

# **A DISAGREGATE MODEL OF CAR OWNERSHIP BASED ON FRENCH NATIONAL TRANSPORT SURVEY**

**Richard GRIMAL**

**French Technical Department of Studies on Roads, Transports and Highways**

## **INTRODUCTION**

French National Transport Survey, conducted from May 2007 to May 2008, provides opportunities for analyzing changes in mobility and car ownership behaviours, by comparison with the previous survey in 1993-1994. After publishing descriptive analysis, in this paper we develop discrete choice models of car ownership from National Transport Survey. Indeed, car ownership modelling appears to be a necessary step in trying to model mobility behaviours, as a determining indicator of mobility potential, most trips being realized by car. More exactly, we are trying to simulate the number of cars in households from information about their socio-demographic characteristics, housing patterns, residential locations, built environment (high- or low-density areas), and indicators of mobility potential such as number of driving licenses. The sample, containing about 20 000 households, is used for model specification, estimation, and simulation, which is also performed on external census files to check model robustness. Eventually, we are trying to measure changes in car ownership determinants by building specific models from previous National Transport Survey in 1994 and applying them to - or estimating them with -, current data. Simulation results provide evidence that accurate forecasts about future car ownership distribution cannot be derived from models specified and estimated on cross-sectional data. It is thus necessary to separate changes in car ownership distribution caused by explanatory variables (for instance socio-demographic change and land use patterns), from changes in the senses and strengths of effects (changing behaviours).

## **1) DRIVING LICENSE, THE MAIN EXPLANATORY VARIABLE OF CAR OWNERSHIP**

In the first section, we implement comprehensive models by identifying the most relevant explanatory variables and categories. In previous works, we already identified some determinant variables of car ownership, such as household size, income per consumption unit, density and centrality of housing area, head of household's age. Here, we implement a more systematic approach by taking into account every possible explanatory factor of car ownership available in datasets. Indeed, some important factors of car ownership may have been omitted by an intuitive approach. Moreover, some variables may be correlated and we don't know *a priori* if they are really influent by themselves or because of their correlation with other explanatory variables. Modelling approaches allow to separate statistically significant effects from non-significant ones, "everything else being equal", based on

assumptions about distribution of error terms. However, before implementing a multinomial model, preliminary binary logistic regression procedures with backward elimination were performed on files containing the response variable and explanatory factors, in order to remove irrelevant variables and categories, appraise signs and strengths of relative effects, and eventually select the most influential variables.

The following variables were considered potentially influent :

- Housing variables : size of urban area, housing type (flat or house), position within urban area, occupational status, availability of secondary housing...
- Socio-demographic characteristics of household : number of working people, head of household's partner existence, number of young children, of students, financial resources...
- Socio-demographic characteristics of head of household : qualifications, work participation, age, sex, socio-professional group.
- Similar information about head of household's partner, if relevant.
- Indicators of mobility potential : number of driving licenses and seasonal transit tickets...

Agregate modalities were defined in order to keep significant effects, from intuitive or *a priori* knowledge and summary frequency analysis. For instance, car ownership is expected to be lower in high-density areas because of alternative transport supply and shorter distances between activity places. We expect it to be higher among couples, but also with young children to accompany to day-nursery or school, or students still living at parent's home, seeking for individual autonomy. It should also increase with work participation and high financial resources. The response variable, household car ownership, was given three levels : car deprivation (no car), single-ownership (one car), and multi-ownership (at least two cars). We were not interested, here, in describing car attributes such as fuel type or weight classes, because we were mainly interested in car ownership as a determiner of mobility potential. Of course, more sophisticated models including car attributes should be implemented in the perspective of analyzing buying behaviours or developing marketing approaches.

In logistic regression, every n-level qualitative variable is factorized into n-1 effects corresponding to differences with the specified reference level. After backward elimination, several effects were removed from models, as they were not significant under the 0.05 significance level, some of them because they were not really influent by themselves, such as residential location in deprived urban areas, head of household's age and sex, distance to closest train station, but others because of redundance with similar explanatory factors, such as number of working adults. Surprisingly, head of household's age was removed from significant effects, despite a low level of car ownership, because in fact low car ownership among old people is mainly caused by low driving license holding.

Signs of effects revealed consistent with a *priori* expectations. Everything else being equal, the most influential factor is household's number of driving licenses. Probabilities of car deprivation strongly increase with residential location in the urban area of Paris, but also with transit seasonal tickets. Symmetrically, students living at home, residential location in peripheric urban areas or high financial resources contribute to high probabilities of multi-ownership.

Influential variables and categories to keep into a multinomial choice model were selected through two criteria : the number of binary regressions where the effect was significant, and the effect strength. An influent effect at least on one level of the response variable was entered into model specification. On the contrary, a weakly significant effect in all regressions was removed. Selection was implemented to develop the most possibly economic models for given performances in terms of goodness-of-fit and forecasting abilities. A series of discrete choice model specifications including both quantitative and qualitative variables were then estimated on National Transport Survey. Alternative-specific constants and parameter values were specified, as shown for instance for the following multinomial logit model described by its utility functions. Indeed, alternative-specific parameters allow variability through alternative's utilities despite constant household's characteristics.

$$U_0 = B0\_LICENSE * n\_permis + B0\_ABONTC * n\_cartabon + B0\_ICENTRE * i\_centre + B0\_IAPPT * i\_appt + B0\_IPARIS * i\_paris + \varepsilon_0.$$

$$U_1 = C_1 + B1\_ICENTRE * i\_centre + B1\_IPARIS * i\_paris + B1\_ABONTC * n\_cartabon + B1\_IAPPT * i\_appt + B1\_IACTIF * i\_pr\_actif + B1\_CONJOINT * i\_conjoint + B1\_CONJACT * i\_conjoint\_actif + B1\_NENFMAJ * n\_enfants\_maj + B1\_IPERIURBAIN * i\_periurbain + B1\_IPROPRIO * i\_proprio + B1\_IQ4 * i\_q4 + \varepsilon_1.$$

$$U_2 = C_2 + B2\_LICENSES * n\_permis + B2\_NENFMAJ * n\_enfants\_maj + B2\_IPERIURBAIN * i\_periurbain + B2\_IPROPRIO * i\_proprio + B2\_IQ3 * i\_q3 + B2\_IQ4 * i\_q4 + B2\_IACTIF * i\_pr\_actif + B2\_CONJOINT * i\_conjoint + B2\_CONJACT * i\_conjoint\_actif + \varepsilon_2.$$

$U_i$	Utility of alternative $i$ (0 = no car, 1 = one car, 2 = two cars or more)
$n\_permis$	number of driving licenses
$n\_cartabon$	number of seasonal transit tickets
$i\_centre$	indicator variable for homeplace in central city
$i\_appt$	indicator variable for apartment
$i\_paris$	indicator variable for homeplace in the urban area of Paris
$i\_pr\_actif$	indicator variable for head of household's work participation
$i\_conjoint$	indicator variable for partner's existence
$i\_conjoint\_actif$	indicator variable for partner's work participation (set to 0 if partner doesn't exist)
$n\_enfants\_maj$	number of major children at home
$i\_periurbain$	indicator variable for homeplace in periferic crowns of urban areas
$i\_proprio$	indicator variable for property owning
$i\_q3$	indicator variable for 3rd quartile of income per consumption unit
$i\_q4$	indicator variable for 4th quartile of income per consumption unit

In this specification, it is assumed that household's number of driving licenses mainly influences utilities of "car deprivation" and "multi-ownership". Probabilities of car deprivation increase with seasonal transit tickets, residential location in city-centre or in the urban area of Paris, living in apartment. On the contrary, probabilities of multi-ownership are supposed to increase with household's number of driving licenses, adult children living at

home, residential location in suburban areas, property owning, high financial resources, head of household's work participation, partner's existence and work participation.

Maximum likelihood estimation of parameter values shows significant improvement by comparison with a trivial model. Log-likelihood comes from  $L(\beta) = -22\,166.7$  to  $L(\beta) = -11\,922.7$  with an estimated adjusted rho\_square to  $\rho^2 = 0.461$ .

Utility parameters				Utility parameters			
*****				*****			
Name	Value	Std err	t-test	Name	Value	Std err	t-test
ASC_1CAR	-1.83	0.0857	-21.37	B1_IAPPT	0.576	0.0631	9.13
ASC_2CARS	-6.94	0.135	-51.38	B1_ICENTRE	0.453	0.0597	7.59
ASC_NOCAR	0.00	--fixed--		B1_IPARIS	0.610	0.0525	11.62
B0_ABONTC	0.734	0.0556	13.21	B1_IPROPRIO	0.465	0.0625	7.44
B0_IAPPT	0.910	0.0901	10.10	B1_IQ4	0.293	0.0785	3.73
B0_ICENTRE	1.21	0.0876	13.82	B2_CONJOINT	0.968	0.108	9.00
B0_IPARIS	1.73	0.0851	20.29	B2_IACTIF	1.76	0.0793	22.19
B0_IQ3	-0.390	0.0754	-5.17	B2_PERIURBAIN	0.569	0.0968	5.88
B0_LICENSE	-2.90	0.0720	-40.25	B2_IPROPRIO	0.920	0.0780	11.79
B1_ABONTC	0.184	0.0317	5.82	B2_IQ3	0.412	0.0504	8.17
B1_CONJACT	-0.285	0.120	-2.37	B2_IQ4	0.998	0.0911	10.96
B1_CONJOINT	0.257	0.0758	3.39	B2_LICENSES	1.89	0.0627	30.11
B1_IACTIF	1.09	0.0641	17.03	B2_NENFMAJ	0.435	0.0925	4.71

Parameter signs are consistent with expectations and parameter values are significant at the 0.05 level. By considering respective strengths of parameter values, household's number of driving licenses appears to be by far the most influential effect on the level of car ownership. Such a result may seem logic as it is the main filter to driving.

	No car	One car	Two cars	Three or more	Total
<b>No license</b>	97.9	2.0	0.1	0.0	100.0
<b>One license</b>	21.4	72.6	5.4	0.6	100.0
<b>Two licenses</b>	2.0	33.7	58.3	6.0	100.0

Table 1 : rate of households (%) from a given level of car ownership, for a given number of driving licenses

Source : National Transport Survey, 2007-2008

Reading guide : among single-licensed households, 5.4 % are holding two cars.

A cross-analysis of household's number of cars and driving licenses reveals a very strong correlation between driving license holding and car ownership. Most license-deprived households don't hold any cars, which is obvious and in some sense comforting. The major case is that the number of cars matches the number of driving licenses, and the minor case that number of cars equals number of driving licenses minus one. Some single-licensed households don't have any car, and some multi-licensed households only have one car. Single-licensed households may be either singles, or couples with only one driving adult. This situation may happen for instance among young couples where

one partner hasn't passed its driving license examination yet. Single-licensed households may also be old couples, where some women didn't have access to driving license. Eventually, multi-licensed households holding only one car may be for instance couples with only one member really needing a car for daily commuting trips.

Given the very strong relationship between driving license holding and car ownership, we can imagine another possible model form, where the choice process for the level of car ownership is dependent on household's number of driving licenses. For instance, we may try to estimate a segmented multinomial logit model, assuming that parameter values may vary from one segment to another. Thus, segmented multinomial logit models were estimated separately for two segments - "single-licensed" and "multi-licensed" households - and compared to the global model for at least one-licensed households, using a likelihood ratio test. License-deprived households were excluded from model estimation as they have no real other choice than being car-deprived.

We obtained for likelihood ratio test,  $LR = -2 (-11\ 406.6 + 4\ 186.9 + 7\ 171.4) = 96.6$ , which is compared to a  $\chi^2$  distribution at the 0.05 level of significance with 24 degrees of freedom, corresponding to the number of supplementary freedom degrees generated by model segmentation. As we have  $LR \gg \chi^2(0.05, 24) = 36.4$ , we can reject the null hypothesis and conclude to an interest for model segmentation. However, as gain from segmentation is moderate, selecting a relevant model requires seeking for the optimal trade-off between goodness-of-fit and complexity. Indeed, a high number of supplementary parameters are generated by segmentation, and a segmented model is more complex to apply.

We can wonder if alternative model forms would not better describe the choice process of car ownership level. For instance, we could assume a strong correlation between single- and multi-ownership, that could be gathered in a "car-owning" nest. Such assumptions correspond to nested-logit models. However, nested-logit models didn't provide very satisfying results, as car deprivation and multi-ownership are negatively correlated, while single-ownership is quite stable. Thus, we asked ourselves if cross-nested logit models wouldn't be more appropriate. Single-ownership is more likely to be correlated with both car deprivation and multi-ownership, as it is suggested by cross-analysis of car ownership and driving license holding.

We specified and estimated cross-nested logit models with two nests, defined as "low-equipment" (car deprivation or single-ownership), and strong equipment (single- or multi- ownership). With the same specification, the best results were obtained with  $\alpha_0 = 0.125$  and  $\alpha_2 = 0.875$ , corresponding to a stronger correlation with multi-ownership, which is partly resulting from exclusion of license-deprived households. However, optimal values of  $\alpha$ 's may be doubtful as log-likelihood curve with varying  $\alpha$ 's is quite flat. In terms of goodness-of-fit, a likelihood ratio test shows that a CNL model is approximately similar to a segmented MNL model. Thus, gains obtained from more complex model forms than simple MNL models look limited, though

better representing the choice process. However, we cannot exclude that alternative specifications would produce better results, for instance with limited subsets of alternatives in each segment.

## II) SIMULATION RESULTS REVEAL GOOD SENSITIVITY TO CONTRASTED SUBSAMPLES :

In this section we will try to forecast car ownership distribution on various subsamples of National Transport Survey, but also on external census data, by simulation. For instance, car ownership distribution was simulated on random subsamples which size was about one third of the original sample. Individual probabilities were calculated for each observation and alternative, given socio-demographic characteristics, housing location, urban environment, and conditions of access to mobility. Average probabilities then provide estimators of car ownership distribution.

Subsample	P(0)	P(1)	P(2)	E=0	E=1	E>=2
T <sub>1</sub>	0.09	0.47	0.44	0.09	0.47	0.44
T <sub>2</sub>	0.09	0.46	0.45	0.09	0.46	0.45
T <sub>3</sub>	0.09	0.46	0.45	0.09	0.46	0.45
<b>Urban area of Paris</b>	0.19	0.48	0.33	0.18	0.49	0.33
<b>Urban areas - &gt; 500 000 dwellers</b>	0.08	0.46	0.46	0.08	0.46	0.46
<b>Midi-Pyrénées</b>	0.06	0.48	0.46	0.05	0.47	0.48
<b>Pays-de-la-Loire (id)</b>	0.06	0.45	0.49	0.05	0.45	0.5
<b>Rhône-Alpes (id)</b>	0.06	0.45	0.49	0.06	0.44	0.5
<b>Urban areas - 100 to 500 000 dwellers</b>	0.06	0.46	0.48	0.05	0.45	0.5
<b>City-centres</b>	0.21	0.53	0.26	0.21	0.54	0.25

*Table 2 : Simulated car ownership distribution on various subsamples of National Transport Survey, compared to real car ownership distribution*

*Reading guide : in the first random sample T<sub>1</sub>, simulated rate of car deprivation P(0) is 9 %, and real rate of car deprivation (E=0) is also 9 %.*

The above array shows results from simulation with the previous CNL model. It is not surprising that car ownership distribution is correctly forecast on random subsamples of National Transport Survey as it was also used for model estimation. It is more interesting to appraise model sensitivity to specific conditions. Indeed, from an operational point of view, we expect disaggregate models to recreate variations in the response variable from individual specificities in various subsamples. In order to appraise model sensitivity, the model was applied to territories of specific profiles. For instance, high-density areas such as the urban area of Paris or city-centres are characterized by very low car ownership. On the contrary, Pays-de-la-Loire and peripheric crowns of urban areas are characterized by high car ownership.

Distributions specific to high-density areas were well recreated by indicator variables introduced in the model from the beginning. Simulated distributions on French administrative areas such as Pays-de-la-Loire and Rhône-Alpes, which weren't used for model estimation through indicator variables, are also

quite satisfying. Upward shift in car ownership distribution for these areas, which are characterized by a high rate of multiownership, is correctly recreated by the model, in sense and even quite well in intensity, the rate of multi-ownership being only a little under-estimated.

Differences between real and simulated car ownership distribution on specific subsamples may be caused by some bias in estimators of parameter values, because of endogeneity problems. More exactly, known factors of multi-ownership may be negatively correlated with error terms, leading to under-estimation of multi-ownership effects. Endogeneity problems won't be solved here, but experiments made on various subsamples might help to improve model specification by taking into account factors that would have escaped to first analysis, as we will see when trying to apply models to census data.

Indeed, models also have to be applied on external data to check their solidity. Ideally, they will have to show low sensitivity to various datasets and high sensitivity to specific subsamples. However, previous models couldn't be applied directly on external data, as some variables used for model estimation were missing in census files that were used for simulation, for instance number of seasonal transit tickets or household income, and modalities could also be different. An interesting aspect of census files, anonymized for confidentiality, is their availability for the year of National Transport Survey, in 2007.

If models are correct, simulation on census files should correctly recreate car ownership distribution, because census files were themselves used for sample stratification in National Transport Survey. Simulation, however, could hardly be applied directly on census files, because they contained nearly 30 millions lines, mobilizing too many computational resources. Instead, we used subsamples produced with random drawings at about 1 for 1 000. Subsamples were then restrained to "main home" and French metropolitan territory, excluding overseas islands. By definition, there is exactly one "main home" per household. People less than fifteen also had to be excluded. These limitations were made to obtain subsamples consistent with National Transport Survey.

Another difficulty was that the number of driving licenses, the main explanatory variable of car ownership, was also missing in census files. It brought us to build a specific model for driving license holding, in order to simulate it from census files. Though this two-stage modeling process was not absolutely necessary from a modelling point of view (we could have built a direct car ownership model from available data), it better recreates the natural logic of access to car mobility through a series of successive filters : driving license, car availability, priorities of car use within household.

The following MNL model was built up for the number of driving licenses, where  $U_0$ ,  $U_1$  and  $U_2$  now stand for respective utilities of 0, 1 or 2 licenses in household:

$$U_0 = B0\_APPT * i\_appt + B0\_ICENTRE * i\_centre + B0\_AIRE\_PARIS * i\_aire\_paris + B0\_PRINFBAC * i\_pr\_infbac + B0\_PRCYCLE23 * i\_pr\_cycle23 + B0\_PR\_FEMME * i\_femme + B0\_AGE65 * i\_age65 + B0\_AGE25 * i\_age25 + \varepsilon_0.$$

$$U_1 = C_1 + B1\_APPT * i\_appt + B1\_ICENTRE * i\_centre + B1\_IPERIURBAIN * i\_periurbain + B1\_PROPRIO * i\_proprio + B1\_AIRE\_PARIS * i\_aire\_paris + B1\_NACTIFS * NACTIFS + B1\_NJEUNENF * n\_jeunes\_enfants + B1\_PR\_FEMME * i\_femme + B1\_AGE25 * i\_age25 + B1\_AGE65 * i\_age65 + \varepsilon_1.$$

$$U_2 = B2\_PROPRIO * i\_proprio + B2\_IPERIURBAIN * i\_periurbain + B2\_NACTIFS * NACTIFS + B2\_NETUDIANTS * n\_etudiants + B2\_PRINFBAC * i\_pr\_infbac + B2\_PRCYCLE23 * i\_pr\_cycle23 + \varepsilon_2.$$

Most variables are common with car ownership model displayed in page 3 and therefore have the same meaning. However, some specific variables were entered because of their availability in census files, namely :

<i>i_pr_cycle23</i>	second and third cycle qualifications (university graduates)
<i>i_pr_infbac</i>	below bachelor
<i>i_pr_femme</i>	indicator variable for head of household being a woman
<i>B0_AGE65</i>	indicator variable for head of household more than 65
<i>B0_AGE25</i>	indicator variable for head of household less than 25
<i>n_jeunes_enfants</i>	number of children aged less than 11
<i>n_etudiants</i>	number of scholars/students aged more than 14

Utility parameters

Utility parameters

\*\*\*\*\*

\*\*\*\*\*

Name	Value	Std err	t-test	Name	Value	Std err	t-test
ASC_1LICENSE	1.49	0.0609	24.51	B1_APPT	0.716	0.0529	13.53
ASC_NOLICENSE	0.00	--fixed--		B1_ICENTRE	0.260	0.0514	5.05
B0_AGE25	0.766	0.133	5.75	B1_PERIURBAIN	0.337	0.0875	3.85
B0_AGE65	0.297	0.0751	3	B1_NACTIFS	0.473	0.0510	9.27
B0_AIRE_PARIS	0.630	0.0749	8.42	B1_NJEUNENF	-0.238	0.0280	-8.49
B0_APPT	0.817	0.0787	10.38	B1_PROPRIO	0.871	0.0649	13.43
B0_ICENTRE	0.405	0.0757	5.36	B1_PR_FEMME	1.17	0.0402	29.22
B0_PRCYCLE23	-1.37	0.116	-11.81	B2_PERIURBAIN	0.501	0.0895	5.60
B0_PRINFBAC	-0.405	0.0589	-6.88	B2_NACTIFS	1.82	0.0499	36.45
B0_PR_FEMME	2.10	0.0626	33.61	B2_NETUDIANTS	0.356	0.0458	7.78
B1_AGE25	0.318	0.104	3.06	B2_PRCYCLE23	0.464	0.0567	8.17
B1_AGE65	-0.556	0.0540	-10.30	B2_PRINFBAC	0.126	0.0420	2.99
B1_AIRE_PARIS	0.369	0.0474	7.80	B2_PROPRIO	2.28	0.0687	33.20

Parameter signs revealed consistent with a priori expectations. Probabilities of license deprivation are higher for young (less than 25) and old (more than 65) people, mainly students and retired. They also increase in the urban area of Paris, city-centres, for a feminine head of household. On the contrary, probabilities for a household of being multi-licensed increase with work participation, property owning, living place in peripheric crowns of urban areas, high qualifications, students living home. Determinants of driving license holding within households are similar to those of car ownership. Log-likelihood comes from  $L(\beta) = -22\ 166.7$  to  $-12\ 735.6$  with model estimation, and adjusted rho-square is  $\rho^2 = 0.424$ .

Car ownership model was also re-estimated with a different specification taking into account variable's availability in census files. Fortunately, loss in goodness-of-fit caused by limited variable availability in census files remains moderate [ $L(\beta) = -12\,132.4$  with  $\rho^2 = 0.452$ ] as they contain most of known explanatory variables.

In the first forecasting results of this two-stage process, the rate of car deprivation was correctly recreated, but multi-ownership was over-estimated. Further examination revealed that both models of driving license holding and car ownership were biased because they were estimated on the unweighted sample, leading to over-estimation of multi-ownership effects. Indeed, sample stratification in national transport survey over-represents multi-ownership, representativeness being ensured by corrective weightings.

Subsample	P(0)	P(1)	P(2)	E=0	E=1	E>=2
T <sub>1</sub>	0.206	0.476	0.318	0.193	0.475	0.332
T <sub>2</sub>	0.205	0.477	0.318	0.192	0.476	0.332
T <sub>3</sub>	0.2	0.477	0.323	0.185	0.480	0.335

Table 3 : Simulated car ownership distribution on random subsamples of census files with model estimated on a stratified sample, compared to real car ownership distribution  
Reading guide : same than table 2

Forecasting results become better after re-estimating the model from an unbiased subsample, at least with respect to car ownership distribution. However, this time the model slightly under-estimates multi-ownership. Model estimation would probably be improved by multi-variable stratification, namely if the estimation sample was consistent with margins calculated on several variables. However, there is probably no perfect solution to design an unbiased model, parameter estimates remaining dependent on the structure of estimation samples.

Subsample	P(0)	P(1)	P(2)	E=0	E=1	E>=2
Ile-de-France	0.335	0.408	0.257	0.316	0.476	0.207
Pays-de-la-Loire	0.137	0.495	0.368	0.120	0.472	0.408

Table 4 : Simulated car ownership distribution on contrasted areas from subsamples of census files, compared to real car ownership distribution  
Reading guide : same than table 2

Model sensitivity was then tested with contrasted conditions of land use and socio-demographic structure. Simulation results are given below for Ile-de-France and Pays-de-la-Loire. While people living in Ile-de-France have a very low car ownership, on the contrary a high level of multi-ownership is observable in Pays-de-la-Loire. Interestingly, car ownership distortion is correctly recreated in sense and also to a large amount in intensity, providing evidence of good model sensitivity. Notably, the rate of car deprivation is quite well recreated. However, multi-ownership is over-estimated in Ile-de-France and under-estimated in Pays-de-la-Loire. Thus, model sensitivity is not completely satisfying.

Despite of limited variable availability in census files, models can be improved with other modalities of explanatory variables or still unexploited variables. For

instance, in Ile-de-France and big cities, we know that car-restraining behaviours arise among young adults. Similarly, occupancy status might influence car ownership because of limited parking facilities, and Ile-de-France is characterized by a high rate of tenants. In Pays-de-la-Loire, on the contrary, specific factors of multi-ownership may have been omitted, for instance a high proportion of houses. To take specific ways of living into account, these supplementary parameters were added to model specification through indicator variables. New model estimation followed by simulation on census files was then performed.

<b>Subsample</b>	<b>P(0)</b>	<b>P(1)</b>	<b>P(2)</b>	<b>E=0</b>	<b>E=1</b>	<b>E&gt;=2</b>
<b>Ile-de-France</b>	0.334	0.436	0.230	0.316	0.476	0.207
<b>Pays-de-la-Loire</b>	0.137	0.482	0.381	0.120	0.472	0.408

*Table 5 : Simulated car ownership distribution on contrasted areas from subsamples of census files –enriched model -, compared to real car ownership distribution  
Reading guide : same than table 2*

More accurate forecasts resulted from model enrichment because a larger amount of heterogeneity was taken into account for model estimation. Further developments are possible, of course, to improve forecasts.

### **III) NEED FOR TEMPORAL PARAMETERIZATION OF MODELS RESULTS FROM BEHAVIOURAL CHANGE THROUGH GENERATIONS**

In this chapter we question model stability through long-term periods. More exactly, we are interested in appraising long-term stability of explanatory factors. We will see that future car ownership distribution cannot simply be derived from simulation with models specified and estimated at a determined period, given assumptions about socio-demographic change, future land use patterns, and access to mobility. Indeed, previous analysis provided clues that car ownership also increased “everything else being equal”, for instance, for a given household size or relative income. On the other hand, new “car-restraining” trends, among young adults and executives in high-density areas, are also suggesting us an assumption of growing rationality leading to greater substitutability between modes. Such interpretation refers to latent psychological constructs and behavioural change that cannot be derived from any socio-demographic projections. In other words, if endogeneity problems are solved, models estimated at a given period may be correct to recreate sensitivity on contrasted samples for short-term forecasting, but may be biased when used for long-term forecasting, because of temporal-dependent unobserved heterogeneity caused by changing attitudes, preferences or lifestyles. Other sources of behavioural instability might be temporal-dependent, general economic conditions that are not specific to given individuals, for instance car ownership price dynamics and household’s buying power.

In order to give further evidence of behavioural instability, a model was estimated from National Transport Survey in 1994. It was then applied on National Transport Survey from 2008, to simulate what car ownership distribution would have been under the assumption of stable behaviours and

effects, given observable changes in explanatory variables. Fortunately, stable survey methodology avoided eliminating too much explanatory variables because of variable's availability.

$$U_0 = B0\_ABONTC * n\_cartabon + B0\_AIREPARIS * i\_aire\_paris + B0\_ICENTRE * i\_centre + B0\_IAPPT * i\_appt + B0\_IQ1 * i\_q1 + B0\_IQ2 * i\_q2 + B0\_ARTAGR * i\_art\_agr\_com.$$

$$U_1 = C_1 + B1\_LICENSE * n\_permis + B1\_IAPPT * i\_appt + B1\_NJEUNENF * n\_jeunes\_enfants + B1\_NETUD * n\_etudiants + B1\_IQ3 * i\_q3 + B1\_IQ4 * i\_q4 + B1\_ACTPR * i\_act\_pr + B1\_CONJOINT * i\_conjoint + B1\_CONJACT * i\_act\_conj.$$

$$U_2 = C_2 + B2\_LICENSES * n\_permis + B2\_IPERIURBAIN * i\_periurbain + B2\_BAC * i\_pr\_bac + B2\_SUPBAC * i\_pr\_supbac + B2\_NJEUNENF * n\_jeunes\_enfants + B2\_NETUD * n\_etudiants + B2\_PRFEMME * i\_femme + B2\_ARTAGR * i\_art\_agr\_com + B2\_IQ3 * i\_q3 + B2\_5065 * i\_5065 + B2\_ACTPR * i\_act\_pr + B2\_CONJOINT * i\_conjoint.$$

In the model above, most variables are common with previous models. More specific variables are:

- $i\_q1, i\_q2$  indicator variables for low-income households (1<sup>st</sup> and 2<sup>nd</sup> quartiles)
- $i\_art\_agr\_com$  indicator variables for craftsmen, farmers and traders
- $i\_act\_pr$  indicator variable for head of household's work participation
- $i\_act\_conj$  indicator variable for partner's work participation

Like in 2008, the optimal model was obtained for  $\alpha_0 = 0.125$  and  $\alpha_2 = 0.875$ .

	<b>E=0</b>	<b>E=1</b>	<b>E&gt;=2</b>
<b>Car ownership in 1994</b>	0.251	0.488	0.262
<b>Car ownership in 2008</b>	0.195	0.473	0.332
<b>Simulated car ownership in 2008 with NTS 94 model</b>	0.216	0.497	0.286

Table 6 : Real car ownership distribution in 2008 compared to simulated one with model estimated on National Transport Survey 94

Reading guide : real rate of car deprivation in 2008 is 19.5 %, while simulated one with model estimated on 1994 data is 21.6 %.

Car ownership distribution in 2008 is not correctly recreated by the model. In particular, multi-ownership is under-estimated. However, the model forecasts a part of car ownership increase, especially decrease in car deprivation. Therefore, car ownership growth comes for a part from socio-demographic factors, land use or changes in access to mobility (notably generalization of driving licenses), but a large amount of car ownership growth still remains unexplained, resulting from unobserved heterogeneity, and illustrating changes in lifestyles and behaviours.

Apart from endogeneity problems, changes in car ownership result from both changing levels of explanatory variables and changing effects in sense and/or strength. The following changes are observed in respective groups of explanatory variables :

- Housing and location effects (perpetuation of urban sprawl) : growing rate of households living in suburban areas, increasing proportion of houses.

- Socio-demographic effects : decreasing rate of craftsmen, farmers and traders, increasing qualifications, increasing part of the 50-65 age group caused by population ageing.
- Mobility potential : larger access to driving licenses, less seasonal transit tickets.
- Household structure : smaller households, more singles, less children and students at home, head of household's feminization, head of household's growing work participation and slightly increasing partner's work participation.

If effects were constant, changes in housing, socio-demographic and mobility potential should induce car ownership growth, while changes in household structure and work participation, would have more uncertain effects. Thus, most changes in explanatory variables contribute to growing car ownership. Differences between forecast distribution and real one in 2008 may be caused by changes in senses and/or intensity of significant effects. In order to reveal such shifts, we tried to estimate new parameter values on National Transport Survey from 2008 with the same model specification. Of course, using the same specification will omit some relevant factors in 2008, but will help analysing apparent changes in the relative strengths of effects.

Parameter name	Parameter value 1994	Parameter value 2008
C <sub>1</sub>	-1,16	-0,79
C <sub>2</sub>	-1,89	-1,64
B0_AIREPARIS	0,44	1,27
B0_ICENTRE	0,45	0,9
B0_IQ1	1,28	0,91
B0_IQ2	0,95	0,44
B1_ACTPR	0,53	0,83
B1_CONJOINT	0,55	0,26
B1_IQ3	-0,6	-0,13
B1_LICENSE	3,1	3,1
B1_NETUD	0,39	0,67
B2_ACTPR	0,56	0,9
B2_CONJOINT	0,5	0,29
B2_IQ3	-0,57	-0,08
B2_LICENSES	3,38	3,43
B2_NETUD	0,39	0,72

From this exercise we can separate effects of stable, increasing or decreasing influence :

- Effects becoming less determinant: household income and head of household's partner's existence.
- Increasingly influent effects : number of seasonal transit tickets, location in the urban area of Paris, or in central cities, head of household's work participation, students in household.

Signs of effects didn't change, but socio-economic factors such as income became less influent, and on the contrary density or location effects are becoming more determining. Income's decreasing influence on car ownership level could be questioned as income setting is quite unsophisticated in model

specification. However, further evidence of income's decreasing influence on car ownership level is given by another model specification, where income was entered as a quantitative variable: 1 for 1<sup>st</sup> decile, 2 for 2<sup>nd</sup> decile, and so on. Alternative-specific parameter values for income were specified in alternatives 1 and 2. With such specification, income-related parameter values are respectively decreasing from 0.155 to 0.118 for alternative 1, and 0.170 to 0.132 for alternative 2. Therefore, income's decreasing influence on car ownership level doesn't look dependent from model specification. This could be partly explained by the development of second-hand market, making car more affordable for low-income workers and middle-class households. Indeed, descriptive statistics reveal that the retired, workmen and employees have been on the way to catching up executives among the fifteen past years. But it could also be interpreted through the general dynamics of good diffusion, in which car market would have gradually come to saturation during the last fifteen years.

Consistently with previous results from descriptive statistics, car ownership dynamics are becoming more dependent on density effects and correlated alternative transport supply. Work participation is also more influent than before, may be because of increasing commuting distances. But observable trends could also be evidence of growing rationality, car ownership becoming more dependent on regular mobility needs deriving from professional situation. Students in household also have a greater influence on car ownership, a fact that could be explained by growing need for autonomy.

Eventually, these trends could be synthetized by the double assumption of individualization and rationalization, giving a plausible global framework for behavioural analysis of mobility-related behaviours, including car ownership. Interestingly, the idea of "rationalization" implies that "rationality" (understood as the idea that an individual is searching for its optimal interest, whatever the way "interest" is defined), is not constant over time. Micro-economic modelling, based on the assumption of individual "rationality", is likely to produce better forecasts if individuals (or households) are indeed growingly "rational". Reasons of growing rationality might be interesting to find, and could be related to economic factors (economic unsafety and growing competition) and/or deep underlying psychological attitudes to be revealed by surveys on stated preferences.

By combining changing values in explanatory variables and changing senses and/or strengths of effects, we can expect influences towards increasing or decreasing car ownership. Of course, changing values of explanatory variables with stable effects, such as increasing driving license holding, or the growing part of households living in houses, don't bring wrong forecasts. Therefore, under-estimation of multi-ownership in 2008 doesn't result, for instance, from growing diffusion of driving license holding through successive cohorts. By elimination, we can infer that under-estimation of car ownership in 2008 mainly comes from changes in alternative-specific constants, and also from two increasingly influent effects on the level of car ownership : head of household's work participation, and students living at home.

Changing alternative-specific constants illustrate increased probabilities of car ownership and multi-ownership, everything else being equal, including for a given level of driving license holding. It means that under-estimated multi-ownership is not caused by cohort effects in driving license holding that are taken into account by the model but may come from hidden factors, either psychological or economic.

The increasing influence of work participation doesn't come out from a simple cross-analysis : on the contrary, car ownership differential caused by work participation is decreasing, as car ownership growth is stronger among retired people. But this is in fact a generational adjustment caused by growing access to driving license among retired people.

Subsample	1994			2008		
	E=0	E=1	E>=2	E=0	E=1	E>=2
<b>1 LICENSE</b>						
Inactive	0.293	0.692	0.015	0.264	0.696	0.04
Active	0.176	0.787	0.037	0.124	0.813	0.063
<b>2 LICENSES</b>						
Inactive	0.051	0.615	0.334	0.031	0.532	0.437
Active	0.024	0.473	0.503	0.016	0.309	0.675

Table 7 : Distribution of household car ownership in 1994 and 2008, depending on number of driving licenses and head of household's work participation

Reading guide : in 1994, among single-licensed households where head of household is active, the rate of single-ownership is 78.7 %.

Among multi-licensed households, difference in car ownership between active and inactive people is effectively increasing, as multi-ownership is increasing faster among working people, showing that car autonomy could be growingly perceived as necessary among working people. It may also be partly explained by changes in household structure. Indeed, head of household's work participation is influent on partner's existence and work participation. In 1994, among inactive heads of household, only 50.7 % had a partner, among which 81.5 % were inactive. On the contrary, 74 % of active heads of household had a companion, among which only 34 % were inactive. This difference notably comes from a greater part of singles among retired people and close ages of partners leading to similar work participation status, but also depends on homogamous social schemes.

Between 1994 and 2008, the following changes happened in household structure :

- More people are living alone.
- Among couples, homogamous schemes are getting stronger : among working heads of household, partner's work participation has increased from 66 to 69 %, and among inactive heads of household, partner's work participation has decreased from 18.5 to 11.5 %.

The reinforcement of homogamous schemes contributes to the stronger influence of head of household's work participation, as the residual effect of partner's work participation is declining.

## CONCLUSION

Given observable results, it would be useful to implement temporal parameterization of car ownership models. Past growth was notably related to driving license dissemination among newly retired people, and need for individual autonomy among couples living in low-density areas, in the context of urban sprawl and women's increasing work participation. Early trends underline eroding socio-economic gaps and new "car-restraining" behaviours among young adults living in metropolitan areas. For instance, in the urban area of Paris, the rate of young adults (aged less than 29) living in "car-deprived" households has increased from 16.8 to 31.1 % between 1994 and 2008. At least in France, location and density effects are becoming more determinant than socio-economic factors in explaining car ownership behaviours.

Individualization and rationalization could provide a general framework to describe fundamental changes in car ownership behaviours today. However, future impacts of a presumably generational effect among young adults are still doubtful. Are new behaviours various signs of a changing relationship to car, becoming a complementary and substitutable way to ensure mobility, in a logic of growing rationality ? Or shall it be rather interpreted as a temporary effect of precarious conditions and expensive urban life, leading young people to delay some investment decisions ? Moreover, the question remains of how interpreting changing parameter values, leading to wrong forecasts of models estimated from cross-sectional data. Straightly interpreting them as changing behaviours or attitudes towards car might probably be too conclusive. In fact, there might probably be three distinct dimensions to consider:

- The general dynamics of car market might have come to saturation, with decreasing marginal utility of supplementary cars among the well-off, and an adjustment process among low-income groups. Homogenization also results from driving license diffusion.
- This adjustment process is facilitated by the second-hand market making car more affordable. Price dynamics, being temporal-dependent, couldn't be pictured in our modelling approach based on cross-sectional demand effects.
- Eventually, behavioural change: psychological latent factors such as desire for autonomy could explain increased multi-ownership, or, on the contrary, rationalization could explain car restriction and greater substitutability between alternatives among the young urban.

One difficulty is about making distinction between psychological factors and economic *inputs* as they could be correlated. For instance, the adjustment process might either be interpreted through psychological attitudes (rationality, need...) or price dynamics. Building dynamic models from panel data including variables about price dynamics but also opinions and attitudes would probably help separating strictly economic from psychological dimensions in changing behaviours.

## BIBLIOGRAPHY

- Bussière, Y., Armoogum, J. et Madre, J.L (1996) Vers la saturation : une approche démographique de l'équipement des ménages en automobile dans trois régions urbaines, *Population*, **51** (4-5) 955-977.
- Button, K., Pearman, A.D, and Fowkes A.S. (1982) *Car ownership modelling and forecasting*, Gower Publishing Company Ltd, London.
- Collet, R., Bourcq, E., Madre, J-L., and Hivert, L. (2010) Long-term automobile ownership and mileage trends by income class in France – 1975-2008, *Proceedings of the 12<sup>th</sup> World Conference on Transport Research*.
- Dargay, J., Madre, J.L., and Berri, A. (2000) Car ownership dynamics seen through the follow-up of cohorts : a comparison of France and the UK, *Transportation Research Record*, **1733** 31-38.
- Dargay, J., and Vythoulkas, C. (1999) Estimation of a dynamic car ownership model : a pseudo-panel approach, *Journal of Transport Economics and Policy*, **33** (3) 287-302.
- Dargay, J. (2001) The effect of income on car ownership : evidence of asymmetry, *Transportation Research Part A : Policy and Practice*, **35** (9) 807-821.
- Gallez, C. (1994) *Modèles de projection à long terme de la structure du parc et du marché de l'automobile*, Thèse de doctorat en sciences économiques.
- Gallez, C. (1994) Identifying the long-term dynamics of car ownership : a demographic approach, *Transport Reviews*, **14** (1) 83-102.
- Janson, J-O. (1989) Car demand modelling and forecasting. A new approach, *Journal of Transport Economics and Policy*, **XXIII** (2) 125-140.
- Jong, J., Fox, J., Peter, M., Daly, A.J., Smith, R. (2004) A comparison of car ownership models, *Transport Reviews*, **24** 379-408.
- Kitamura, R. (2009) A dynamic model system of household car ownership, trip generation, and modal split : model development and simulation experiment, *Transportation*, **36** 711-732.
- Madre, J-L. (1990) Long-term forecasting of car ownership and car use, in Jones, P., Ed., *Developments in dynamic and activity-based approaches to travel analysis*, Oxford Studies in Transport, Avebury, Aldershot.
- Madre, J.L. and Gardes, F. (2005) Why long-term elasticities differ from cross-sectional elasticities ?, *Proceedings of the 33<sup>rd</sup> European Transport Conference*.
- Pochet, P. (2003) Mobilité et accès à la voiture chez les personnes âgées : évolutions actuelles et enjeux, *Recherche Transports Sécurité*, **79-80**.
- Whelan, G. (2007) Modelling car ownership in Great Britain, *Transportation Research Part A : Policy and Practice*, **41** (3) 205-219.