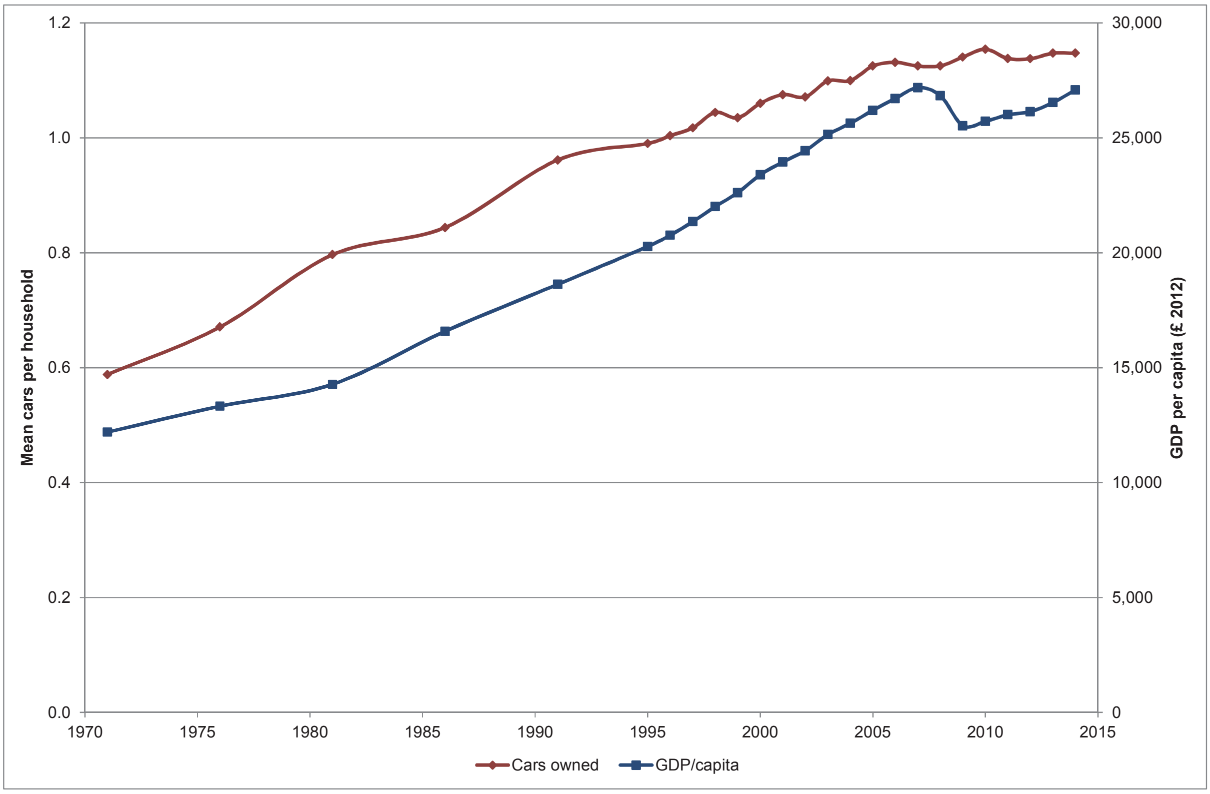
Source: Estimation samples of FES, EFS and NTS data detailed earlier in Section 3.1.

The proportion of households owning one car remained remarkably constant between 1971 and 2014, at around 45 per cent. However, the proportion of households owning no cars fell from just under half to just under one-quarter, and correspondingly the proportions of multi-car households increased considerably.

The net effect of these changes on mean car ownership per household is plotted in Figure 3 alongside changes in real GDP per capita over the same period.[4] Again, these are unweighted figures.

[4] www.ons.gov.uk, GDP data: ABMI series, population data: CDID series.

Figure 3: Average household car ownership and GDP/capita by year



Sources: Car ownership estimation samples of FES, EFS and NTS data detailed earlier in Section 3.1, GDP: ABMI series, Quarterly National Statistics (downloaded from www.ons.gov.uk).

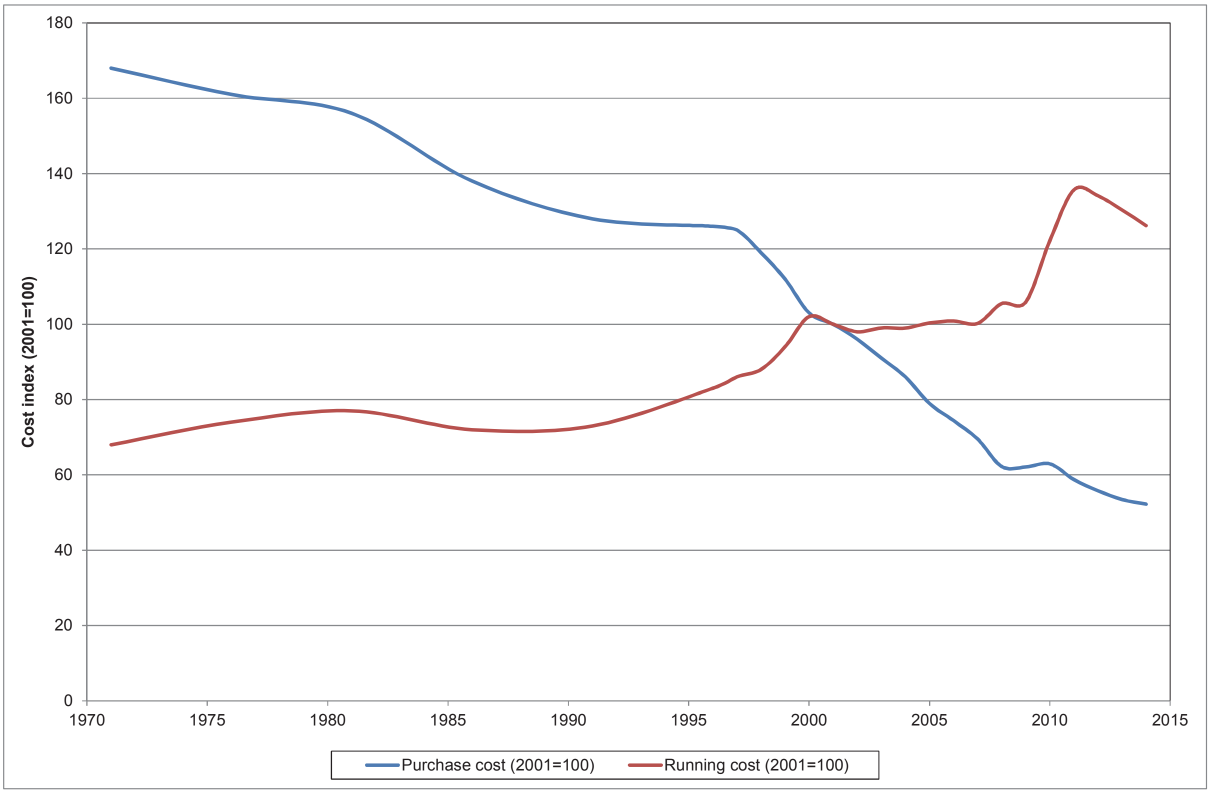
Total car ownership rose fairly steadily between 1981 and 2007. Car ownership more or less levelled off from 2007, but as the GDP per capita line illustrates this was in the context of a significant fall in GDP between 2007 and 2008. By 2014 GDP per capita was still slightly below the 2007 peak.

3.2. Cost data

The previous NATCOP models incorporated terms that accounted for changes in national average purchase and running costs over time. Cost data for the 1971–2004 period was assembled by MVA (2004). Cost data for the 2005–2014 period was assembled from transport expenditure survey statistics.[5] The variation in purchase and running costs over the entire 1971–2014 period is plotted in Figure 4.

[5] https://www.gov.uk/government/statistical-data-sets/transport-expenditure-tsgb13, accessed 23/10/15.

Figure 4: Real purchase and running costs by year



Source: MVA (2004) and transport expenditure survey statistics.

The general trend over the period was for purchase costs to decline in real terms but for running costs to increase in real terms. There were significant changes in purchase and running costs from 2001, the base year for the previous version of NATCOP; in particular, purchase costs fell significantly but there were increases in running costs (increases in tax and insurance costs in addition to increases in fuel cost). These changes are explored further in Section 4.2.1 by investigating changes in the different components of running costs.

4. Review of previous model performance and specification

This Chapter presents a review of the performance and specification of the previous version of NATCOP (2001 base). The review was undertaken during Phase 1 of this project to inform the development of the new version of NATCOP in Phase 2. A key part of the review work was to compare the predictions of the previous (2001 base) version of the model for 2011 to observed Census car ownership data.

Section 4.1 presents a validation of the predictive performance of the models by the five area types distinguished in the old model, and the performance of the models by population density, relating closely to the different area types. Section 4.2 reviews some of the exogenous inputs to the model, specifically purchase and running cost information and company car ownership inputs. Section 4.3 presents a review of the treatment of saturation and investigates the impact of public transport accessibility on car ownership. Section 4.4 discusses analysis of the impact of access to public transport on car ownership, and Section 4.5 considers the introduction of parking space terms into the model specification. Section 4.6 discusses the treatment of licence holding. Finally, Section 4.7 provides a set of recommendations for model development.

4.1. Validation by area type and population density

The NATCOP predictions for 2011 have been compared to observed car ownership levels from the 2011 Census. The Census information assembled for the validation is at district level, and so the NATCOP predictions for 2011, which are at the 2496 TEMPRO v6 zone level, have been aggregated up to district level.

4.1.1. Validation by area type

Table 3 presents a validation of the NATCOP predictions for zero-, one-, two- and three-plus-car household states across the five area types currently represented in the model. The table also presents a validation of total household car ownership by area type. The validation deliberately works with the probability of each car ownership state rather than with total households by state to remove the effect of differences between the observed and predicted number of households by area type.

In Table 3 'Obs' is observed, 'Pred' is predicted, 'Error' is the percentage error in the prediction (assuming the Census to be correct), 'Non-met' is non-metropolitan districts and PD is the population density in the district. So, for example in London the model predicts 35 zero-car households per 100 households (i.e. a cell value of 0.35), whereas 42 zero-car households are observed per 100 households (i.e. a cell value of 0.42).

Table 3: Validation of previous NATCOP predictions for 2011 by area type

		London	Metropolitan Districts	Non-met, PD > 10	Non-met, 2 < PD <=10	Non-met, PD<=2	Overall
P(0)	Obs	0.42	0.32	0.29	0.20	0.17	0.26
	Pred	0.35	0.28	0.25	0.18	0.16	0.23
	Error	-17%	-13%	-15%	-11%	-10%	-13%
P(1)	Obs	0.41	0.42	0.43	0.42	0.43	0.42
	Pred	0.45	0.47	0.48	0.47	0.48	0.47
	Error	11%	13%	11%	11%	12%	12%
P(2)	Obs	0.14	0.21	0.22	0.29	0.30	0.24
	Pred	0.16	0.21	0.22	0.28	0.29	0.24
	Error	17%	-3%	0%	-3%	-4%	-1%
P(3+)	Obs	0.04	0.05	0.06	0.09	0.10	0.07
	Pred	0.04	0.04	0.05	0.07	0.08	0.06
	Error	-1%	-15%	-11%	-17%	-23%	-17%
Total cars	Obs	0.82	1.02	1.07	1.30	1.37	1.16
	Pred	0.91	1.03	1.10	1.28	1.32	1.16
	Error	12%	2%	3%	-2%	-4%	0%

Looking first at the Overall column, which gives the total predictions for all of Great Britain, it can be seen that the model over-predicts the percentage of households owning one car by 12 per cent, and under-predicts the zero car and the multiple car ownership states (particularly the P_{3+} state). The net effect of the under-predictions of zero and multiple car ownership is that the total car ownership prediction (1.16) matches the observed value very closely.

The under-prediction of zero-car households is important in the context of making forecasts of public transport demand, because individuals in zero-car households are more likely to be public transport users than those in car-owning households.

Looking next at how the model performs between different area types, for area types other than London a similar pattern is observed, with total cars predicted reasonably well but a consistent pattern of over-prediction of one-car households and under-predictions of zero- and multiple-car-owning states. However, for London there is a significant (12 per cent) over-prediction of total car ownership, and in contrast to the other area types two-car households are over-predicted, which contributes to the overall over-prediction of car ownership.

To give more insight into the over-prediction of car ownership in London, the results have been broken down into Inner and Outer London in Table 4.

Table 4: Validation of previous NATCOP predictions for 2011 for London

		Inner London	Outer London	London
P(0)	Obs	0.57	0.31	0.42
	Pred	0.46	0.26	0.35
	Error	-19%	-15%	-17%
P(1)	Obs	0.35	0.44	0.41
	Pred	0.43	0.47	0.45
	Error	21%	6%	11%
P(2)	Obs	0.07	0.19	0.14
	Pred	0.09	0.22	0.16
	Error	41%	12%	17%
P(3+)	Obs	0.01	0.06	0.04
	Pred	0.02	0.05	0.04
	Error	45%	-8%	-1%
Total cars	Obs	0.53	1.02	0.82
	Pred	0.68	1.08	0.91
	Error	28%	6%	12%

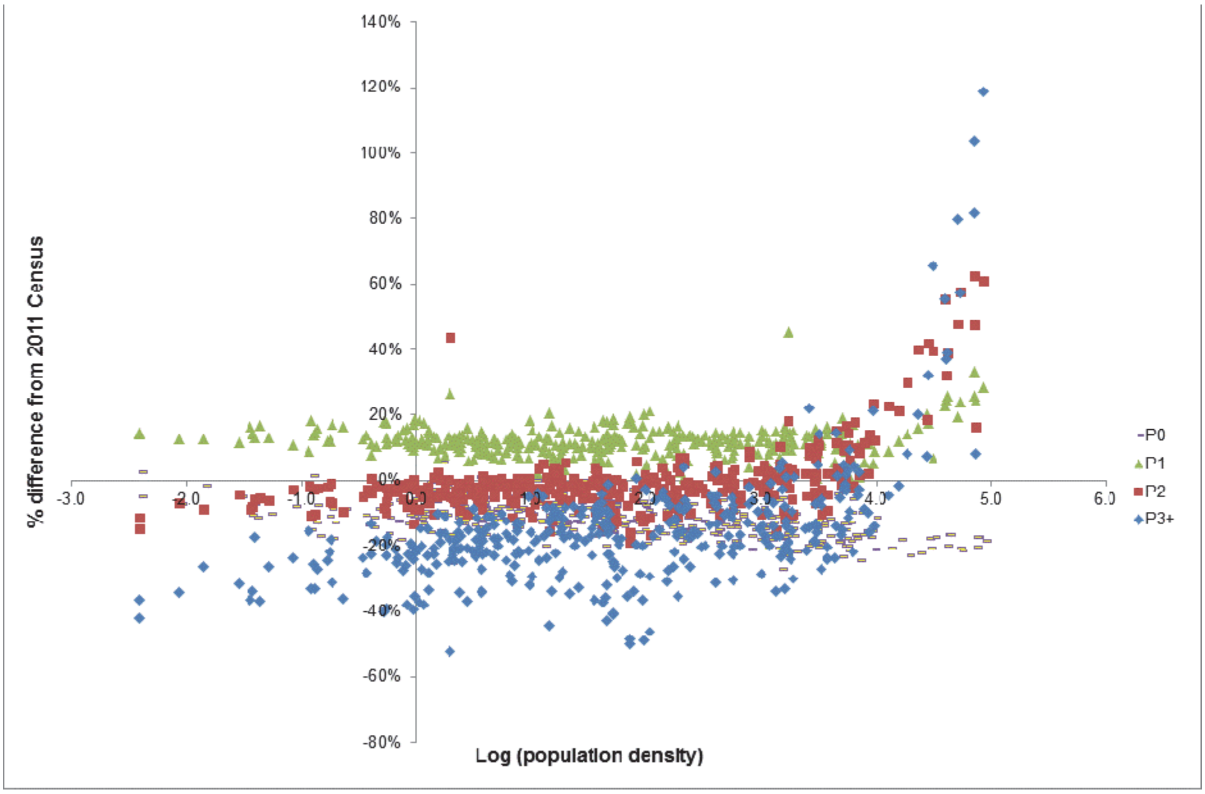
Table 4 highlights that the over-prediction of car ownership in London observed in Table 3 is largely due to an over-prediction of car ownership in Inner London. It can be seen from Table 4 that there is a particular problem of over-prediction of multiple-car-ownership households in Inner London.

As a result of this analysis separate area types for Inner and Outer London were tested in the new model. The findings from these tests are documented in Section 5.2.

4.1.2. Validation by population density

The brief for this work highlighted that NATCOP is known to over-predict car ownership in denser urban areas, particularly in London, and this was confirmed by the analysis presented in Section 4.1.1. Therefore validation of the 2011 NATCOP forecasts was undertaken by examining predicted and observed car ownership levels according to the 2011 population density of the district. The errors in the zero-, one-, two- and three-plus-car probabilities (P0, P1, P2 and P3+ respectively) have been plotted against population density in Figure 5. A logarithmic scale has been used for population density (measured as population per hectare) to account for the high densities observed in some urban areas.

Figure 5: Validation of 2011 NATCOP predictions by population density



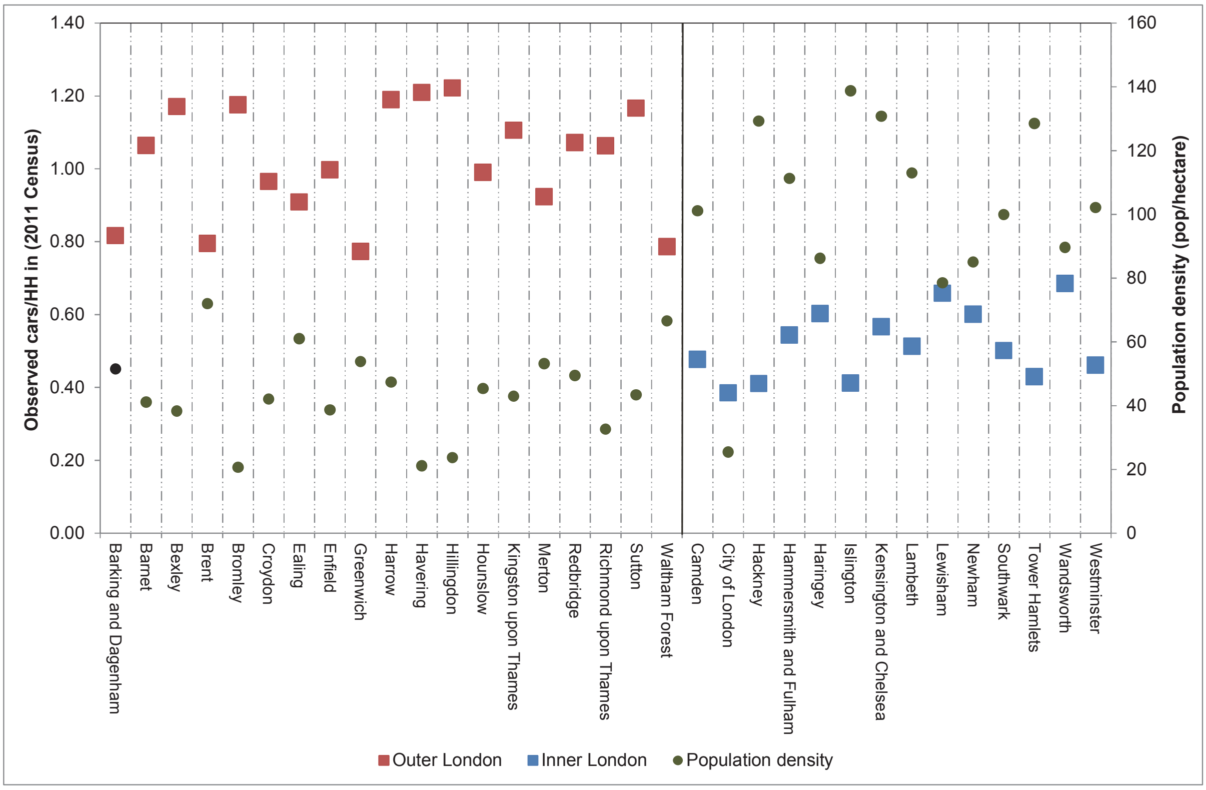
It can be seen that the over-prediction of one-car households highlighted in Table 3 persists across the whole range of observed population densities: the green triangles in Figure 5 show the one-car household predictions. Furthermore, there is no clear evidence of a change in the level of error as a function of population density except for the very highest population densities (over 4 on the log scale, equivalent to 55 persons per hectare). Zero-car households are under-predicted across the range of observed population densities (shown as yellow rectangles), with under-predictions ranging from around -10 per cent in the least dense areas to around -20 per cent in the densest areas.

By contrast, the errors in multiple car ownership (the red and blue squares in Figure 5) show a clear relationship with population density, moving from under-prediction at the lower population densities to over-prediction at the highest population densities. This is consistent with the pattern of over-prediction in multiple car ownership in Inner London highlighted in Table 4.

4.1.3. Interaction between area type and population density

To investigate the interaction between the London area types and population density the relationship between observed 2011 car ownership and population density was investigated by London borough. This analysis is presented in Figure 6, which plots the boroughs on the x-axis, the mean observed car ownership per household for the borough on the left-hand y-axis (as red squares for Outer London boroughs and as blue squares for Inner London boroughs), and the population density of the borough on the right-hand y-axis (as green dots).

Figure 6: Observed 2011 car ownership and population density by London borough



It can be seen from Figure 6 that car ownership is consistently lower in Inner London, even for boroughs with medium-high population density such as Haringey and Lewisham. Thus there seems to be an 'Inner London' effect that applies in addition to the population density effect. This is likely to reflect a combination of factors including higher congestion, constraints on parking supply, the impact of the congestion charge, and high levels of public transport accessibility.

Overall, the analysis clearly highlights a need to distinguish Inner and Outer London area types and to improve (reduce) the predictions of multiple car ownership in the densest areas in the Phase 2 re-estimation work. In addition, further investigations of terms of exogenous inputs were undertaken to explore whether these might account for differences between predicted and observed levels of zero car ownership. These are described in the following sections. The changes that have been made to the model specification to realise these improvements are discussed in Chapter 5.

4.2. Review of exogenous model inputs

4.2.1. Purchase and running costs

The NATCOP models incorporate purchase and running cost terms. The parameters for these two terms were constrained in the model estimation procedure so that the models replicated elasticity estimates from other research (Whelan et al., 2001). When MVA re-estimated the models in 2007 they constrained the purchase and running cost parameters to values that replicated the same elasticity estimates.

In this section we examine how purchase and running costs, which are inputs to the model, have changed over the 2001 to 2011 period in comparison to the 2011 values that were assumed when applying the previous (2001 base) NATCOP model.

Table 5 summarises the 1971 to 2000 historical purchase and running cost values assembled by MVA for the last re-estimation work (2007), as well as the values that were assembled by MVA for forecasting with the previous 2001 base model. These indices are expressed relative to base values of 100 for the 2001 base year.

Table 5: Purchase and running cost indices

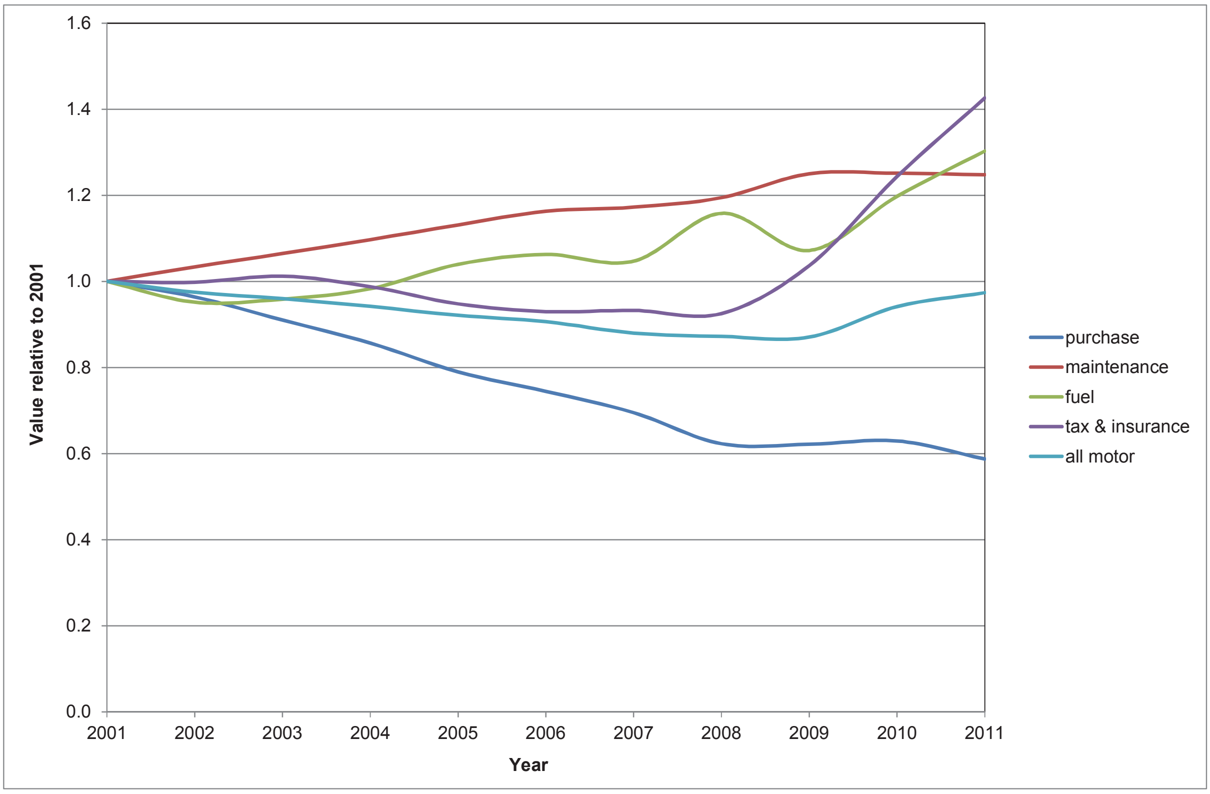
Year	Purchase cost	Running cost
1971	168	68
1976	161	74
1981	156	77
1986	138	72
1991	128	73
1996	126	83
1997	125	86
1998	119	88
1999	112	94
2000	103	102
2001	100.0	100.0
2006	74.4	100.0
2011	57.3	100.0
2016	46.9	100.0
2021	46.9	100.0
2026	46.9	100.0
2031	46.9	100.0
2036	46.9	100.0
2041	46.9	100.0

It can be seen that while significant reductions in purchase costs have been assumed in forecasting relative to the 2001 values, it has been assumed that running costs remain at 2001 levels for all forecast years.

For 2011, the actual observed value for the purchase cost index is 58.7, i.e. the observed reduction in purchase costs was forecast well. Therefore in the remainder of this section we have focused on analysing how running costs have changed over the period in comparison to the assumption made when applying the model that they remain at constant 2001 levels for all forecast years.

Changes in running cost come about as a result of changes of a number of different constituent components. Figure 7 plots the evolution of the various components of running cost between 2001 and 2011. The purchase cost index has also been plotted for comparison, and the 'all motor' series is all motoring costs (i.e. both purchase and running costs).

Figure 7: Observed changes in vehicle running and purchase cost indices, 2001–2011



Sources: Table TSGB0123, Retail Prices Index, Transport Components, UK Department for Transport.

It can be seen that maintenance, fuel and tax and insurance costs were all higher in 2011 than in 2001, showing increases ranging between 25 per cent and 43 per cent. Tax and insurance costs increased significantly over the period.

Overall it can be seen that that the assumption of constant running cost has under-estimated increases in car running costs that will have acted to dampen some of the growth in car ownership that would have occurred if running costs had remained constant. In the context of recent (2015–2016) declines in fuel prices, any predictions of future running costs will contain significant uncertainty. This issue is discussed further in *Software Developer's Note and QA* (D21).

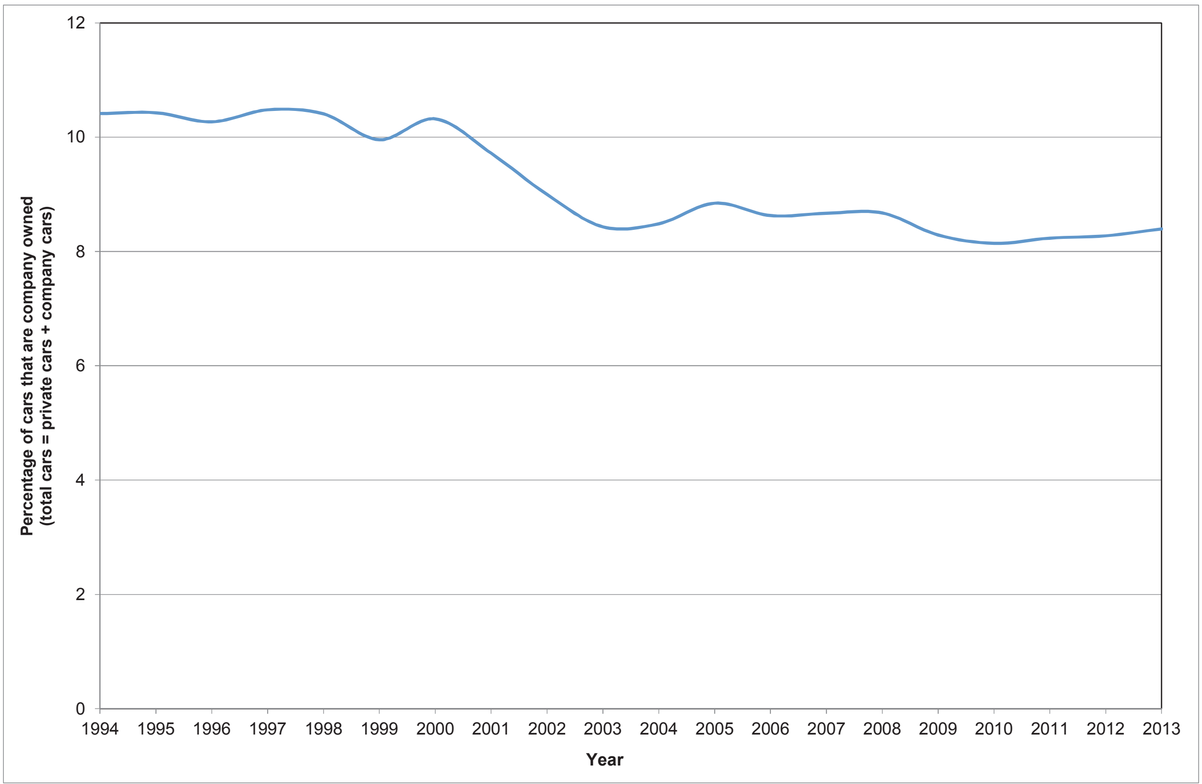
A limitation of the cost index information that has been used is that it does not represent variation in these costs between individuals of different ages, or between different areas. However, an issue is whether it is possible to first capture such information, and second forecast changes in that information over time. This is discussed further in Chapter 5.

4.2.2. Company car ownership

When ITS Leeds undertook the original development work on the disaggregate NATCOP models they investigated the impact of company car ownership on total car ownership (Whelan et al., 2001). Terms were included in the multiple car ownership models (P_2+, P_{3+}) to account for the higher probability of households owning multiple cars if they owned company cars. No term was included for the P_1+ model on the basis that households with these characteristics (i.e. above-average incomes) would be expected to own at least one car anyway.

When applying the previous 2001 base version of NATCOP, it was assumed that company car ownership remained fixed at 2001 levels. However, company car ownership levels actually declined noticeably after 2001, as shown by Figure 8.

Figure 8: Trends in company car ownership, 1994–2013



Source: DfT car statistics tables (accessed 18/03/15)
https://www.gov.uk/government/statistical-data-sets/veh02-licensed-cars

It can be seen from Figure 8 that company car ownership fell noticeably in the early 2000s from a figure of around 10 per cent of total cars to just over 8 per cent of total cars. This means that the assumed company car ownership level for 2011 will have been over-predicted. This over-prediction will contribute to the general pattern of over-prediction of multiple car ownership observed in Table 3.

It is also noteworthy that company car ownership levels and changes to these vary across the country. In 1995/97, the prevalence of company cars on a per-capita basis was 32 per cent greater in the South East than the rest of Great Britain, but by 2008/10 the situation had reversed and the figure was 6 per cent lower in the South East than elsewhere (Le Vine and Jones 2012; Rohr and Fox 2015). Le Vine and Jones (2012) conclude that the drop in company car activity by Londoners was sharp enough to be a major contributor to London's falling traffic levels in recent decades; this may also have had a substantial impact on car ownership levels.

The suggestion at the stakeholder event held at the Department on 13 March 2015 was that the fall in company car ownership was a structural change that occurred as a result of taxation policy, and that the change had now played out. That suggestion is consistent with the trend shown in Figure 8, which shows company car ownership levelling off at just over 8 per cent of total cars. The new models discussed in later chapters were re-estimated using a 2011 base year, and in application company car ownership levels observed in 2011 will be retained for future years. Thus the re-basing to 2011 ensures that the forecasts of the new model will not be impacted by the fall in company car ownership in the early 2000s.

4.3. Review of saturation and treatment of income

In Inner London, it is believed that the saturation levels currently represented in the model do not adequately represent the constraints on car ownership levels, specifically in denser urban areas and in particular Inner London.[6] Following the analysis presented in Section 4.1, in the new 2011 base version of NATCOP separate area types are used for Inner and Outer London, and where possible saturation rates have been estimated separately for those two areas, which allows the model specification to directly capture differences in saturation levels between Inner and Outer London. These results are discussed further in Section 5.2.

When ITS originally developed the NATCOP models they tested models without saturation effects (Whelan *et al.*, 2001). In these models, a logarithmic form for income gave the best fit to the data, and this specification has the effect that the marginal impact of increasing income on car ownership reduces as incomes increase, which is similar to imposing a saturation level. When saturation was directly incorporated into the model, it was found necessary to move to a linear specification for income in order to estimate saturation levels significantly different from one. However, the presence of an explicit saturation term retains the feature that the marginal impact of income reduces as incomes increase. This feature is consistent with the analysis by the NTM team (noted in the brief), which has found no sudden break between income and car ownership, but rather a long weakening of the relationship.

Overall, we are satisfied that the current model specification is sound in that it directly incorporates segmentation and gives the expected result that the marginal impact of income reduces as income increases. The representation of saturation in London has been enhanced in the new model by differentiating Inner and Outer London; in particular this change will reflect the much lower saturation rates in Inner London. Furthermore, tests have been undertaken to ensure that the saturation terms remain significant with the extended estimation dataset. These tests are reported in Section 5.2.

4.4. Impact of public transport accessibility

The brief for this work noted that the model may potentially be improved by providing information on public transport supply or parking space provision.[7] To investigate the impact of public transport supply,

[6] This comes from the brief for this work, specifically paragraph 5.6.1 of Appendix B, Specification, which states 'The model is known to over-forecast car ownership in dense urban areas, particularly London, and will require investigation into improving that capability.'

[7] Paragraph 5.6.1 of Appendix B, Specification states 'The model may potentially be improved by providing information on public transport supply or parking space provision that may impact on the decision to own and operate a vehicle, particularly in those denser areas.'

regressions were run to investigate the relationship between the number of cars per household and the walk time to the nearest bus and rail services. The regressions were estimated using 2002–2010 NTS data.

The regression that was estimated is detailed in Equation (4.1).

$$\begin{aligned} Cars/HH &= \beta_{Constant} + \beta_{BusWalk} BusWalk + \beta_{TrainWalk} TrainWalk \\ &+ \beta_{InnerLon} IF(InnerLon) + \beta_{OuterLon} IF(OuterLon) + \beta_{Year} Year \end{aligned} \tag{4.1}$$

where: *Cars/HH* is the number of cars per household

BusWalk is the walk time to the nearest bus service

TrainWalk is the walk time to the nearest train service

InnerLon is a constant applied if the household is resident in Inner London

OuterLon is a constant applied if the household is resident in Outer London

Year is a constant for the year (2002=1, 2010=9) to reflect the trend increase in car ownership.

The resulting parameter estimates are given in Table 6.

Table 6: Access to public transport regression results (rho-squared = 0.0324)

Parameter	Estimate	t-ratio
$\beta_{Constant}$	0.9955	110.0
$\beta_{BusWalk}$	0.0084	16.5
$\beta_{TrainWalk}$	0.0022	14.7
$\beta_{InnerLon}$	-0.5888	-40.8
$\beta_{OuterLon}$	-0.0910	-7.7
β_{Year}	0.0101	8.5

The regression results indicate significant effects whereby as walk time to the nearest public transport service increases (i.e. as access to public transport worsens) car ownership increases. These effects are significant after accounting for the lower levels of car ownership in Outer London, and the much lower levels of car ownership in Inner London. However, the low rho-squared value indicates that the overall ability of public transport accessibility to explain the observed variation in household car ownership is low.

It is noteworthy that the magnitude of the bus walk time parameter is 3.8 times that of the train walk time parameter, i.e. access to bus services gives a noticeably better explanation of car ownership at a household level than access to train services.

On the basis of these results, tests were carried out to investigate whether the NATCOP model specification would be enhanced by adding public transport accessibility into the model specification. These tests are documented in Section 5.5.

4.5. Consideration of adding parking space terms

The possibility of testing parking space provision at the home location in the model specification was briefly considered during the Phase 1 review. The conclusion was that the possibility of testing a term in the enhanced model specification should be considered, but this was caveated by the view that assembling a dataset that could be forecast into future years was likely to be a considerable challenge.

4.6. Improved treatment of licence holding

The brief for this work noted that:

> Recent behavioural trends in car ownership, particularly the decline in young males owning driving licences (and a relative increase in female drivers) are not captured in the current model's methodology. It should be considered how this may improve the forecasts and if it is warranted to be included in the model. This also suggests that it may be necessary to review the age segmentation within this model, and indeed the Scenario Generator.[8]

The current trends in licence holding are more complex, with reductions observed for younger adults (particularly men), and increases observed for older adults (particularly women). These trends are likely to play out differently in the different area types represented in NATCOP. Williams and Jin (2013) analysed 1981–2011 Census data and found that by 2011 the 25–44 age group were more strongly concentrated in high-density areas. If this trend were to continue alongside lower licence holding for younger persons this would impact upon car ownership in high-density areas.

An approach that has been successfully used in the Sydney Strategic Travel Model (STM) to account for effects of this type is to develop a cohort forecasting model (Tsang & Daly, 2010), and the PRISM West Midlands model also uses a simple cohort approach to reflect changes in aggregate licence holding (Fox et al., 2014). A cohort model could be developed in spreadsheet form using historical NTS data.

Therefore at the end of Phase 1 it was recommended that the cohort approach be adopted for the new version of NATCOP.

4.7. Summary of recommendations for model development

Table 6 summarises the findings from the review of model performance and specification, and outlines the recommendation made for updating and enhancing the model based on the review. In Table 6 'o/p' stands for 'over-predicts' and 'u/p' stands for 'under-predicts'.

[8] Paragraph 5.6.1 of Appendix B of brief.

Table 7: Summary of findings and recommendations for model development

Section		Findings	Recommendations
4.1	Validation by area type and population density	Previous models over-predict one-car households and under-predict zero-, two- and three-plus-car households, across all area types	Re-base models to 2011 to reflect observed 2011 shares of households by 0, 1 2, and 3+ cars
4.1.1	Validation by area type	Performance is worst in London; further analysis shows Inner London particularly problematic	Split the London area type into separate Inner and Outer London area types
4.1.2	Validation by population density	Validation by population density demonstrates previous models over-predict multi-car households in densest areas	Test continuous population density terms
4.1.3	Interaction between area type and density	Analysis of variation in car ownership by population density between Inner London and Outer London suggests area type effect in addition to density	Test both area type and population density terms
4.2	Review of exogenous model inputs		
4.2.1	Purchase and running costs	Purchase costs fell significantly between 2001 and 2011 but this fall predicted well in the 2011 inputs Running costs assumed to remain fixed between 2001 and 2011 whereas running costs increased significantly over the period, particularly tax, insurance and fuel costs	Review purchase and running cost assumptions post-2011 for forecasting with new model (2011 base), can draw on observed data for 2016
4.2.2	Company car ownership	Company ownership assumed to remain constant at 2001 levels when applying the previous model; however, car ownership fell significantly in the early 2000s	Re-basing the model to 2011 will ensure that drop in company car ownership is reflected in forecasts generated with the new model
4.3	Review of saturation and treatment of income	Review concluded previous treatment of saturation and income was sound; specifically it directly incorporates saturation and the marginal impact of income reduces as income increases	Retain previous treatment of income and saturation
4.4	Impact of PT accessibility	Increased access to PT correlates with lower car ownership	Test significance of this effect in estimation alongside the model parameters including area type and population density
4.5	Consider parking space terms	Consider testing parking space provision at the home location in model specification	Investigate whether a term can be tested noting that forecasting data of this type is likely to present considerable challenges
4.6	Improved treatment of licence holding	Treatment of licence holding would be improved by representing variation between age and gender cohort in the base year, and by providing a tool to forecast changes in licence holding by age band and gender in the future	Develop spreadsheet-based licence cohort model

5. New model development

This chapter describes the development of the new NATCOP model. Section 5.1 discusses choice data availability, as some of the variables in the new NATCOP model can only be specified from some of the choice data used for model estimation. Section 5.2 describes how the specification of saturation rates by area and household type was determined. Section 5.3 discusses how variation in car ownership behaviour across London and with population density is represented in the new model specification. Section 5.4 describes how running and purchasing cost coefficients have been constrained to match exogenous elasticity estimates.

5.1. Data availability

The choice data assembled for the model development work was described in Section 3.1. Most variables included in the models were defined for all years of data; however, the company car ownership terms in the previous version of NATCOP were only estimated from NTS data, and furthermore area type information was not available for the 1976 and 1981 FES data. The new variables added in this work following the Phase 1 review can only be defined from NTS data; these variables are discussed further in subsequent sections of this chapter.

Table 8 summarises which variables are available by year and data type. The three groups of variables that have been added to the model specification are shown at the bottom of the table.

Table 8: Model variables by dataset and year

Variable group	1971 FES	1976 and 1981 FES	1986–2000/01 FES	2000/01–2004/05 EFS	1999–2014 NTS
LPA (annual average)	√	√	√	√	√
Household income	√	√	√	√	√
Household type	√	√	√	√	√
Five original area types	√		√		√
Number of adults in household	√	√	√	√	√
Number of workers in household	√	√	√	√	√
Purchase and running cost indices	√	√	√	√	√
Company car ownership					√
Licences per adult by age and gender					√

Variable group	1971 FES	1976 and 1981 FES	1986–2000/01 FES	2000/01–2004/05 EFS	1999–2014 NTS
Inner and Outer London area types					√
Population density					√

The treatment of variables not available for all years of data is discussed further in the subsequent sections of this chapter.

5.2. Saturation terms

The saturation terms in the model vary by area and household type, but as per the previous versions of NATCOP there has been some aggregation in the final model specifications.

The starting point for the saturation tests was to estimate saturation terms separately for each combination of area and household type and then aggregate terms on the basis of those results.

Table 9 presents the full set of saturation terms estimated in the P_{1+} model for each area and household type combination. The saturation rates give an upper bound for the proportion of households owning one or more car for a given area and household type combination (please refer to Equation [2.1] for the mathematical formulation). As Inner and Outer London area types cannot be distinguished from the FES data a single London saturation term is estimated from this data.

Table 9: Full set of saturation rates by area and household type, P_{1+} model

Household type	London (FES data)	Inner London	Outer London	Metropolitan districts	Non-met dist >10 pers/ha	Non-met dist 2–10 pers/ha	Non-met dist <2 pers/ha
One adult, not retired	0.69	0.48	0.75	0.97	0.84	0.91	0.93
One adult, retired	0.51	0.52	0.75	0.53	0.77	0.81	0.79
One adult, with children	0.86	0.74	0.83	0.93	0.91	0.92	0.93
Two adults, retired	0.81	0.71	0.91	0.87	0.93	0.95	0.96
Two adults, no children	0.85	0.59	0.88	0.95	0.94	0.98	0.98
Two adults, with children	0.97	0.86	0.96	0.99	0.98	0.99	0.99
3+ adults, no children	0.90	0.64	0.94	0.99	0.96	0.98	0.99
3+ adults, with children	0.94	0.90	0.96	0.95	0.98	0.99	0.99

It can be seen that the saturation rates for Inner London area types are consistently lower than those for Outer London, and in turn the saturation rates for the four non-London area types are in all but one case higher than those for Inner London. Therefore the saturation rates have been merged into three area type groups:

- Inner London
- Outer London
- Non-London area types (metropolitan districts, non-metropolitan districts).

The variation in the saturation rates with household type is in line with expectations, with higher saturation rates in households with more adults and households with children. Based on the degree of difference between different household types the saturation rates have been merged into four groups:

- One adult, not retired
- One adult, retired
- One adult, with children and two adults, retired
- All two-adult and three-plus-adult household types.

The final aggregations are indicated by the coloured shading in Table 9.

Table 10 presents the saturation terms estimated before aggregation over area and household types for the P_{2+} model. Again, the aggregations used later in the final model specification are shown by the coloured shading.

Table 10: Full set of saturation rates by area and household type, P_{2+} model

Household type	London (FES data)	Inner London	Outer London	Metropolitan districts	Non-met dist >10 pers/ha	Non-met dist 2–10 pers/ha	Non-met dist <2 pers/ha
One adult, not retired	0.19	0.15	0.17	0.23	0.18	0.18	0.21
One adult, retired	0.14	0.54	0.13	0.13	0.16	0.15	0.16
One adult, with children	0.21	0.06	0.12	0.26	0.10	0.11	0.17
Two adults, retired	0.47	0.77	0.94	0.42	0.72	0.77	0.74
Two adults, no children	0.45	0.40	0.78	0.79	0.81	0.88	0.89
Two adults, with children	0.67	0.59	0.94	0.88	0.92	0.94	0.95
3+ adults, no children	0.65	0.67	0.92	0.83	0.89	0.94	0.95
3+ adults, with children	0.70	0.86	0.99	0.79	0.89	0.94	0.94

In general London saturation levels are lower than those for the other four area types; however, for households with a single adult and two retired adults the difference between Inner and Outer London saturation rates is not consistent. Therefore Inner and Outer London area types have been merged for calculation of saturation terms for these household types. For the final four household types the split into Inner London, Outer London and the rest has again been used.

The aggregation of household types varies from the P_{1+} model. In particular, as might be expected single-adult households have much lower saturation rates than multiple-adult households, and the presence of children is important in influencing the saturation rates in multiple-adult households. As saturation rates approach one the implication is that the fraction of the population that will never consider owning a car tends to zero.

The household type segmentation used for the saturation terms in the final model is:

- All one-adult household types
- Two adults, retired; two adults, no children
- Two adults, with children and all three-plus-adult household types.

For the P_{3+} model it was not possible to estimate a full set of saturation rates due to the lack of data. As illustrated in Figure 2, the fraction of households observed to own three-plus cars is just 5 per cent in 2011, and considerably lower in the older FES data. Therefore, a model was estimated where the saturation rates were aggregated over the eight household types allowing investigation for area type variation only. The results from this model are shown in Table 11.

Table 11: Saturation rates by area type, P_{3+} model

Household type	London (FES data)	Inner London	Outer London	Metropolitan districts	Non-met district>10 pers/ha	Non-met district 2–10 pers/ha	Non-met district <2 pers/ha
All household types	0.45	*1.00*	*1.00*	0.66	0.82	0.83	0.87

The estimated saturation rates for Inner and Outer London were not statistically significant, and the results effectively implied that the model could not estimate a saturation rate from the available data. A lower rate was estimated for metropolitan districts, but given the issues for the London area types it was decided to pool across all area and household types in the final model and estimate a single saturation rate. This is consistent with the treatment of saturation in the P_{3+} model by ITS Leeds in the original model development work (Whelan *et al.*, 2001) and MVA when they re-estimated the model (MVA, 2007).

5.3. Variation across London and with population density

The models reflect differences in observed behaviour between area types in three ways, first through variation in the estimated saturation rates $S_{1,ah}$, $S_{2,ah}$ and $S_{3,ah}$ in Equations (2.1) to (2.3), second through variation in the income sensitivity modifiers by area type c_{a1}, c_{a2} and c_{a3} in Equations (2.4) to (2.6), and in the new model specification through continuous population density terms.

5.3.1. Final saturation rates by area and household type

Table 12 presents the variation in the estimated saturation rates across the six area types used in the new P_{1+} model. The t-ratios given in brackets express the significance of the estimated parameter relative to a value of one.

Table 12: Final saturation rates by area and household type, P_{1+} model

Household type	Inner London		Outer London		Non-London	
One adult, not retired	0.46	(10.8)	0.74	(17.9)	0.90	(52.3)
One adult, retired	0.49	(4.9)	0.75	(10.3)	0.79	(29.4)
One adult with children, two adults retired, two adults no children	0.61	(18.6)	0.88	(34.0)	0.97	(120.5)
Two adults with children, 3+ adults no children, 3+ adults with children	0.78	(20.1)	0.95	(36.9)	0.99	(114.3)

Table 13 presents the variation in the estimated saturation rates across the six area types used in the new P_{2+} model. Again, the t-ratios express the significance of the estimated saturation rate relative to a value of one.

Table 13: Final saturation rates by area and household type, P_{2+} model

Household type	Inner London		Outer London		Metropolitan districts		Non-London	
One-adult households	0.16	(2.4)	0.16	(2.4)	0.22	(2.9)	0.17	(7.3)
Two adults retired, Two adults no children	0.49	(8.7)	0.82	(8.7)	0.76	(9.7)	0.87	(35.4)
Two adults with children, 3+ adults no children, 3+ adults with children	0.72	(7.5)	0.94	(9.7)	0.83	(12.8)	0.93	(46.4)

For the P_{3+} model, as discussed in Section 5.2 only a single saturation term was estimated because it was not possible to estimate differences by area and/or household type from the relatively small fraction of households observed to own three or more cars. This saturation term was 0.721 with a t-ratio of 15.8 relative to a value of one.

5.3.2. Variation in income sensitivities by area and household type

The income sensitives in the model vary by both area and household type. In both cases a base level is defined and then differences relative to the base level are estimated. Inner London is the base area type level in the new model, and for household type 1 (one adult) remains the base category.

The variation in the income modifiers by area type is summarised in Table 14. It is emphasised that for a given area type the modifiers express the *difference* between the sensitivity in that area type and the base level. It is also noted that the utility functions in the models are on the car-owning alternatives. Therefore a more positive income term implies a larger marginal impact of income on car ownership.

Table 14: Variation in income modifiers by area type

Model	Base: AT 1, HH 1 (one adult, not retired, Inner London)	AT 2: Outer London	AT 3: Metropolitan districts	AT 4: Non-met districts >10 pers/ha	AT 5: Non-met districts 2–10 pers/ha	AT 6: Non-met districts <2 pers/ha	London (FES data)
P_{1+}	0.101 (10.6)	*-0.002 (0.2)*	-0.025 (2.7)	-0.019 (2.0)	*0.004 (0.4)*	0.021 (2.1)	-0.016 (1.6)
P_{2+}	0.008 (2.6)	*0.000 (n/a)*	0.017 (8.1)	0.013 (9.1)	0.019 (13.2)	0.025 (7.5)	0.025 (7.5)
P_{3+}	*0.000 (n/a)*	*0.000 (n/a)*	*0.000 (n/a)*	*0.000 (n/a)*	0.002 (3.9)	0.004 (5.4)	*0.000 (n/a)*

Note: numbers that are shown in zero italics are parameters that are not significantly different from zero

Note that in the P_{3+} model the base level was not significantly different from zero. This means that to be plausible, any identified effects have to be positive to ensure that the marginal impact of income on car ownership is positive. For this reason, *negative* income modifiers for area types 2, 3 and 4 were fixed to zero; of these only the term for area type 3 was significantly different from zero (t=2.3) and so the impact of constraining these parameters on the overall model fit was modest.

For Outer London there is no significant difference in the income sensitivities relative to Inner London across all three models.

For metropolitan districts and higher-density non-metropolitan districts (>10 persons/hectare) there is a higher income sensitivity in the P_{1+} model but a lower income sensitivity in the P_{2+} model; again both effects are relative to Inner London.

For medium-population-density non-metropolitan districts (2–10 persons/hectare) significant income modifiers were identified in the multiple-car-ownership models (i.e. the P_{2+} and P_{3+}) relative to Inner London.

Finally, in low-population-density non-metropolitan districts (<2 persons/hectare) significant positive income modifiers, implying higher income sensitivities, were identified in all three models.

The variation in the income modifiers by household type is summarised in Table 15.

Table 15: Variation in income modifiers by household type

Model	Base: AT 1, HH 1 (one adult, not retired, Inner London)	HH 2: one adult, retired	HH 3: one adult with children	HH 4: two adults, retired	HH 5: two adults, no children	HH 6: two adults, with children	HH 7: three or more adults, no children	HH 8: three or more adults, with children
P_{1+}	0.101 (10.6)	*0.001 (0.1)*	-0.028 (12.7)	0.058 (23.4)	0.016 (7.7)	0.010 (5.2)	-0.018 (8.4)	-0.021 (9.5)
P_{2+}	0.008 (2.6)	*0.000 (n/a)*	*0.000 (n/a)*	0.008 (2.8)	0.010 (3.6)	0.006 (2.2)	0.023 (7.7)	0.013 (4.3)
P_{3+}	*0.000 (n/a)*	*0.000 (n/a)*	*0.000 (n/a)*	*0.000 (n/a)*	*0.000 (n/a)*	*0.000 (n/a)*	0.032 (25.3)	0.021 (21.2)

Note: numbers shown in italics are parameter estimates that are not significantly different from zero.

As per the discussion of Table 14, for the P_{3+} model the zero base value means that any income modifiers need to be significantly greater than zero for the income elasticities to be plausible.

There are no significant differences in the income sensitivities between HH 1 and HH 2, i.e. one-adult households without children.

For one-adult households with children the marginal impact of income is lower than for one-adult households without children, which is logical as households with children have a greater requirement for car ownership. No significant income modifier effects were identified for the P_{2+} and P_{3+} models, which is logical given that few single-adult households will own multiple cars.

For two-adult households (HH 4, HH5 and HH 6) significant positive income modifiers are observed for the P_{1+} and P_{2+} models implying a higher marginal impact of income for these household types relative to the one adult not retired households. No effect was identified for the P_{3+} model, which is consistent with the fact that a low fraction of two-adult households will choose to own three or more cars.

For three-adult households (HH 7 and HH 8) significant income effects were identified in all three models. For the P_{1+} model, the marginal impact of income is lower than for single person without children households; this is likely to reflect the fact that a high fraction of three-plus-adult households will own at least one car. However, positive income effects were identified in the P_{2+} and P_{3+} models, which means that the marginal impact of income is higher than the base level for these models.

5.4. Running and purchasing cost coefficients

Plausible coefficients for the running and purchasing costs could not be directly estimated from the year-specific car ownership data, where there is no cross-sectional variation because the indices are GB-wide and vary only with year. Therefore, consistent with the approach used in both the ITS and MVA estimations, the running and purchasing cost coefficients were constrained to generate the elasticity properties of the previous models (Whelan et al., 2001; MVA, 2007). The elasticities were constrained to the previous elasticities rather than to more recent values due to a lack of any more recent evidence on purchase and running cost elasticities (see for example Dunkerley et al., 2015).

The approach used to constrain the running and purchase cost parameters was taken from the 1999 NRTF work described in Whelan (1999), which was itself referenced in the original 2001 ITS NATCOP project. In the 1999 NRTF work, the elasticities which varied over time were derived from an underlying aggregate power growth model.

A simple model was used to estimate the purchase and running cost elasticities:

$$\varepsilon_t = \sum_m d_m cost_t \left(1 - \frac{P_{m,t}}{S_m}\right) W_{m,t} \tag{5.1}$$

where: ε_t is the estimated elasticity in year t

m is the sub-model (P_{1+}, P_{2+}, P_{3+})

d_m is the estimated cost parameter for sub-model m

$cost_t$ is the purchase or running cost index in year t

$P_{m,t}$ is the market share for sub-model m in year t

S_m is the saturation level for sub-model m

$W_{m,t}$ is the weight for m in year t.

The saturation level S_m does not vary with time; however, the weights $W_{m,t}$ vary as a function of year t. They are calculated as a function of the observed market shares. The full formulae used to make these calculations are detailed in Annex 1 of Whelan (1999).

Computationally, the starting point for the recalibration of the purchase and running cost coefficients for this work was the d_t values obtained in the 2001 ITS work (which developed 1991 base models) and the 2007 MVA work (which developed 2001 base models). However, the final reports from these studies did not contain any information on the saturation rates S_m that were assumed in order to derive their elasticity values.[9] Therefore, we adopted a two-stage approach to calibrate the running and purchase cost coefficients in the new model to be as consistent as possible with the ITS and MVA values and the underlying aggregate power growth dating back to the 1999 NRTF work on which the approach is based:

1. Take the running and purchase cost parameters d_t reported by MVA for their 2001 base model, and from these infer the global saturation rates $S_{m,t}$ that replicate the elasticity values reported for 1991 and 2001 in their study; and
2. Use the global saturation rates inferred from step 1 to derive the 2011 base parameters that replicate the elasticity estimates obtained by MVA in 2007.

The purchase and running cost parameters reported in the 2007 report for a 2001 base model are detailed in Table 16 (the decimal places used vary as per the source ITS and MVA reports).

[9] Note that these are global saturation rates per model, i.e. without the segmentation of saturation rates by area and household type used in the final model specifications.

Table 16: Purchase and running cost parameters for 2001 base model

Parameter d_t	P_{1+}	P_{2+}	P_{3+}
Purchase	-0.0125	-0.00408	-0.00095
Running	-0.006	-0.00196	-0.00046

Source: MVA (2007).

MVA calculated elasticities using these parameters for two points in time, 1991 and 2001; these are tabulated in Table 17.

Table 17: Running and purchasing cost elasticities from 2001 base elasticity model

	Purchasing cost	Running cost
1991	-0.34	-0.10
2001	-0.17	-0.08

Source: MVA (2007).

The 2001 values for running/purchase costs d_t (Table 16) were input into Equation (5.1) and the saturation rates S_m that best matched the observed running/purchase cost elasticities ε_t for 1991 and 2001 (Table 16) were calculated using a least squares approach. The inferred saturation rates are given in Table 18.

Table 18: Inferred saturation rates

	P_{1+}	P_{2+}	P_{3+}
S_m	0.812	0.652	0.400

These saturation rates were then used in Equation (5.1) together with purchase/running cost indices that use 2011 as the base year, and again least squares was used to infer the parameter values d_t that best match the 1991 and 2001 elasticity values quoted in Table 17. These values are detailed in Table 19, and differ from the 2001 base values calculated by MVA because they are specified to work with indices that use a 2011 base year.

Table 19: Purchase and running cost parameters for 2011 base model

Parameter d_t	P_{1+}	P_{2+}	P_{3+}
Purchase	0.00752	0.00084	0.00654
Running	0.00010	0.00054	0.01194

This allowed elasticity values for 2011 to be calculated, which are detailed in Table 20 as are the values obtained when applying the model to 1991 and 2001. The table demonstrates that the calibration has successfully identified 2011 base parameters able to reproduce the elasticity values obtained by MVA from the 2001 base model (Table 17).

Table 20: Running and purchasing cost elasticities from 2011 base elasticity model

	Purchasing cost	Running cost
1991	-0.34	-0.10
2001	-0.17	-0.08
2011	-0.08	-0.05

For both purchase and running cost elasticities it can be seen that the 2011 elasticity value shows a further reduction in elasticity relative to the 2001 value. Referring to Equation (5.1) it can be seen that the closer ownership levels get to saturation, the lower the elasticity value. Furthermore, Equation (5.1) illustrates that changes in the cost index will impact on the elasticity value. Purchase costs fell significantly between 2001 and 2011 whereas running costs increased, and this explains why the purchase cost elasticity has fallen by more in percentage terms than the running cost elasticity.

5.5. Public transport accessibility terms

The 1999–2014 NTS data records three public transport accessibility variables:

- Walk access time to the nearest bus service (minutes)
- Walk access time to the nearest train service (minutes)
- Bus access time to the nearest train service (minutes).

Analysis undertaken in Phase 1 of the project (presented in Section 4.4) indicated a relationship between the two walk access time variables and cars per household. Therefore the three public transport accessibility variables were tested alongside other variables in the model specification, including the saturation terms and other terms that vary with area type, that may capture some of the variation in public transport accessibility between households resident in different areas.

Table 21 summarises the impact of adding these three terms to the model in terms of increase in model fit, and the individual parameter estimates. Parameter estimates that are not significantly different from zero are shown in italics.

Table 21: Impact of incorporating PT accessibility terms in models

	P_{1+} model		P_{2+} model		P_{3+} model	
Observations		211,346		151,402		57,491
Gain in log-likelihood		165.7		143.3		35.3
Bus walk time parameter	0.0339	(9.5)	0.0233	(9.3)	0.0154	(5.8)
Rail walk time parameter	0.0022	(4.9)	0.0025	(6.8)	0.0019	(3.4)
Bus access to rail parameter	0.0041	(3.7)	0.0016	(1.7)	*1.10e-4*	*(0.1)*

The addition of the three public transport accessibility terms to the three models has resulted in statistically significant improvements in the fit to the data (measured by log-likelihood); however, the increases are relatively modest given the very large household sample sizes used to estimate the models.

The PT access parameters are all the expected sign, i.e. positive, indicating that higher car ownership is observed for households with higher access times to public transport services after correcting for other area type differences captured in the model specification.

It was decided not to take forward these terms for implementation, on the basis that the improvements in model fit are relatively modest and it would be difficult and time-consuming to make forecasts of how public transport accessibility will change in the future. However, the area type and population density terms will indirectly account for variation in public transport accessibility because public transport accessibility is positively correlated with population density.

5.6. Parking terms

While the NTS trip data collects parking information at the journey destination, the household data provided to us for the estimation work does not record information on parking costs and/or resident permits schemes. Even if such data were available, a further issue is that it would be time-consuming to assemble future-year forecasts of changes in these variables because it would require contacting individual local authorities.

For these reasons, explicit parking terms were not added to the models. However, the area type and population density effects will indirectly capture effects such as increased difficulties in parking in urban and denser areas, and in particular in Inner London.

5.7. Improved treatment of licence holding

One of the key improvements incorporated in the new NATCOP model is an enhanced treatment of licence holding. This work was documented in full in a separate deliverable, D11: *NATCOP Model Development Note*. Therefore this section focuses on the changes that have been made to the NATCOP models so that they can be fed by forecasts of licence holding by age and gender cohort from the new licence projection spreadsheet, as full documentation of the development of the licence cohort model was presented in D11.

In the previous car ownership model, the licence holding rates were incorporated by using an average LPA measure which was a GB-wide average value varying only by year. Therefore it only reflected the aggregate licence holding changes over time, not cross-sectional variation in licence holding between households.

The impact of licence holding changes has been enhanced relative to the current version of NATCOP by incorporating the cross-sectional variations of licence holding predicted by the new cohort model into the new NATCOP model specification.

To achieve this, LPA is included in the base NATCOP model estimation as two different terms: the individual LPA term by age–gender cohort and area type to reflect cross-sectional variation in licence holding, and the difference between the individual LPA terms and the annual average LPA to reflect longitudinal changes in licence holding. Therefore the variation of the licence holding by age-gender and different area type over the years has been reflected in the new car ownership model.

For the implementation, for each household, an average LPA is calculated by summing over age–gender cohorts the number of the adults multiplied by the projected LPA. For each year of NTS data:

$$LPA_h = \frac{\sum LR_{ci} \times N_{ci}}{\sum N_{ci}} \qquad (6.1)$$

where: h represents the household

LR_{ci} is the licence holding rates for the age-gender cohort ci

N_{ci} represents the number of adults for the cohort ci.

Therefore the future changes in licence holding rates will affect the average LPA calculated for the household, and so lead to changes in the predicted probabilities of the household owning a car.

A key point with the implementation of this approach is that the required disaggregate age-gender information is only available for the 1999–2014 NTS data. Thus for other data only the longitudinal LPA term is applied. For the 1999–2014 NTS data, an additional longitudinal term was tested to account for the mean contribution of the cross-sectional term. The a priori expectation is that this term will be negative as the cross-sectional term will capture some of the longitudinal effect.

The LPA parameters in the final model specifications are summarised in Table 22.

Table 22: LPA parameter estimates

Model	Cross-sectional LPA, 1999–2014 NTS only		Longitudinal LPA, all years of data		Longitudinal LPA, 1999–2014 NTS only	
P_{1+}	2.766	(44.7)	1.269	(10.9)	-0.109	(2.7)
P_{2+}	2.231	(23.8)	5.442	(40.4)	*0.000*	*(n/a)*
P_{3+}	*0.000*	*(n/a)*	2.784	(8.3)	*0.000*	*(n/a)*

Note: numbers shown in italics are parameter estimates that are not significantly different from zero.

Consistent with the previous NATCOP model, a significant longitudinal LPA term has been identified in all three models. The magnitude of the term demonstrates that licence holding has an important impact on car ownership across all three models, but that the effect is strongest for the P_{2+} model, followed by the P_{3+} model, i.e. the multiple-car-ownership models.

Significant cross-sectional effects have been identified in the P_{1+} and P_{2+} models and the magnitude of the effect in both models shows that cross-sectional variation in licence holding has an important effect on predicting car ownership. This result demonstrates the enhanced explanatory power introduced with the cross-sectional LPA terms. The improvements to the overall model fit are shown in Table 23.

Table 23: Impact of incorporating cross-sectional LPA terms on model fit

	P_{1+} model	P_{2+} model
Observations	211,346	151,402
Gain in log-likelihood	1,009.8	297.8

It can be seen that the addition of the cross-sectional LPA term results in a substantial improvement in the fit of the model to the observed car ownership choices for both the P_{1+} and P_{2+} models.

As noted above, the longitudinal LPA terms estimated from the 1999–2008 NTS data only account for any remaining time trend effect given that for the 1999–2008 NTS data both cross-sectional and time trend terms are applied. It can be seen from Table 22 that a relatively small term in magnitude has been identified for the P_{1+} model, and that no statistically significant term was identified for the P_{2+} model. For the $P3_{+}$ model only, no cross-sectional LPA term was identified, and so there is no reason to expect an additional longitudinal effect of the 1999–2008 NTS data.

5.8. Final model results

The model results incorporating the findings documented in Section 5.1 to 5.7 are summarised in Table 24 to Table 26. The following column headings used in these tables denote for each model coefficient:

- description: description of the model term
- form: whether the term is a constant or varies linearly with the variable
- label: the name of the coefficient label in the ALOGIT model estimation files
- used in application: whether the term is carried forward for implementation
- value: the coefficient value
- t-ratio: the t-ratio for the coefficient (coefficient value / standard error)

The P_{1+} model results are presented in Table 24, the P_{2+} model results in Table 25 and the P_{3+} model results in Table 26.

Table 24: Model results, P_{1+} model (v46)

Description	Form	Label	Used in application?	Value	t-ratio
alternative specific constant (all data)	constant	basc1	N	-0.7868	-11.1
alternative specific constant (1999–2001 NTS)	constant	basc2	N	-2.2373	-25.7
alternative specific constant (2002-2014 NTS)	constant	basc3	Y	-2.5917	-30.3
linear income (base)	linear	binc_b	Y	0.1014	10.6
linear income, HH type 2	linear	binc_hh2	Y	0.0005	0.1
linear income, HH type 3	linear	binc_hh3	Y	-0.0282	-12.7
linear income, HH type 4	linear	binc_hh4	Y	0.0577	23.4
linear income, HH type 5	linear	binc_hh5	Y	0.0158	7.7
linear income, HH type 6	linear	binc_hh6	Y	0.0101	5.2
linear income, HH type 7	linear	binc_hh7	Y	-0.0175	-8.4
linear income, HH type 8	linear	binc_hh8	Y	-0.0214	-9.5
linear income, area type 2	linear	binc_at2	Y	-0.0016	-0.2
linear income, area type 3	linear	binc_at3	Y	-0.0255	-2.7
linear income, area type 4	linear	binc_at4	Y	-0.0186	-2.0
linear income, area type 5	linear	binc_at5	Y	0.0038	0.4
linear income, area type 6	linear	binc_at6	Y	0.0206	2.1
linear income, London area type	linear	binc_FL	Y	-0.0162	-1.6
linear Income, area type missing	linear	binc_at0	Y	-0.0073	-0.8
number of workers in the household	linear	bemploy	Y	0.4041	29.5
purchase cost index	linear	bpur	Y	-0.0075	n/a
running cost index	linear	brun	Y	-0.0001	n/a
household-level LPA	linear	bLPA	Y	2.7656	44.7
GB-level LPA	linear	bLPA_T	Y	1.2592	10.9
GB-level LPA, term for years 1999 onwards	linear	bLPA_T2	Y	-0.1087	-2.7
population density (estimated from 2002–2014 data)	linear	bpopden	Y	-0.0075	-8.3
population density (dummy for missing years)	constant	bmpopden	N	-0.1907	-1.2

Table 25: Model results, P_{2+} model (v22)

Description	Form	Label	Used in application?	Value	t-ratio
alternative specific constant (all data)	constant	basc1	N	-5.3297	-31.7
alternative specific constant (1999–2001 NTS)	constant	basc2	N	-7.0651	-37.6
alternative specific constant (2002–2014 NTS)	constant	basc3	Y	-7.1377	-56.4
linear income (base)	linear	binc_b	Y	0.0077	2.6
linear income, HH type 2	linear	binc_hh2	Y	0.0000	n/a
linear income, HH type 3	linear	binc_hh3	Y	0.0000	n/a
linear income, HH type 4	linear	binc_hh4	Y	0.0083	2.8
linear income, HH type 5	linear	binc_hh5	Y	0.0103	3.6
linear income, HH type 6	linear	binc_hh6	Y	0.0061	2.2
linear income, HH type 7	linear	binc_hh7	Y	0.0227	7.7
linear income, HH type 8	linear	binc_hh8	Y	0.0125	4.3
linear income, area type 2	linear	binc_at2	Y	0.0000	n/a
linear income, area type 3	linear	binc_at3	Y	0.0174	8.1
linear income, area type 4	linear	binc_at4	Y	0.0125	9.1
linear income, area type 5	linear	binc_at5	Y	0.0190	13.2
linear income, area type 6	linear	binc_at6	Y	0.0227	15.2
linear income, London area type	linear	binc_FL	Y	0.0254	7.5
linear Income, area type missing	linear	binc_at0	Y	0.0271	16.5
one company car in the household	dummy	bcc1	Y	2.0266	23.2
number of workers in the household	linear	bemploy	Y	0.4030	35.1
purchase cost index	linear	bpur	Y	-0.0008	n/a
running cost index	linear	brun	Y	-0.0005	n/a
household-level LPA	linear	bLPA	Y	2.2306	23.8
GB-level LPA	linear	bLPA_T	Y	5.4421	40.4
GB-level LPA, term for years post 1998	linear	bLPA_T2	N	0.0000	n/a
population density (estimated from 2002–2014 data)	linear	bpopden	Y	-0.0059	-6.1
population density (dummy for missing years)	dummy	bmpopden	N	-0.2781	-2.0

Table 26: Model results, P_{3+} model (v22)

Description	Form	Label	Used in application?	Value	t-ratio
alternative specific constant (all data)	constant	basc1	N	-2.5928	-8.0
alternative specific constant (1999–2001 NTS)	constant	basc2	N	-3.1366	-9.3
alternative specific constant (2002–2014 NTS)	constant	basc3	Y	-2.9180	-11.9
linear income (base)	linear	binc_b	N	0.0000	n/a
linear income, HH type 2	linear	binc_hh2	n/a	0.0000	n/a
linear income, HH type 3	linear	binc_hh3	n/a	0.0000	n/a
linear income, HH type 4	linear	binc_hh4	n/a	0.0000	n/a
linear income, HH type 5	linear	binc_hh5	n/a	0.0000	n/a
linear income, HH type 6	linear	binc_hh6	n/a	0.0000	n/a
linear income, HH type 7	linear	binc_hh7	Y	0.0321	25.3
linear income, HH type 8	linear	binc_hh8	Y	0.0207	21.2
linear income, area type 2	linear	binc_at2	n/a	0.0000	n/a
linear income, area type 3	linear	binc_at3	n/a	0.0000	n/a
linear income, area type 4	linear	binc_at4	n/a	0.0000	n/a
linear income, area type 5	linear	binc_at5	Y	0.0022	3.9
linear income, area type 6	linear	binc_at6	Y	0.0037	5.5
linear income, London area type	linear	binc_FL	n/a	0.0000	n/a
linear Income, area type missing	linear	binc_at0	N	0.0013	2.7
one company car in the household	constant	bcc1	Y	0.3029	5.9
two or more company cars in the household	constant	bcc2	Y	1.3698	9.6
number of workers in the household	linear	bemploy	Y	0.3929	20.2
purchase cost index	linear	bpur	Y	-0.0065	n/a
running cost index	linear	brun	Y	-0.0119	n/a
household-level LPA	linear	bLPA	Y	0.0000	n/a
GB-level LPA	linear	bLPA_T	Y	2.7842	8.3
population density (estimated from 2002–2014 data)	linear	bpopden	Y	-0.0122	-7.2
population density (dummy for missing years)	constant	bmpopden	N	0.2179	0.9

6. Summary and recommendations

A full executive summary was presented at the start of this report, and furthermore a separate deliverable was provided at the end of Phase 1 that summarised recommendations for the Phase 2 model development phase. Therefore this section presents a summary of the outcome of the Phase 2 model development (Chapter 5) as well as discussing some recommendations for further work.

6.1. Summary of Phase 2 model development

The NATCOP models have been updated to reflect a 2011 base year and enhanced to improve their predictive ability. The updates and improvements have been made in the light of the Department's experience in the use of the previous model as detailed in the brief for this work and of the Phase 1 review of the performance of the previous model.

Estimation data

The dataset for model estimation retains the previous approach of combining FES, EFS and NTS data. More recent NTS data has been utilised for this work reflecting the data that has become available since the models were last updated in 2007.

The new variables that have been used to enhance the model specification during this work have all used the NTS data alone. This helps to illustrate the value of recent NTS data for transport modelling projects of this type, where understanding individual-level or household-level decisions is key.

Incorporating behavioural variation by area and household type

As per the previous versions of NATCOP the models incorporate an explicit representation of area type that varies by both area and household type. However, the area types have been enhanced to represent Inner London and Outer London separately; the other four area types for other metropolitan and non-metropolitan areas have been defined in the same way as in the previous version of the model.

Again consistent with earlier versions of the models, variation in income sensitivities by area and household type are explicitly represented in the new models.

Enhanced treatment of multiple ownership in high-density areas

A key finding from the Phase 1 review was that the models over-predicted car ownership in high-density areas, and in particular in Inner London. As noted above, the models have been enhanced to represent separate Inner and Outer London area types. The models have been further enhanced to incorporate a continuous population density term applied across all area types, i.e. including Inner London.

The impact of public transport accessibility and parking constraints

The brief for this work suggested that the impact of public transport and parking constraints on car ownership should be considered. It should be noted that the area type terms (present in both the previous and new models) and the population density terms (added as part of the current model enhancements) will capture a mixture of different effects including PT accessibility and constraints on parking, in particular through the estimation of significantly lower saturation effects in higher-density areas.

Tests of terms measuring households' accessibility to PT demonstrated that these yielded a significant improvement in the ability of the models to predict the car ownership choices observed in the 1999–2014 NTS data. However, it was judged that the considerable difficulty in forecasting changes to these variables in the future did not justify their retention in the final model specification.

It was not possible to identify a suitable variable to explicitly represent parking constraints at the home location from the 1999–2014 NTS data supplied for this work, and this combined with the difficulties in forecasting how these constraints might evolve over time meant that no parking constraint variable was tested as part of this work.

Improved treatment of licence holding

One of the key enhancements made to the car ownership model specification is the incorporation of forecasts of licence holding by age band and gender cohort. This enables the enhanced models to take account of cross-sectional variation in licence holding – which will evolve differently for different age-gender cohorts over time – in addition to the longitudinal licence holding term that has always been present in the NATCOP model specifications.

6.2. Recommendations for further work

This work was undertaken in response to a brief that set out a particular approach to the validation of the previous version of NATCOP in Phase 1. Specifically, the 2011 predictions of that model were compared to the car ownership levels observed in the 2011 Census. As one of the peer reviewers highlighted at the end of Phase 1, the most rigorous way to validate the predictive performance of the models for 2011 would be to replace all forecasts of input variables for 2011 with observed values. This approach would allow the analyst to fully separate the impact of errors in the input data from problems with the underlying model specification. While such a validation was beyond the scope of the current work it is worth considering for any future updates of the model.

Owing to a limited amount of recent evidence, the car ownership elasticities with respect to running and purchasing costs assumed in the current model are based on published values that date back to 2001. Given the general volatility in some of the running cost components and the importance of these to the model, we recommend that it would be valuable to undertake a more comprehensive literature review to identify more appropriate elasticity values for the new 2011 base year.

References

Anowar, S., N. Eluru and L. Miranda-Moreno. 2014. ‘Alternative Modeling Approaches Used for Examining Automobile Ownership: A Comprehensive Review.’ *Transport Reviews* 24(4): 441–73.

Daly. 1999. ‘How much is enough? Saturation effects using choice models.’ *Traffic Engineering and Control*.

De Jong, G., J. Fox, A. Daly, M. Pieters and R. Smit. 2004. ‘Comparison of Car Ownership Models.’ *Transport Reviews* 24(4): 379–408.

Department of Transport. 1978. *Report of the Advisory Committee on Trunk Road Assessment*. London: HMSO.

Dunkerley, F., C. Rohr and A. Daly. 2015. *A Rapid Evidence Assessment of Road Traffic Demand Elasticities in the UK*. Cambridge: RAND Europe.

Fox, J., S. Patil, B. Patruni and A. Daly. 2014. *PRISM 2011 Base: Frequency and Car Ownership Models*. Cambridge: RAND Europe.

Le Vine, S. and P. Jones. 2011. *On the Move: Making Sense of Car and Train Travel Trends in Britain*. London: RAC Foundation.

MVA. 2007. *Continuous Improvement: Updating National Car Ownership Model*. Report for Department for Transport.

Rohr, C. and J. Fox. 2014. *Evidence Review of Car Traffic Levels in Britain: a rapid evidence assessment*. Cambridge: RAND Europe

Tsang, F. and A. Daly. 2010. *Forecasting Car Ownership in the Sydney Area*. Presented to European Transport Conference, Glasgow.

Williams, I. and Y. Jin (2013) *The Impacts of Urban Densification of Transport*, slide presentation, European Transport Conference, Frankfurt.

Whelan, G. 1999. *A Recalibration of the NRTF Car Ownership Models*. Report to the Department of the Environment, Transport and the Regions.

Whelan, G., K. Fox and A. Daly. 2001. *Updated Car Ownership Forecasts*. Final report to the Department of the Environment, Transport and the Regions.

Whelan, G. 2001. *Methodological Advances in Modelling and Forecasting Car Ownership in Great Britain*. European Transport Conference, Cambridge.

Whelan, G. 2007. 'Modelling Car Ownership in Great Britain.' *Transportation Research Part A: Policy and Practice* 41(3): 205–219.

Appendix A – Saturation estimation methodology

The approach used in the models to allow direct estimation of saturation effects works as follows (Daly, 1999). If each choice *a* has an attractiveness function U_a, as defined in Equations (3.1) to (3.3), then an artificial alternative is set up with attractiveness function U_{ab}:

$$V_{ab} = V_a + \log \theta_b \tag{A.1}$$

where: θ_b is positive.

Then, *n* composite alternatives are defined, each being a nest containing the original alternative *b* and the *n* artificial alternatives with the same constant θ_b. The composite utility of nest b^* is then given by:

$$\exp V_{b*} = \exp V_b + \sum_a \exp V_{ab} \tag{A.2}$$

$$\exp V_{b*} = \exp V_b + \theta_b \sum_a \exp V_a \tag{A.3}$$

The choice probability for the nest b^* is given by:

$$p_{b*} = \frac{\exp V_b + \theta_b \sum_a \exp V_a}{\sum_c \left(\exp V_c + \theta_c \sum_a \exp V_a \right)} \tag{A.4}$$

$$p_{b*} = \frac{\exp V_b + \theta_b \sum_a \exp V_a}{\left(1 + \sum_c \theta_c\right) \sum_a \exp V_a} \tag{A.5}$$

The minimum fraction choosing alternative b^* is when $V_b \to -\infty$:

$$\min\left[p_{b*}\right] = \frac{\theta_b}{1 + \sum_c \theta_c} \tag{A.6}$$

The maximum fraction choosing alternative b^* is when $V_b \to \infty$:

$$\max\left[p_{b*}\right] = \frac{1 + \theta_b}{1 + \sum_c \theta_c} \tag{A.7}$$

This means that the parameters θ_b define the fraction of the population that is captive to that alternative, which gives the minimum choice fraction. The maximum choice fraction is given by the captive fraction plus the choices made by the rest of the population that is not captive to other alternatives.

In the NATCOP context, a series of binary choices are modelled and in each case only one of the two alternatives has a minimum choice fraction.

Appendix B – Quality Assurance

RAND Europe QA has been used on this project at the following stages:

- Agreement of a QA plan at the outset of the project, where the two independent QA reviewers were nominated, and a risk table drawn up summarising risks and mitigation measures;
- Periodic discussions with the continuous reviewer to ensure that that project is on track; and
- Double review of all final outputs, including this deliverable.

Each RAND Europe report deliverable (including this document) has been scored against RAND's quality standards, which are detailed in Figure 9.

Figure 9: RAND's quality standards

1: The problem should be well formulated and the purpose of the study should be clear.

2: The study approach should be well designed and executed.

3: The study should demonstrate understanding of related studies.

4: The data and information should be the best available.

5: Assumptions should be explicit and justified.

6: The findings should be important, advance knowledge and bear on important policy issues.

7: The implications and recommendations should be logical, warranted by the findings, and explained thoroughly, with appropriate caveats.

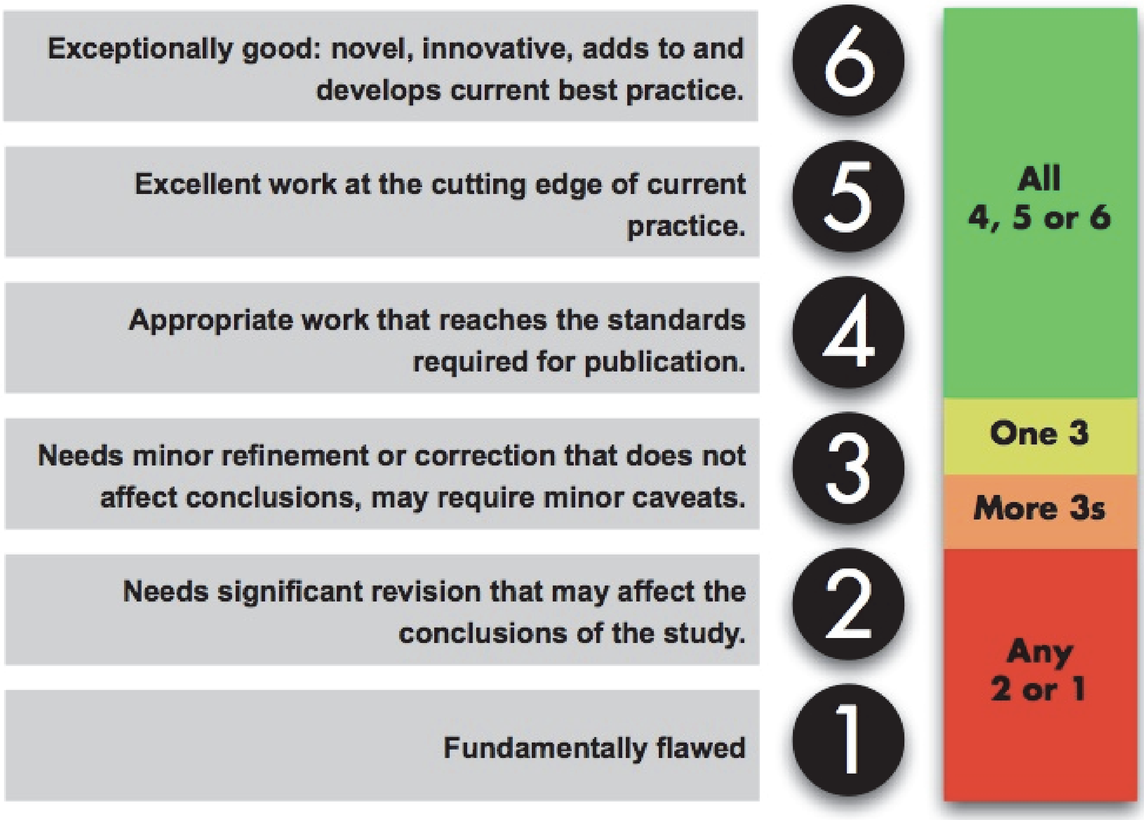
8: The documentation should be accurate, understandable, clearly structured and temperate in tone.

9: The study should be compelling, useful, and relevant to stakeholders and other decisionmakers.

10. The study should be objective, independent, and balanced.

To ensure each of these ten quality standards are assessed, and an appropriate level of quality is met, RAND Europe reports are scored on a numerical scale from 1 to 6. Only when the reviewer(s) are satisfied that the report has met the minimum standards for publication (4 or higher in all categories) can it be released. The scoring ladder that defines the interpretation of each numerical score is given in Figure 10.

Figure 10: RAND Europe's quality scoring system



Appendix C – Licence cohort model

RAND EUROPE

MEMO

To : Pawel Kucharski

From : James Fox, Bhanu Patruni, Andrew Daly, Hui Lu

Subject : Licence cohort model

Date : 22 January 2016, updated 21 April 2016 and 13 October 2016

Reference : PR-2285-DfT

1 Introduction

This note documents a licence cohort model that has been developed to forecast changes in (car driver) licence holding by area type, age band and gender over time. These forecasts will be used as inputs to the new national car ownership model (NATCOP) that better takes account of future changes in licence holding.

Section 2 summarises the data that has been used to develop the model and presents analysis showing how licence holding has evolved historically. It also presents analysis of licence holding for 2011 (the base year for the revised NATCOP model), including analysis of variation between area types.

Section 3 describes how the cohort model operates, presenting equations showing how acquisition and loss rates are used to predict licence holding for a given cohort as a function of that cohort's licence holding in the previous period.

Section 4 documents the calibration of the model, which involves calculating acquisition and loss rates from historical licence holding data, and then making adjustments to those rates to smooth out variations and ensure that the cohort models give plausible forecasts.

Section 5 documents validation and sense checking of the models. The models have been validated by making a short projection from 2011 to 2014 and those projections have been compared to observed NTS data. Additionally a longer term forecast to 2041 has been made and checked to ensure that it is plausible.

Section 6 explains how the new cohort model will be integrated with the new national car ownership model to better account for future changes in licence holding.

Finally, Section 7 summarises the work and some caveats associated with the projections given by the model.

2 Data

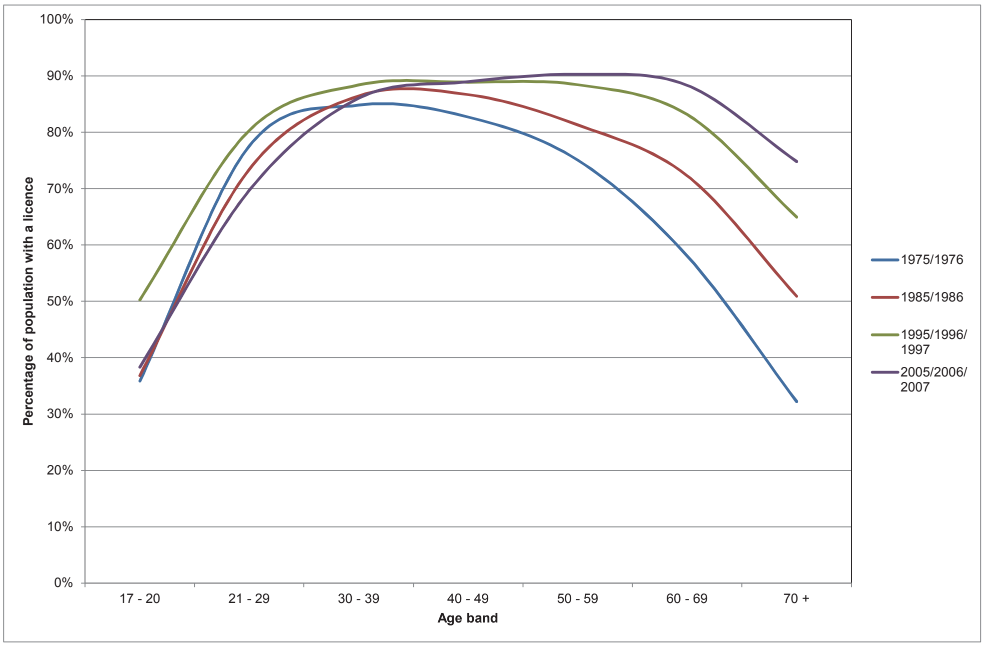
Two data sources have been assembled for the licence cohort analysis:

1. historical NTS data on licence holding analysis available from published tables, providing licence holding in 1975/1976, 1985/1986 and 1989/1991[1]; and
2. more recent NTS continuous survey data, supplied by the Department to enable the development of the new car ownership models, that can be analysed to give licence holding rates for the period covering 1995 to 2014.

2.1 Changes in licence holding over time

This data has been analysed to examine how licence holding has evolved over the past 40 years. Figure 1 plots male licence holding by age band at approximately ten year intervals from 1975/1976 to 2005/2006/2007. For each age band you can see how licence holding has changed across years (coloured lines). You can also see the profile of licence holding across age bands for any one year and trends in these patterns.

Figure 1: Historical male licence holding by age band



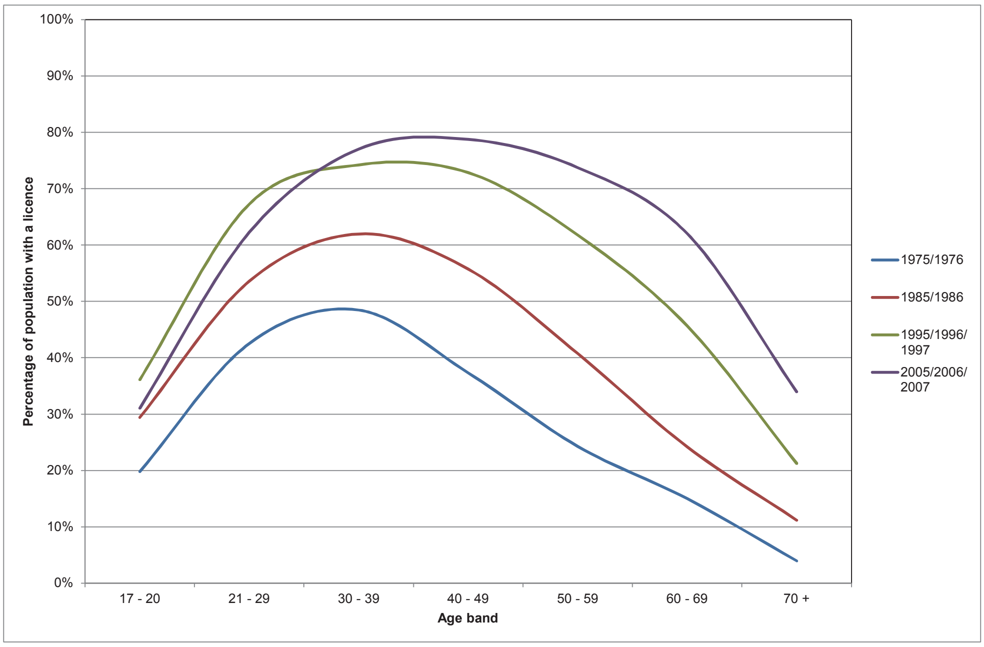
The largest changes in licence holding over time have occurred for persons aged 50 and above, where licence holding has increased over time as individuals have retained

[1] https://www.gov.uk/government/statistical-data-sets/nts02-driving-licence-holders, accessed 22/12/15. Please note that the historical data was presented in a summary table of the licence holding rate by gender at a 10 year age intervals.

their licences into older age. However, some changes can also be observed for young adults, specifically licence holding in the 17–20 and 21–29 age bands is lower in 2005/2006/2007 than in 1995/1996/1997.

Figure 2 plots observed licence holding for females over the same years.

Figure 2: Historical female licence holding by age band



Female licence holding has changed more substantially than male licence holding over the 30 year span covered by this analysis, in particular substantial increases in licence holding for working age females have been observed as well as substantial increases for older females, whereas for working age males the key changes have been for older people.

Consistent with the analysis for males, licence holding for the youngest two age bands is lower in the most recent 2005/2006/2007 data compared to the previous 1995/1996/1997 data. This trend for younger people to delay licence acquisition is also observed in 2010/2011/2012 data,[2] as illustrated in Table 1.

[2] This most recent data was not plotted in Figure 1 and Figure 2 because only data at 10-year intervals has been plotted.

Table 1: Observed licence holding percentage for younger cohorts

		1995/1996/1997	2000/2001/2006	2005/2006/2007	2010/2011/2012
males	17–19	50 %	36 %	38 %	35 %
	20–24	78 %	66 %	64 %	61 %
females	17–19	36 %	30 %	31 %	32 %
	20–24	65 %	57 %	54 %	55 %

It can be seen that male licence holding in these age bands has reduced more than female licence holding, though from higher initial levels. This may be because some of the factors that are believed to contribute to these changes, such as increases in insurance costs, have impacted more on males.[3] Furthermore, female licence holding changes only slightly between 2005/2006/2007 and 2010/2011/2012. This trend for delayed licence acquisition is discussed further in Section 4.

This pattern of changes is in line with other evidence. For example, Kuhnimhof *et al.* (2012)[4] analysed changes in driving licence information in Germany, France, Great Britain, Norway, the US and Japan. They observed that the share of young licenced drivers[5] had decreased noticeably in four of these countries, whereas for France and Germany no significant change was observed. Their analysis also demonstrated greater falls in licence holding for males than females, consistent with the analysis of NTS data presented in Table 1.

2.2 Base year licence holding

Separate cohort models have been developed for the six area types to be used in the new version of NATCOP, specifically:

1. Inner London
2. Outer London
3. Metropolitan Districts
4. Non-Metropolitan Districts, population density >10 persons/ha
5. Non-Metropolitan Districts, population density 2–10 persons/ha
6. Non-Metropolitan Districts, population density <2 persons/ha

---

[3] At least historically, going forward insurance cost for young males and females are likely to be equally high.

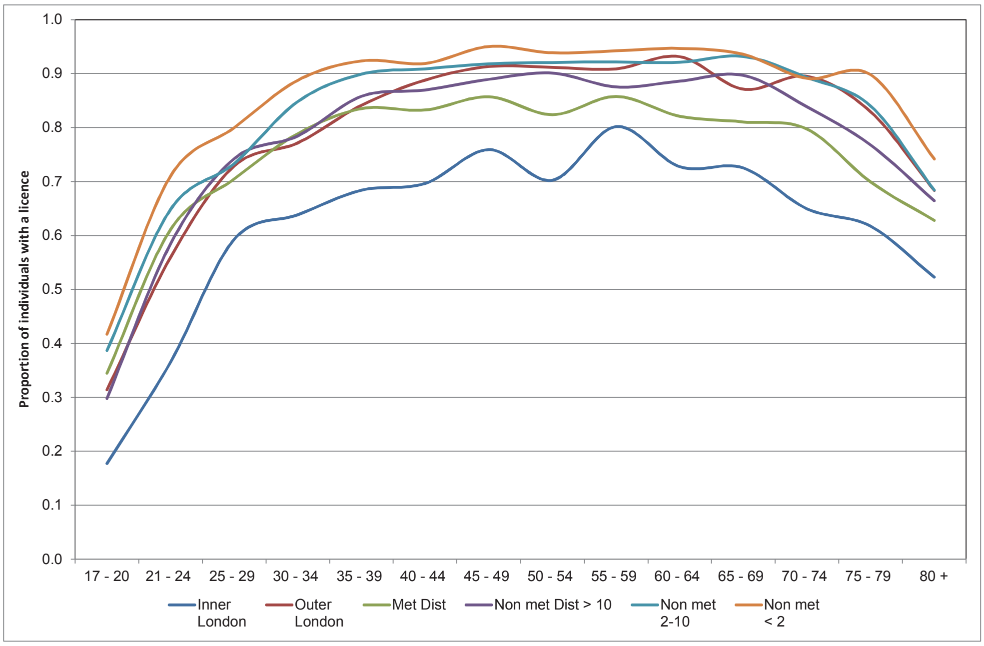
[4] Kuhnimhof, T., J. Armoogum, R. Buehler, J. Dargay, J. Denstadli and T. Yamamoto (2012) Men Shape a Downward Trend in Car Use among Young Adults—Evidence from Six Industrialized Countries, *Transport Reviews*, 32(6), 761–779.

[5] Their definition of 'young' varied between countries, depending on the age at which you can acquire a licence. The upper end was 29 except for Norway, where 34 was used because the available age band was 25–34.

This ensures that area type variations in licence holding impact upon the predictions of car ownership, which also incorporate area type variation.

Analysis has been undertaken to examine variation in licence holding across these area types, split by age band and gender for the base year (2011). To ensure sufficiently large samples the 2011 values were calculated as a five year average of 2009, 2010, 2011, 2012 and 2013 NTS data. The data is presented in Figure 3 and Figure 4. In these figures the 'Non-Metropolitan District, population density >10 persons/ha' area type is abbreviated 'Non Met > 10', similarly for the 2–10 persons/HA and < 2 persons/HA Non-Metropolitan District area types.

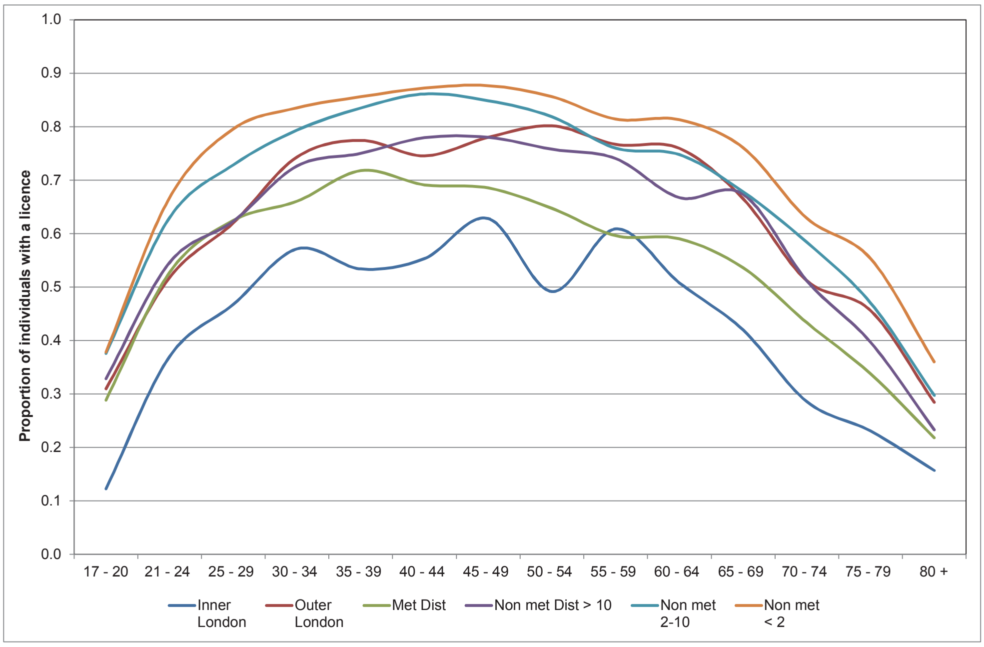
Figure 3: 2011 male licence holding by age band and area type



The patterns of variation by area type are in line with expectations, with the lowest levels of licence holding in Inner London and the highest levels in the lowest density non-metropolitan districts. It is interesting to note that licence holding in Outer London is higher than licence holding in other metropolitan districts; this may be an income effect, and additionally the metropolitan districts will include population in their inner areas which are likely to have lower average licence holding.

There is some volatility in the licence holding rates for the 45 to 64 age bands due to small sample sizes, this issue is discussed further below.

Figure 4: 2011 female licence holding by age band and area type



The patterns of variation of licence holding for females between area types are similar to those observed for males, but the variation in licence holding within area types is greater for females than it is for males.

As per the male rates there is some volatility in the licence holding rates for the 45 to 64 age bands.

To address the volatility issue, some smoothing was undertaken using the following adjustments:

- for the 45–49 and 50–54 age bands, an overall average rate across the 45–54 age band was calculated
- for the 55–59 age band, an overall average rate across the 50–54, 55–59 and 60–64 age bands was calculated
- for the 60–64 age band, an overall average rate across the 55–59, 60–64 and 65–69 age bands was calculated

The smoothed licence holding rates are plotted in Figure 5 and

Figure 6.

Figure 5: Smoothed 2011 male licence holding by age band and area type

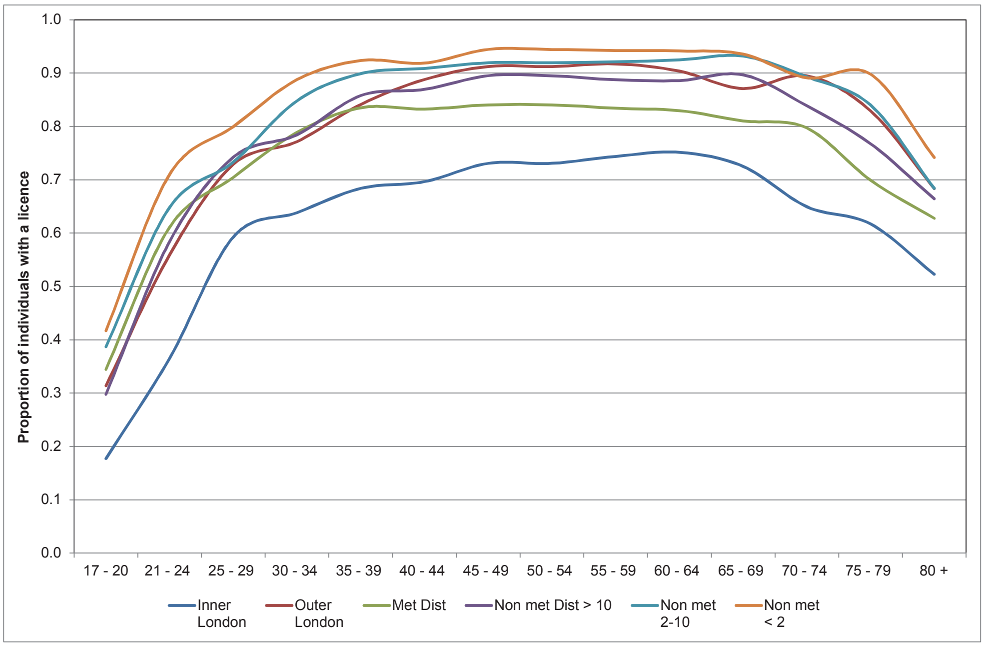
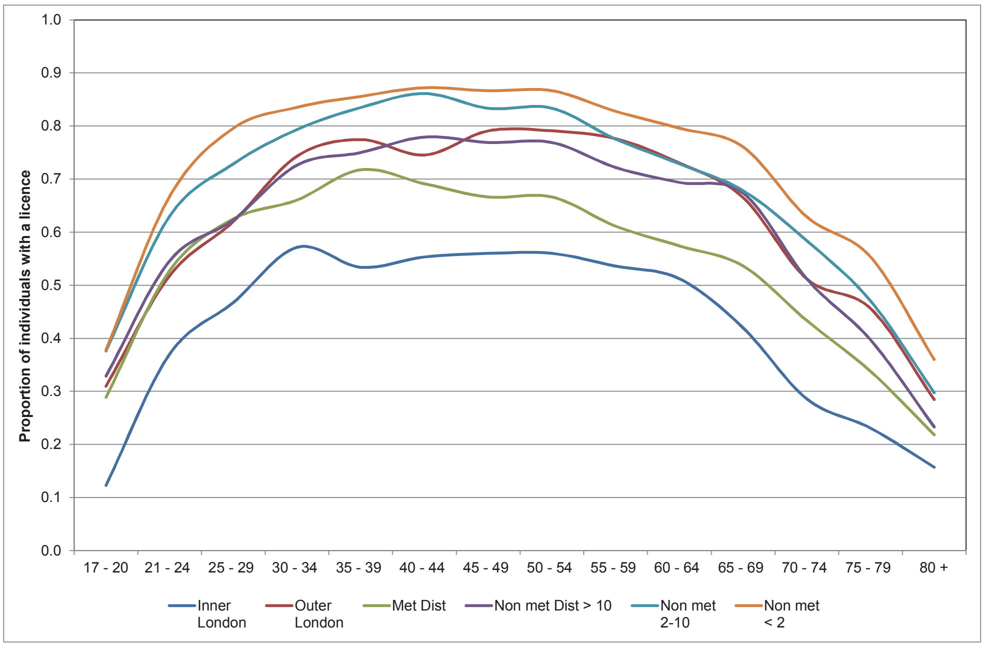


Figure 6: Smoothed 2011 female licence holding by age band and area type



The licence holding rates from the smoothed curves are summarised in Table 2 and Table 3.

Table 2: Smoothed 2011 male licence holding rates by age band and area type

	Area type						
	Inner London	Outer London	Met Dist	Non met Dist > 10	Non met 2-10	Non met < 2	National
17 - 20	0.177	0.314	0.344	0.298	0.387	0.417	0.355
21 - 24	0.365	0.559	0.611	0.584	0.648	0.710	0.605
25 - 29	0.593	0.729	0.705	0.742	0.733	0.800	0.714
30 - 34	0.638	0.772	0.789	0.784	0.848	0.888	0.799
35 - 39	0.684	0.841	0.835	0.858	0.899	0.923	0.857
40 - 44	0.696	0.888	0.832	0.869	0.909	0.919	0.873
45 - 49	0.731	0.912	0.840	0.895	0.919	0.944	0.895
50 - 54	0.731	0.912	0.840	0.895	0.919	0.944	0.895
55 - 59	0.761	0.927	0.841	0.881	0.919	0.945	0.898
60 - 64	0.768	0.913	0.837	0.880	0.923	0.944	0.900
65 - 69	0.725	0.871	0.810	0.897	0.932	0.936	0.891
70 - 74	0.649	0.895	0.797	0.838	0.893	0.891	0.854
75 - 79	0.617	0.829	0.699	0.768	0.840	0.898	0.810
80 +	0.523	0.683	0.628	0.664	0.684	0.742	0.689

Table 3: Smoothed 2011 female licence holding rates by age band and area type

	Area type						
	Inner London	Outer London	Met Dist	Non met Dist > 10	Non met 2-10	Non met < 2	National
17 - 20	0.123	0.310	0.288	0.329	0.376	0.378	0.322
21 - 24	0.370	0.518	0.527	0.546	0.630	0.667	0.554
25 - 29	0.468	0.620	0.625	0.623	0.729	0.796	0.654
30 - 34	0.571	0.743	0.661	0.726	0.793	0.835	0.732
35 - 39	0.534	0.774	0.718	0.750	0.835	0.856	0.778
40 - 44	0.553	0.745	0.691	0.779	0.861	0.872	0.791
45 - 49	0.560	0.791	0.666	0.769	0.833	0.866	0.783
50 - 54	0.560	0.791	0.666	0.769	0.833	0.866	0.783
55 - 59	0.536	0.776	0.611	0.722	0.775	0.828	0.742
60 - 64	0.512	0.730	0.574	0.694	0.729	0.796	0.709
65 - 69	0.420	0.665	0.536	0.674	0.678	0.761	0.666
70 - 74	0.286	0.512	0.433	0.512	0.584	0.629	0.546
75 - 79	0.231	0.456	0.338	0.397	0.470	0.553	0.451
80 +	0.157	0.284	0.218	0.233	0.297	0.360	0.282

These smoothed licence holding rates have been used as the base year rates when predicting future licence holding rates by area type.

3 Cohort model formulation

The cohort model follows an approach that has been successfully used to predict licence holding in a number of large-scale transport models, for example in the Sydney Strategic Model since 2000[6] and the Dutch National Model since prior to 2000. The basic model assumes that licence holding for a cohort (defined by birth date) is equal to the licence holding for the same cohort in the previous time period plus net acquisitions that have occurred over the period, and that licence holding can never exceed a saturation level because for various reasons a fraction of the population will never acquire licences.

The key formula is:

$$P_{c,t} = P_{c,t-1} + A_c(S - P_{c,t-1}) \tag{3.1}$$

$$0 \le A_c \le 1$$

where: $P_{c,t}$ is the licence holding fraction for cohort c at time t

$P_{c,t-1}$ is the licence holding fraction for the same cohort c at time $t-1$

A_c is the net acquisition rate for cohort c, assumed fixed over time

S is the saturation level of the licence holding

The model is applied over fixed time intervals of five years, consistent with the approach used in the Sydney application.

Two modifications are made to Equation (3.1) to apply the model to younger and older adults respectively:

1. for younger adults (17-24), the formula is applied based on the holding of the *previous* cohort at time *t-1*; and
2. for older people (over 60), losses rather than acquisitions are observed and so the formula is modified so that the change is calculated based on the number of people who currently have licences, rather than those who might still acquire them.

The full set of equations used in the model are detailed in Equations (3.2) to (3.4).

Young adults (17–24):

$$P_{c,t} = P_{c-1,t-1} + A_c\left(S - P_{c-1,t-1}\right) \tag{3.2}$$

[6] Daly, A. and F. Tsang (2010) Forecasting Car Ownership in the Sydney Area, presented at European Transport Conference, Glasgow. http://abstracts.aetransport.org/paper/index/id/3334/confid/16.

Main working age adults (25–59):

$$P_{c,t} = P_{c,t-1} + A_c\left(S - P_{c,t-1}\right) \tag{3.3}$$

Older adults (60-plus):

$$P_{c,t} = P_{c,t-1}\left(1 + L_c\right) \tag{3.4}$$

where: L_c is the net rate of gain of licences (i.e. L_c is expected to be negative)

The acquisition and loss rates are calculated from historical changes in licence holding, as detailed in Equations (3.5) to (3.7).

Young adults (17–24):

$$A_c = \frac{k}{n} * \frac{\left(P_{c,t} - P_{c-1,t-1}\right)}{\left(S - P_{c-1,t-1}\right)} \tag{3.5}$$

Main working age adults (25–59):

$$A_c = \frac{k}{n} * \frac{\left(P_{c,t} - P_{c,t-1}\right)}{\left(S - P_{c-1,t-1}\right)} \tag{3.6}$$

Older adults (60-plus):

$$L_c = \frac{k}{n} * \frac{\left(P_{c,t} - P_{c,t-1}\right)}{P_{c,t-1}} \tag{3.7}$$

where: k is the age difference of successive cohorts in years, i.e. k=5

n is the time interval in years between two sets of the observed data; i.e. for observations in 2006 and 2011, n=5

acquisition rates are not negative, i.e. $A_c \geq 0$

loss rates not *positive*, i.e. $L_c \leq 0$

Given that four different sets of the most recent NTS data were available, and that the earlier data presented in a more aggregate way at 10 years age intervals for a given year which smoothed out some significant variation by age groups (for instance, the younger age group 21–24 and 25– 29 and the older age group of 70+), only the more recent NTS data collected from 1995 onwards has been used to calibrate the acquisition and loss rates.

4 Calibration of the cohort model

4.1 Base licence holding rates

The historical licence holding rates that have been used to calibrate the model are summarised in Table 4 for males and for Table 5 females. For each time point, three years of NTS data have been used to give sufficient large sample sizes to provide reliable estimates of the licence holding rate. It is noted that for the 1990 data, only more aggregate age band information is available for the 21 to 29, 60 to 69 and 70-plus age ranges.

Table 4: Historical male licence holding rates

Age band	1990 (1989/1991)	1996 (1995–1997)	2001 (2000–2002)	2006 (2005–2007)	2011 (2010–2012)
17–20	0.52	0.50	0.36	0.38	0.35
21–24	0.82	0.78	0.66	0.64	0.61
25–29		0.82	0.80	0.75	0.71
30–34	0.88	0.89	0.87	0.84	0.80
35–39		0.88	0.89	0.88	0.86
40–44	0.89	0.89	0.90	0.88	0.87
45–49		0.89	0.90	0.90	0.90
50–54	0.85	0.91	0.88	0.90	0.89
55–59		0.86	0.88	0.91	0.90
60–64	0.78	0.84	0.86	0.90	0.90
65–69		0.83	0.83	0.87	0.89
70–74	0.58	0.77	0.80	0.82	0.85
75–79		0.66	0.68	0.78	0.81
80 +		0.44	0.52	0.62	0.69

Table 5: Historical female licence holding rates

Age band	1990 (1989/1991)	1996 (1995–1997)	2001 (2000–2002)	2006 (2005–2007)	2011 (2010–2012)
17–20	0.35	0.36	0.30	0.31	0.32
21–24	0.64	0.65	0.57	0.54	0.55
25–29		0.69	0.70	0.68	0.65
30–34	0.67	0.76	0.77	0.75	0.73
35–39		0.72	0.76	0.79	0.78
40–44	0.66	0.74	0.77	0.79	0.79
45–49		0.72	0.77	0.78	0.80
50–54	0.49	0.67	0.72	0.74	0.77
55–59		0.56	0.69	0.74	0.74
60–64	0.33	0.50	0.58	0.66	0.72
65–69		0.41	0.54	0.57	0.67
70–74	0.15	0.32	0.36	0.47	0.55
75–79		0.20	0.27	0.36	0.45
80 +		0.09	0.15	0.19	0.28

Equations (3.5) to (3.8) are then applied to calculate the acquisition and loss rates. For some age bands, smoothing or averaging is undertaken to ensure that the changes in the acquisition/loss rates with age are plausible and that the signs of the rates are correct, i.e. acquisition rates cannot be negative and loss rates cannot be positive.

4.2 Saturation rates

The saturation rates were determined mostly by observing the peak licence holding rate at any age for the historical data. In some cases, small adjustments were made to the values to improve the validation, for example lowering the national saturation rate from 0.93 to 0.92 was observed to significantly increase acquisition rates for males in their 30s and this in turn was judged to give rise to more plausible forecasts. On the basis that male and female licence holding rates are increasingly converging the same saturation rates are used for males and females.

The final values are detailed in Table 6.

Table 6: Saturation rates by gender and area type

Area type	Males	Females
Inner London	0.92	0.92
Outer London	0.95	0.95
Metropolitan Districts	0.87	0.87
Non-met districts, > 10 person/ha	0.92	0.92
Non-met districts, 2–10 person/ha	0.95	0.95
Non-met districts, < 2 person/ha	0.97	0.97
National	0.92	0.92

It can be seen that for non-metropolitan districts the assumed saturation rates increase as population density reduces, as might be expected.

4.3 Calculation of national acquisition and loss rates

The national acquisition and loss rates were calculated on the 'Acquistion_Rates' tab of a spreadsheet named 'natcop_lic_proj_base_v18_to_DfT.xlsx' that was delivered to the Department on 24 June 2016. An explanation of the licence cohort spreadsheet model is included in the Appendix.

Table 7 summarises the acquisition and loss rates (per year, based on the average changes over a 5 year time period) that have been calculated for males.

Table 7: Male licence acquisition and loss rates (per year)

Age band	A_c / L_c					
	1996	2001	2006	2011	weighted average	final
17–20	-0.0304	-0.3429	0.0422	-0.0528	-0.0960	0.0000
21–24		-0.8206	-0.0958	-0.1171	-0.3445	0.0000
25–29		0.1158	0.3207	0.2680	0.2348	0.2943
30–34		0.5089	0.3952	0.3052	0.4031	0.4031
35–39	0.0667	0.2366	0.0915	0.1692	0.1410	0.1692
40–44	0.2125	0.5705	-0.4658	-0.0582	0.0648	0.0925
45–49	-0.1688	0.4564	-0.3618	0.5551	0.1202	0.0925
50–54	0.4140	-0.1458	-0.2090	-0.4682	-0.1023	0.0000
55–59	0.0887	-2.3799	0.6064	0.1075	-0.3943	0.0000
60–64	-0.0168	-0.0017	0.0224	-0.0103	-0.0016	0.0000
65–69	0.0514	-0.0007	0.0118	-0.0059	0.0141	0.0000
70–74		-0.0347	-0.0163	-0.0170	-0.0227	-0.0227
75–79		-0.1099	-0.0204	-0.0139	-0.0480	-0.0480
80 +		-0.2037	-0.0928	-0.1194	-0.1386	-0.1386

For some age bands, the final rates were determined as follows:

- for the 17–20 and 21–24 band bands, the acquisition rates were set to zero (so base rates will remain fixed in the future) on that basis that we do not know whether the trend for delayed licence acquisition will continue into the future – sensitivity tests could be run to investigate the impact of different assumptions here
- for the 35–39 age band, following testing of the impact of using the average acquisition rate calculated across years the acquisition rate for the most recent 2011 data was used to reflect a 'catch up' effect following delayed licence acquisition when these individuals were younger
- for the 40 to 49 age range, an average of the 40–44 and 45–49 age bands was used to smooth the acquisition rates
- for the 50 to 69 age range, the rates were smoothed by setting them all to zero

Table 8 summarises the acquisition and loss rates that have been calculated for females.

Table 8: Female licence acquisition and loss rates

Age band	A_c / L_c					
	1996	2001	2006	2011	weighted average	final
17–20	0.0213	-0.1161	0.0228	0.0193	-0.0132	0.0000
21–24		-0.2819	-0.0653	0.0256	-0.1072	0.0000
25–29		0.2018	0.3292	0.2905	0.2739	0.2739
30–34		0.3640	0.2243	0.2024	0.2636	0.2636
35–39	0.1876	0.0186	0.0981	0.1651	0.1173	0.1400
40–44	0.2653	0.2729	0.1941	0.0158	0.1870	0.1400
45–49	0.2253	0.1520	0.0392	0.0414	0.1145	0.1400
50–54	0.0415	0.0251	-0.1778	-0.0964	-0.0519	0.0000
55–59	0.1578	0.0604	0.0576	0.0305	0.0766	0.0000
60–64	0.0266	0.0481	-0.0341	-0.0272	0.0033	0.0000
65–69	0.2401	0.0679	-0.0232	0.0038	0.0722	0.0000
70–74		-0.1115	-0.1237	-0.0444	-0.0932	-0.0768
75–79		-0.1377	0.0042	-0.0419	-0.0585	-0.0768
80 +		-0.2807	-0.3091	-0.2260	-0.2719	-0.2719

For some age bands, the final rates were determined as follows:

- for the 17–20 and 21–24 band bands, again the acquisition rates were set to zero (so base rates will remain fixed in the future) on that basis that we do not know whether the trend for delayed licence acquisition will continue into the future
- for the 35 to 49 age range, an average of the 35–39, 40–44 and 45–49 age bands was used to smooth the acquisition rates
- for the 70-plus age range, a weighted average of the 70–74, 75–79 and 80-plus rates was calculated that took account of the fraction of population in each age band – this was done because with higher age-band specific loss rates for the 80-plus band the predicted drop of in licence holding in 2014 was judged to be too high

5 Model validation and sense checking

The models have been validated and sense checked in two ways:

1. first, by making a short cohort forecast from 2011 to 2014, and validating the projections against observed 2014 NTS data; and
2. second, by projecting forward to 2051 and checking that the pattern of changes appears plausible.

5.1 Validation of 2014 predictions against observed NTS data

A limitation of the 2014 validation is that the observed 2011 data was calculated as an average of 2009–2013 data to ensure sufficient sample sizes to calculate the base year licence holding rates, whereas the 2014 data is a single year of data.[7] This means that the validation data is prone to uncertainty around the observed rate.

[7] We do not have data beyond 2014, so cannot for example take a 2013–2015 average.

Figure 7: Validation of 2014 predictions, Inner London, males

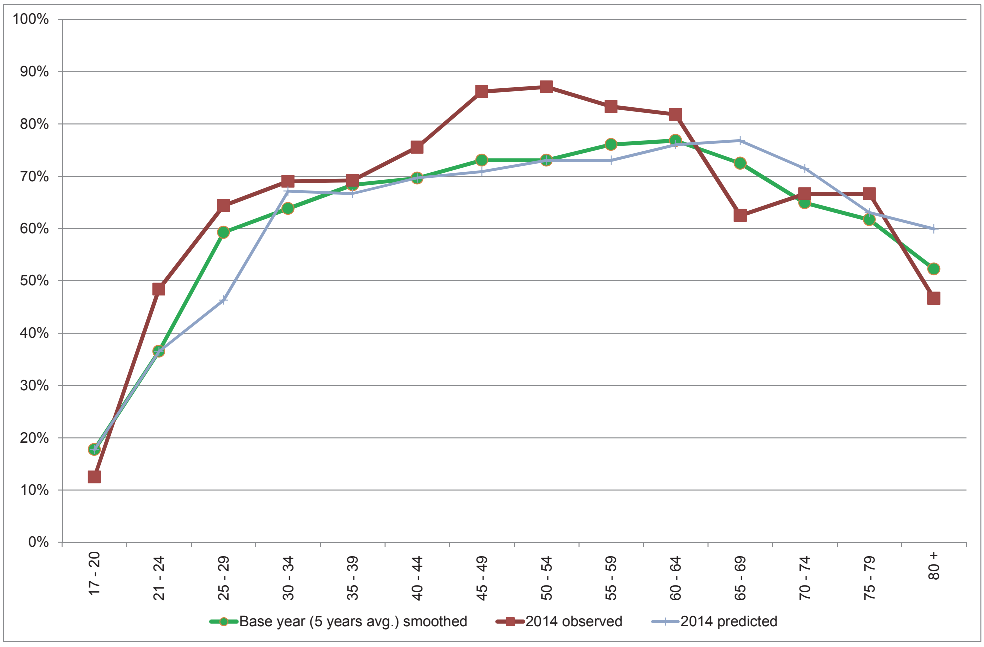


Figure 8: Validation of 2014 predictions, Inner London, females

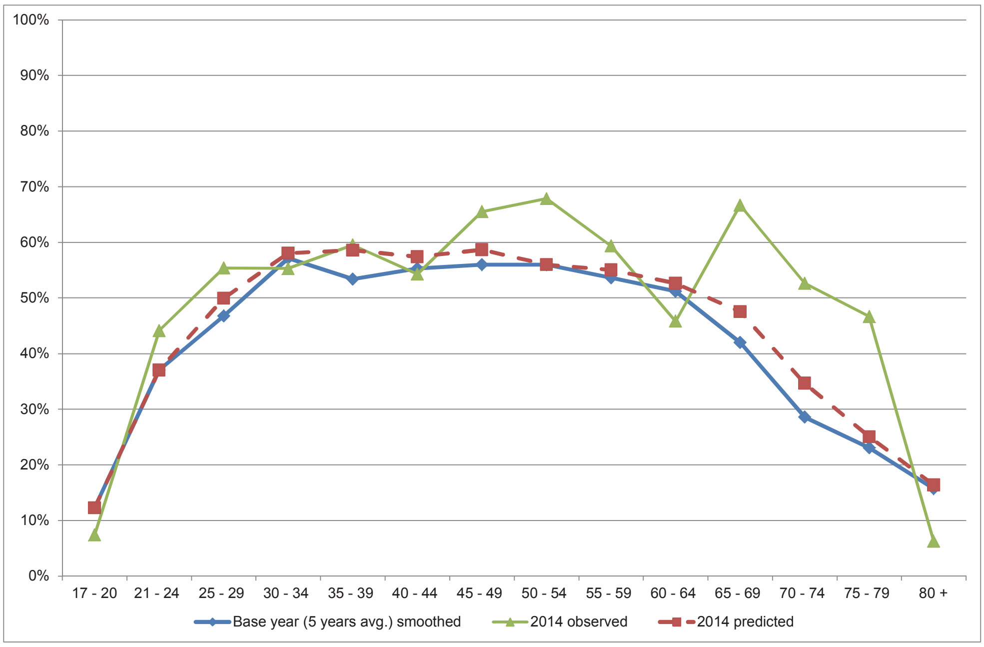


Figure 9: Validation of 2014 predictions, Outer London, males

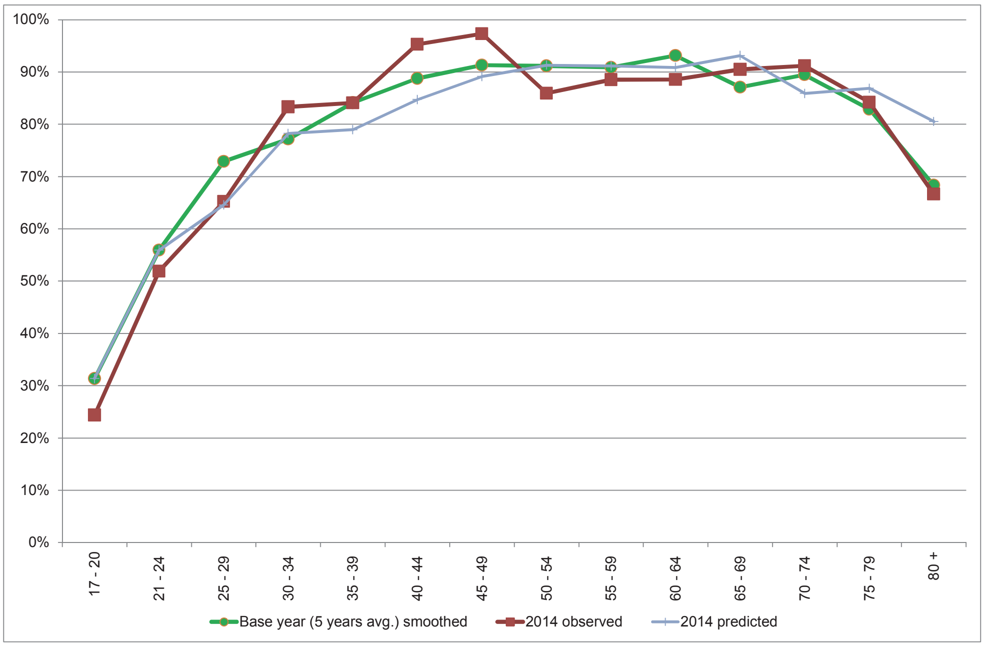


Figure 10: Validation of 2014 predictions, Outer London, females

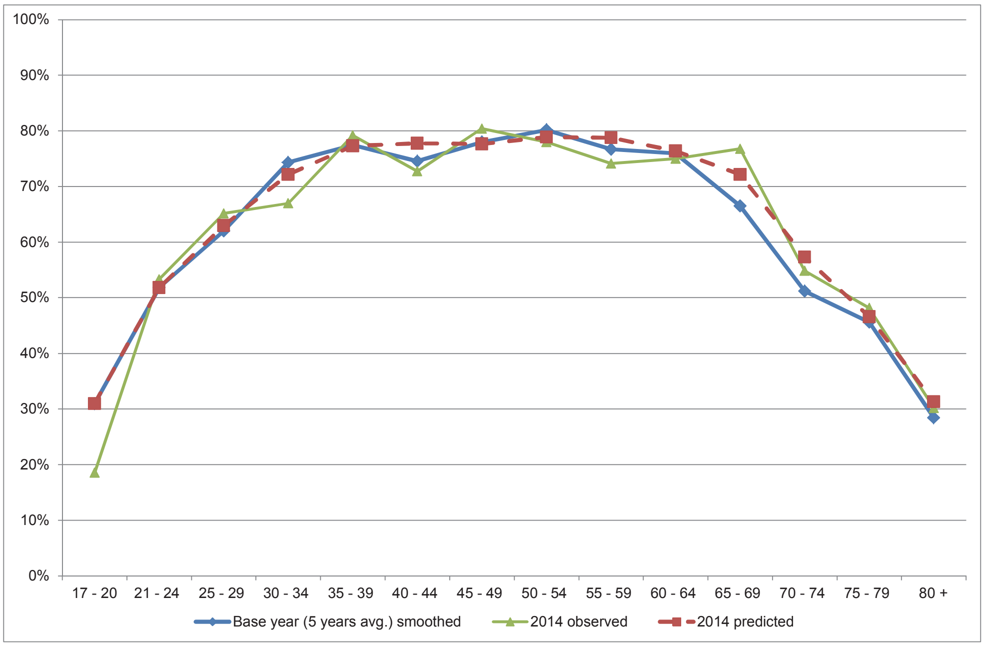


Figure 11: Validation of 2014 predictions, Metropolitan Districts, males

Figure 12: Validation of 2014 predictions, Metropolitan Districts, females

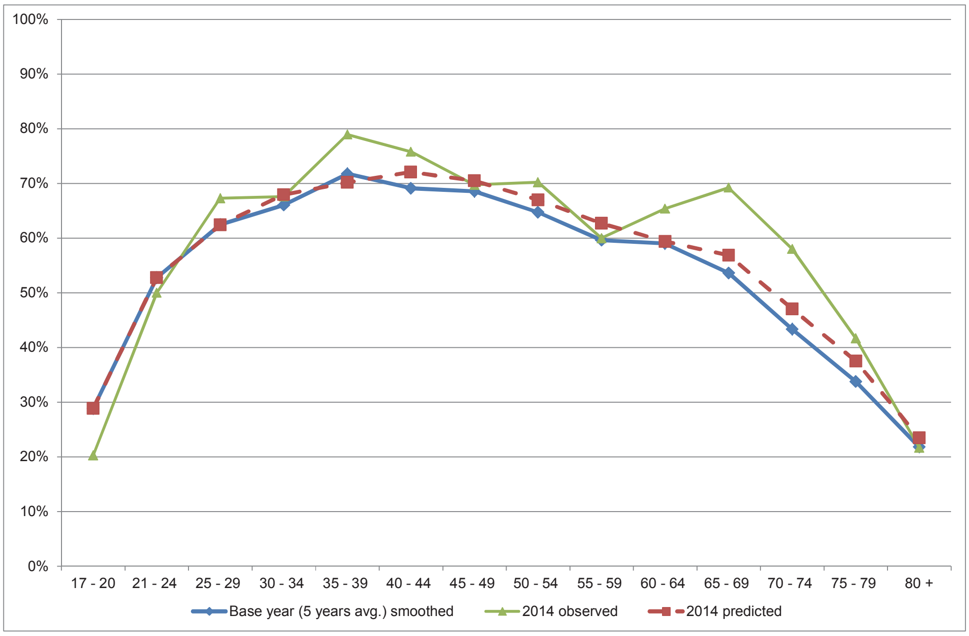


Figure 13: Validation of 2014 predictions, non-Metropolitan Districts >10 pers/HA, males

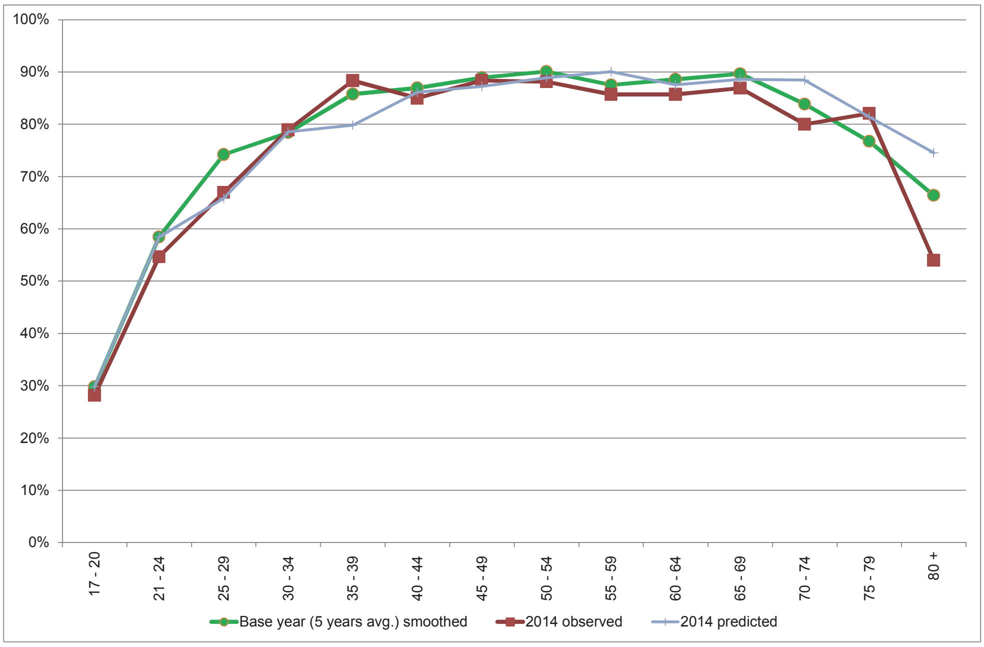


Figure 14: Validation of 2014 predictions, non-Metropolitan Districts >10 pers/HA, females

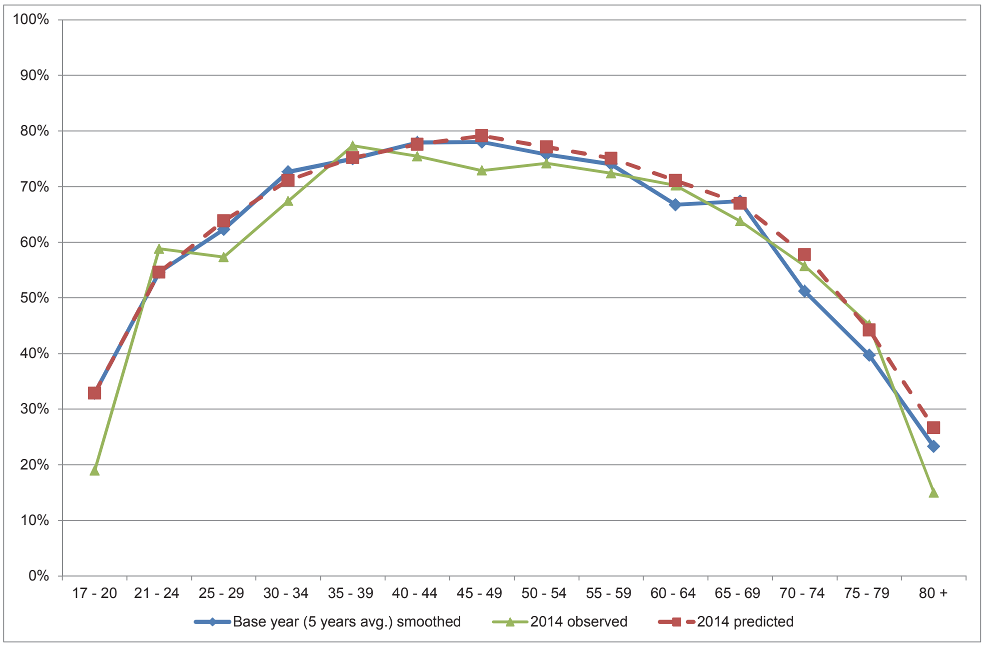


Figure 15: Validation of 2014 predictions, non-Metropolitan Districts 2–10 pers/HA, males

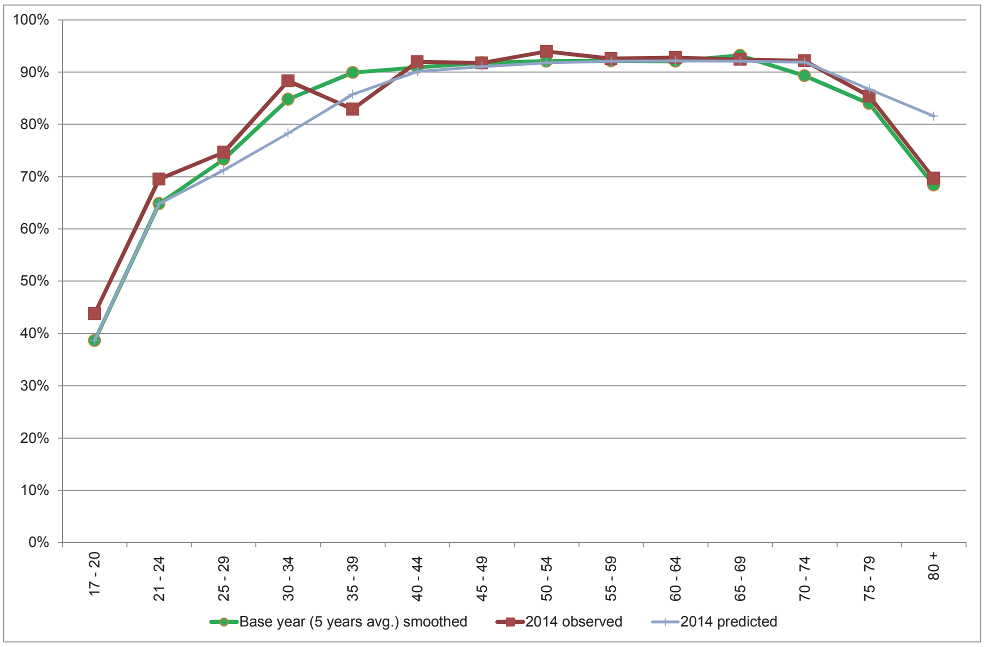


Figure 16: Validation of 2014 predictions, non-Metropolitan Districts 2–10 pers/HA, females

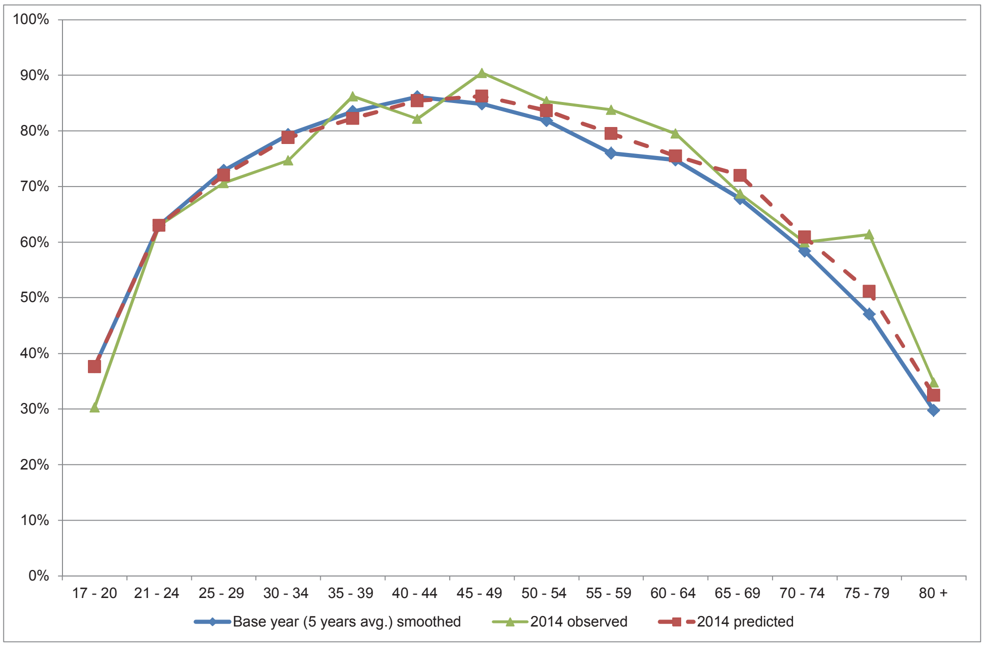


Figure 17: Validation of 2014 predictions, non-Metropolitan Districts < 2 pers/HA, males

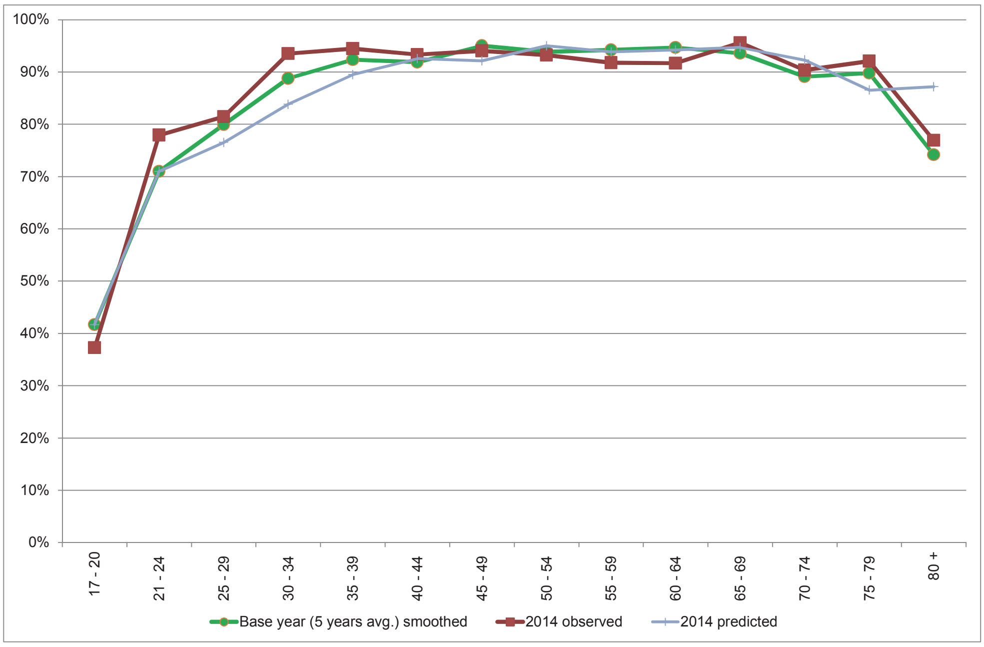
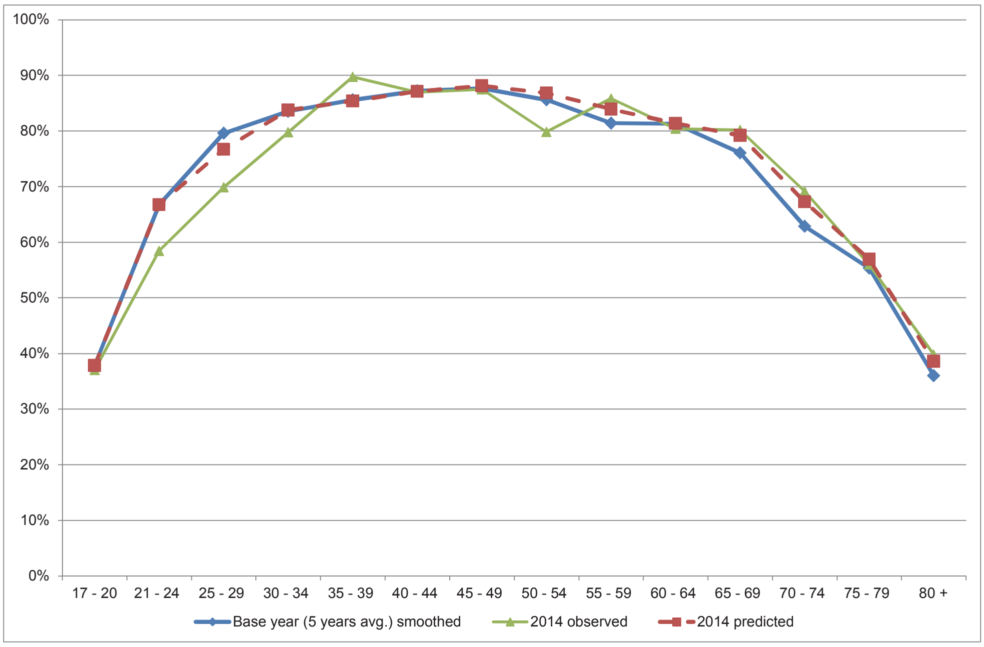


Figure 18: Validation of 2014 predictions, non-Metropolitan Districts < 2 pers/HA, females



A number of general patterns emerge from the validation:

- The considerable volatility in the 2014 observed values which are based on one year of NTS data in comparison to the smoother values, helping illustrate why we used five year averages for the 2011 base rates;
- The cohort effect for older female licence holding whereby higher licence holding is predicted in 2014 than 2011 as females retain their licences into older age; and
- The relationship between licence holding and population density in both base and projected plots, namely more dense areas have lower observed licence holding in the base year and that difference is projected forward into the future.

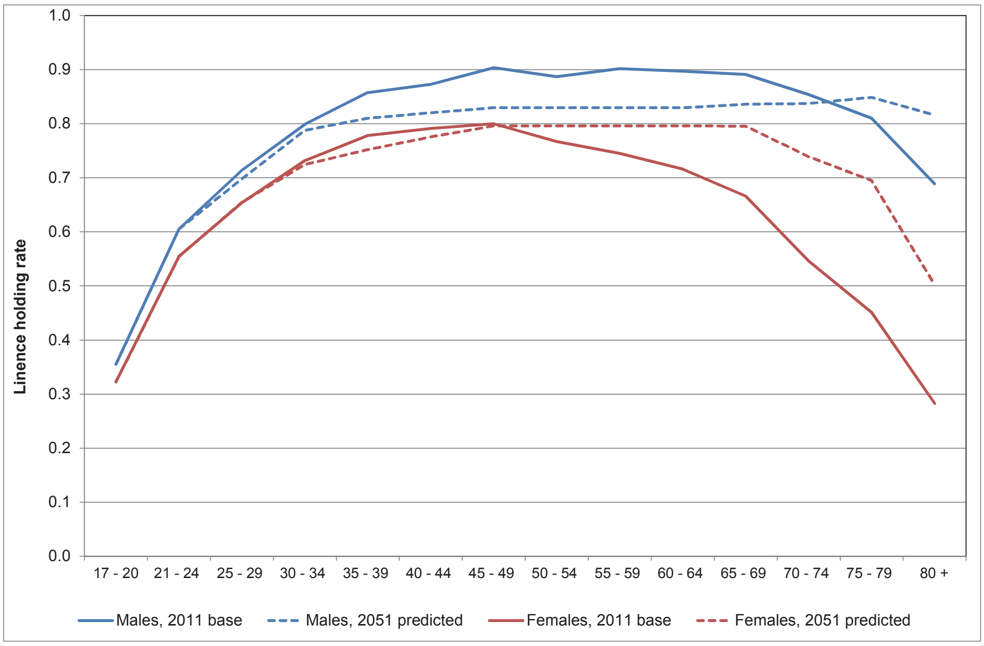
5.2 Sense check of 2051 projections

The cohort model predicts licence holding rates at five year intervals from the 2011 base year through to 2051.

National projections are presented here, the separate cohort spreadsheets for each area types create their own projections which vary as a function of the base rates and saturation rates by area type.

Figure 19 compares the 2051 projections for males and females to the observed 2011 base rates.

Figure 19: 2051 projections for males and females



The projections show the impact of the cohort effect for older people and particularly older females, whose licence holding is predicted to increase as individuals retain licences as they age.

The impact of delayed licence acquisition can be seen during working ages, where for males lower licence holding is predicted than is currently observed. This prediction is sensitive to assumptions on the extent to which licence holding in this age range will 'catch up' to compensate for delayed acquisition for younger persons. The Department's view on this prediction is welcomed, sensitivity tests of different assumptions are one option.

The final trend that is observed is the increases convergence between male and female licence holding levels, with only the oldest two cohorts showing significantly higher male licence holding.

6 Integrate the licence holding model into the car ownership model

In the previous car ownership model, the licence holding rates was incorporated in the model by an average Licence per adult (LPA) measure which was a GB-wide average value that varies only by year. Therefore it only reflects the aggregate licence holding over time, but not the cross-sectional variation in licence holding between households.

Differently from present studies, the impacts of licence holding changes have been better represented by incorporating the cross-sectional variations of licence holding into the model.

To achieve this, LPA is included in the base model estimation as two different terms: the individual LPA term (by age-gender cohort and area type) and the difference between the individual LPA term and the annual average LPA. Therefore the variation of the licence holding by age-gender and different area type over years has been reflected in the new car ownership model.

For the implementation, for each household, an average LPA is calculated, by summarising the number of the adults by which multiplied the licence holding rates projection of their age-gender cohorts. For a given time period t,

$$LPA_h = \frac{\sum LR_{ci} \times N_{ci}}{\sum N_{ci}}$$

where: h represents the household

LR_{ci} is the licence holding rates for the age-gender cohort ci

N_{ci} represents the number of adults for the cohort ci

Therefore the future changes in licence holding rates will affect the average LPA calculated for the household, and so lead to the changes in the predicted probabilities of the household owning a car.

7 Summary

A licence cohort model has been developed from NTS data that provides forecasts of licence holding by age band and gender. These forecasts are provided separately by the six area types used in the new version of NATCOP, reflecting the significant variations in licence holding between more densely and less densely populated areas. The base year for the cohort model is the 2011 base year used in the new version of NACTOP, and the licence holding forecasts are available for five year intervals through to 2051.

Analysis was undertaken of licence holding rates by age band and gender at 10 year intervals between the mid-1970s and the mid-2000s. This analysis demonstrated that for males the main change has been increases in licence holding for older males, whereas for females significance increases in licence holding have occurred for people of working age as well as for older females.

The licence holding models have been validated by making a short projection from 2011 to 2014 and comparing those projections to observed 2014 data. Overall the forecast were plausible, an issue for the analysis was considerable volatility in the observed 2014 rates which were based on a single year of NTS data. The volatility issue was overcome for base year rates by taking a five year average of the rates for years straddling the 2011 base year.

The projections to 2051 were also reviewed to check that their plausibility. The 2051 projections clearly show increases in licence holding for older people, and particularly older females, due to cohort effects as a result of individuals retaining their licences as the move from working age into retirement. The impact of delayed licence acquisition for younger age groups is also apparent in the 2051 forecasts; the model is sensitive to the assumptions around licence holding for younger age groups.

Given that the trend for young people to delay licence acquisition is a recent one, a number of assumptions have had to be made in order to forecast how this trend may play out in future. Specifically, for the 17–20 and 21–24 age bands it has been assumed that the current licence holding rates will remain fixed in the future (i.e. there is no further reduction in licence holding for these groups), and for some later age bands slightly higher acquisition rates are assumed for forecasting to reflect a ‘catch up’ effect. The model has been set up in way that readily allows alternative assumptions to be tested.

Appendix: Explanation of the licence cohort spreadsheet model files

The data folder <ToDfT>[8] contains the licence cohort projection files for each region. Below is the explanation for data sources in the calculation:

File **< natcop_lic_proj_base_v18_to_DfT.xlsx >** contains the calculation of the national acquisition rates using the NTS data. All the formulas are remained in this file to facilitate the understanding of the linkage between the data source and the calculation.

File <**Area_type_summary_v3.xlsx**> contains the smoothing process for the licence holding rates for the 6 area types. The file contains detailed explanation. The original rates of each area type are drawn from the analysis of the five years NTS data (2009-2013). The data for the NTS data by region is saved in the folder <Data>.

In the <**Data**> folder, there are 6 files which contain the licence hold rate data for each cohort and each area type from the NTS data. In tab <raw_data_2>, columns CZ to DE show the sample size and calculation of the 5 year licence holding rate for that particular region.

Back to the main folder <ToDfT>, the 6 area type files <**NATCOP_Lic_Proj_xxx_v2.xlsx**> include the licence holding rate projection by each area type. In each file:

Tab <**input**> contains the basic input for the licence projection. Among them,

- 'Saturation Level S' is determined by the highest level (slightly higher than the highest level) of the licence holding rate for each area type.
- Acquisition rates: are the rates obtained from the national acquisition rates
- Weights and survey intervals are used in the national acquisition rates calculation.

Tab < **Acquisition_rates_national**> contains the national acquisition rates for each age cohort by gender. The final rates are saved in column I.

In tab <**validation**>,

- Column C (5 years sample size) is from the area calculation file in the folder <Data> as mentioned above in tab <raw_data_2>.
- Column D (base year – 5 year avg. smoothed) is from file <**Area_type_summary_v3.xlsx**>, tab < 2011 holding rates> columns <M – R>.

[8] The files discussed in the Readme were sent to DfT on 15th September 2016.

- Column E (2014 observed) is also from the area calculation file in the folder <Data>, tab <validation> column E. Formula is kept in the file to show the linkage between the data sources.
- The rest columns kept the formulae which are shown in column H.

Then in tab < **Projections**>, column D is same as the column D in tab <validation>. The formulae are kept for the other columns.