7. Scenarios for developing a new car ownership model

7.1 Preferred model

A model that can handle all the above questions does not exist and is not likely to be developed in the very near future. We compared the model types with the requirements and our opinion is that the model type that can fulfill most requirements is the dynamic car ownership model (either a dynamic vehicle transactions model or a dynamic Markov-type panel model) with type choice conditional on vehicle transactions. This we regard as the preferred model, both because it represents the state-of-the-art in car ownership modeling and because it scores best when compared to the requirements described in chapter 6. This is worked out below.

Passenger car model

The proposed model is a passenger car model. Maybe vans can be included; we shall return to this issue below.

Number of cars per household

In a dynamic vehicle transaction model, such as the DVTM or the model for California of Brownstone et al., the number of cars per household is predicted on the basis of current car ownership of the household. A duration model predicts the time (e.g. in months) until the next vehicle transaction and the type of transaction (e.g. replacement, disposal, adding a car). In application the model is used in discrete time steps, for instance a year. For every household that does not transact in this year, the vehicle ownership situation of year t+1 will be equal to what it was in year t. For other households there will be a transaction and, if this involves replacing a car or adding a car, the conditional type choice model will be used to get new type choice probabilities. In this way the duration model can be used step by step, each time predicting transactions on the basis of the car ownership situation of the previous year. Vehicle scrappage transactions could also be integrated in such a model: with the passage of time, vehicles age and scrappage (other than accident-related scrappage) becomes more likely.

Alternatively, a panel model might be used to predict the car transactions from year to year. A panel model of household transitions between the different car ownership situations (Markov model) can serve the same purpose as a duration model. The alternatives in such a discrete choice model are the transitions that are possible from
year \( t \) to year \( t+1 \), such as from one car to two cars, or from one specific car to another car (replacement), or scrappage.

Both for a duration model and a panel model of vehicle transactions, short run predictions (up to five years ahead) might be done without updating the population in the sample used. For medium and long run forecasts, the population needs to be updated.

The most sophisticated method for this is dynamic micro-simulation of ‘birth’ and ‘death’ of households and changes within households. This can be done by using duration models for the time that a household spends in a certain state (household lifecycle stages). Such duration models for household demographic and socio-economic changes can be combined in a consistent way with duration models for vehicle holdings, as has been done in the Californian car ownership project. Because duration models predict changes in continuous time, they can give all intermediate time steps. If one uses Markov models for car ownership changes, then the time steps need to be determined by the researcher (e.g. years, five-year periods). As soon as the time interval has been chosen, the Markov model cannot predict for shorter time intervals. The micro-simulation of household change needs inputs from the medium and long term CPB scenarios (e.g. on income and population over time), but also additional restrictions to remain consistent with the CPB scenarios.

A simpler method is to use the model for a specific sample recursively and afterwards reweigh the sample to reflect the changes in the household distribution between the present and the situation 10, 15 or 20 years ahead (based on information from the CPB scenarios). The latter method avoids the spurious accuracy and complication of modelling the generation and termination of households, but loses the dynamic aspect of aging of the households themselves.

This model will produce one of the vehicle type distinctions mentioned in the previous chapter: first and second car in the household. Including third cars would complicate the model substantially, since the number of transaction types would increase considerably.

**Vehicle types**

Conditional on specific vehicle transactions, the discrete vehicle type choice model will be applied. We recommend using as smallest observable unit (choice alternative) the brand-model-vintage combination, e.g. Opel Astra, 1.8 diesel of 1999. A similar distinction was used in the DVTM, which had about 1000 alternatives. Most of the vehicle type choice models in the literature also use brand-model-(vintage) alternatives instead of more aggregated vehicle categories. This distinction is not proposed here because we want to predict by brand (interesting for General Motors, not so much for government), but because:

- This specification is clear, for the researchers but especially for the consumers: this is the kind of vehicle alternative that we can refer to when interviewing a respondent. Moreover, this is the kind of choice alternative that many consumers will have in mind when deciding on the type of vehicle.
This specification can be aggregated in many different ways to yield most of the required outcomes:

- Fuel type (diesel versus petrol, for LPG, which is built in after producing the car, an extra distinction needs to be added, which can be done in a new sample survey)
- Weight
- Vintage
- New or second hand
- Type code of industry
- Year that car type was officially approved
- Energy consumption label, safety label.

Also average emission rates and fuel consumption for the brand-model-vintage combination can be used to give outcomes on these variables.

Specific BPM-exempt accessories, safety equipment and equipment affecting energy consumption usually vary even within brand-model-vintage combinations. To get the distribution in for these attributes, a post-processing of the outcomes of the car type model will probably be necessary using exogenous fractions, which depend on scenario assumptions.

Vans can be included as a number of special brand-model-vintage combinations, if data on the household possession of these would be available.

**Business cars**

The separate identification of business cars is specified by the client as being of the highest priority. Moreover, the factors influencing the ownership of business cars, as well as the types of cars used for this purpose, are different from those affecting cars privately owned, so that accurate and policy-sensitive forecasting of numbers and types of cars can be improved by making this distinction. In the Sydney discrete choice system, as in FACTS, business car ownership is predicted first and independently of private car ownership, which is then predicted conditional on the business car holdings of the household. However, in the UK New Cars model of Page et al., the total number of cars is predicted first and the manner of car holding (business or private) is predicted conditional on this. The choice between the two approaches depends on the results obtained from empirical models of the current situation in The Netherlands.

Either way, business cars can be treated in the same way as private cars with respect to the choice of type, dependence on socio-economic factors etc.. This could also include vans owned by companies, but being used by households for a mixture of private and business purposes.

As in the case of the Sydney and the UK model for business cars, the data to be used can be a household survey of (changes in) car ownership. It is possible to develop a model for company cars based on the production structure of the economy, as happened in FACTS, but operating, policy-sensitive models can also be constructed.
on the same data as for private cars. In the Sydney model for example, the variables influencing company car ownership are: household income, age and gender of the head of the household, cost of parking and composition of the household. Occupation and sector were tested, but not significant.

*Forecasting horizon*

This model, can be applied both for the short and long term, but because it starts from the present situation, it is more suited for short run forecasts, and the further one gets away from the present the more synthetic and less reliable the outcomes will get. Given the complexity of refreshing the population sample for long run forecasts, a simpler approach for the total car ownership in the long run (e.g. on the basis of a cohort model or a pseudo-panel model) might be considered. This could then provide the control total for the LMS. However, it is not desirable to have different car ownership forecasts that are not consistent with each other over time. Therefore we prefer to generate all forecasts from the same dynamic transactions (or panel transitions) model.

It may be possible to find ways of eliminating some of the intermediate steps without losing consistency with the short-term forecasting model. For example, it may be possible to work with longer steps. Alternatively, it may be possible to ‘jump’ several years and then allow the model to find a new stable situation in a short period (e.g. 5 years) before the forecast horizon. Elimination of intermediate steps is not primarily desirable to reduce the run time of the model, more it is to avoid dependence of the results on possibly minor features that might multiply over a long series of applications of the model.

*Car use*

A car use regression equation, for a limited number of car types can be added to the above model, as has been done in the DVTM and the Californian model (with fixed and variable car cost). These equations can be estimated by instrumental variables.

Alternatively, depending on the accuracy required, the LMS can be used. The advantage of using the LMS is that far more detail of the nature of car use is obtained (length of journeys, location of emissions (including noise), purpose of journeys and hence potential for reimbursement etc.). The disadvantage is that it is more time-consuming. However, the main time consumption in the LMS is in the calculation of capacity restraint on the road network. While this is burdensome, it also makes the forecasts more accurate and a car use model that omitted this equilibration would be likely to be inaccurate for large changes in car cost. When capacity ‘feed-back’ can be ignored, the LMS can be run quite quickly.

*Household types*

In models on disaggregate data such as the above models, many household types can be distinguished. It seems probable that the distinctions that are required will be contained within the quite rich data specification of the OVG sample and therefore that the new models can be connected to the LMS prototypical sampling.
For input into the LMS travel choice models, a subset of these distinctions is required and problems of compatibility do not arise.

Financial outputs

Different income groups and the way these are affected can be distinguished in such models. The impacts on the government revenues can also be calculated by summing over all households and vehicle types.

New car types

If in the future new car types become available (and older cars disappear), the attribute values of the new types need to be described in an exogenous car type file, which will be the choice set for application of the car type model in a future year (can be done for various scenarios). The existing car type model coefficients will give the way people trade-off these attributes. Hybrid and electric cars can be included in the forecasts, but only if attributes that are especially relevant for such cars have been included in the model (range, re-charging options). This is only possible if the model is (partly) based on stated preference data, with such alternatives and attributes. SP data were used in the model of Brownstone et al. and in Hensher and Greene (2000).

Demand-supply equilibration

Neither the Dutch DVTM or the dynamic car ownership model for California contains a demand-supply equilibrium mechanism. For new cars in The Netherlands, such a mechanism is clearly not needed, since the new car market is an international market in which the Dutch demand fluctuations will not have an important effect. However, on the second-hand market, the assumption that supply follows demand for The Netherlands might be too strong (although within the EU it has become rather easy to import and export used cars). Supply of second-hand cars can be derived from the car fleet by vintage of previous years in combination with scrappage equations. This can be confronted with second-hand car demand, as happens in the FACTS RAS-procedure, with a feedback to demand in case of disequilibrium. We recommend that the new model will have a RAS-like demand-supply equilibrium mechanism.

Spatial component

It might be possible to find significant location-specific variables, such as degree of urbanisation when estimating the above models. However an allocation of cars to 1308 zones, as happens in the LMS, would be too much for a dynamic car ownership model that should also work with more than 1000 vehicle types. This could make the model untraceable, slow and less stable. We recommend that the allocation to zones will take place after applying the car ownership model, e.g. using procedures as are in the LMS now.

An important theoretical issue is how accessibility impact on car ownership. When this impact has been investigated, it has been found to be significant (e.g. in Zuidvleugel, Stockholm, Sydney and Rich’s Denmark model), but it is not included in many large-scale model systems, such as the LMS. The inclusion of such an effect is useful in allowing a wide range of variables, including variable car costs, to be
included in the car ownership model in a way that is internally consistent and satisfies economic theory. A potential problem is that it is not certain the accessibility influences car ownership: it may be the case that car owning households choose to live in areas where car travel is relatively superior, while non-car-owning households choose to live in areas well served by public transport. Rich’s model may give some insight into this issue. There is also considerable complexity in this approach, but the alternative is to give substantial weight to location, e.g. described by urbanisation level, which gives no ‘handle’ for policy or insight into behaviour.

Policy measures

The impacts of variabilisation, subsidies and fuel tax policies could be tested with the preferred model as described. In a model structure such as the DVTM, changes in the variable car cost will have an influence on:

- The timing of the vehicle transactions (e.g. postponing an acquisition or accelerating a replacement)
- The vehicle type choice (e.g. a higher probability of choosing a fuel-efficient car in case of higher variable car cost)
- Vehicle use (e.g. a reduction in annual kilometrage in case of higher variable car cost).

A change in the fixed car cost will have an impact on the same choices. In order to get reliable results for large changes in variable and/or fixed car cost – larger than have been observed in the RP data – SP questions about reactions to large cost changes need to be asked. A paper reporting on such SP experiments is Rosenberg et al., 1997.

Changes in the taxation rules can have an impact on the choice between company/lease cars and privately owned cars. The representation of lease cars is required as a first priority for the new model. It appears to be of little interest to predict how businesses finance their cars (should a firm purchase the cars or lease the cars?), the lease car issue that is of interest to the – government- users of a car ownership model relates to private car ownership: ownership, type choice and use of lease cars by households may react differently to policy measures than that of privately owned cars.

The choice of leasing or outright ownership can be represented in the model in an appropriate structure with the choice of the type of car itself (possibly including vans). The advantage of leasing in terms of fiscal benefit will depend on annual kilometrage, percentage of business use and income, all of which are variables that are required in any case. A difficulty may arise in obtaining sufficient detail about other tax deductions available to the individual, which may well influence his or her benefit from leasing.

To include the impact of accelerated scrappage subsidies in the model it is necessary to base the scrappage transactions decisions on SP data (as in de Jong et al, 2001).
7.2 Fall-back option 1

Developing the above model would certainly require a major effort and could take several years. A fall-back option, which would involve considerably less effort, and could also be considered for a model to be used while the above-mentioned system would be under construction, is a further extension and adaptation of FACTS.

It may be possible to add a policy-sensitive car use equation (also with a random component) to FACTS, to replace the random procedure in FACTS 3.0. This would make car use more sensitive to policy changes. The possibility of linking to the LMS, with the advantages and disadvantages discussed above, remains.

Also, it might be considered to replace the FACTS car type choice component by a disaggregate type choice model, and let the existing FACTS structure explain the number of cars owned. FACTS has been quite successful for this and less so for type choice (but please note that the possibilities of FACTS for future runs for large cost changes will be rather limited). This type choice model then could give more detailed outcomes than the present 18 car types and cost and other influences on type choice would be separated. This is in our view the most urgent improvement needed for FACTS. Most of the resources required for developing the preferred model (especially collection of new data, both RP and SP; estimation of the type choice model) are also needed for the type choice model in fall-back option 1. In other words, the most expensive part of the preferred option is also included in this fall-back option. This makes fall-back option 1 easier and somewhat faster to develop and less sophisticated than the preferred model, but not much cheaper than the preferred option (see section 7.4).

The main difference between the preferred model and fall-back option 1 is that the former contains a disaggregate dynamic component for the number of cars (the duration or Markov models) and the latter does not: the number of cars follows from the FACTS mechanisms. Both models will have a new vehicle type choice model. The preferred model will be more suitable than FACTS for giving the short run (1-5 years) impact over time and will also be able to produce more policy sensitive forecasts of the number of cars than FACTS.

7.3 Fall-back option 2

Both the preferred option and option 2 require the acquisition of new data on choice of vehicle brand-model-vintage combinations, which would lead to considerable survey cost. The basic idea of fall-back option 2 is to start from presently available data, notably the OVG, and construct new models that can be based on these data sources and nevertheless will lead to an improvement of FACTS in the main problem areas. The current OVG questionnaire contains the following questions on the car, which the main user of the car should answer:

- Fuel type (LPG, diesel, petrol)
- Approximate annual kilometrage
- Vintage (year of interview, 1 year before, 2 years before, …11 year before, older)
- Weight in kg
• Private or company ownership
• Lease-car or not.

We would prefer a vehicle type choice model with brand-model-vintage as choice alternatives, for reasons described above. However, the OVG does contain all the distinctions that are of the highest priority (see chapter 6) for a new car ownership model. It will be difficult if not impossible to add the ‘priority 2’ distinctions when using OVG, because these rely on finer segmentations than the OVG can give.

Advantages of using the OVG are:

• It contains all the information on persons and households that is used in the LMS.
• It has a very large sample size and can also be used as repeated cross section (or even pseudo-panel); the questions on the car have been the same for several years now.
• It already exists: no new data collection needed.

Disadvantages of using the OVG (in comparison to new data collection, tailored to generate the data required for the preferred model) are:
• The choice alternatives in reality are more detailed; variables for the number of elementary alternatives will have to be added to the type choice model and could become very important; variables for brand-royalty and similar effects cannot be estimated.
• The OVG gives the number and type of cars at a certain point in time, not the car type choice conditional on a vehicle transaction (though by selecting the cars of the same year as the survey, type choice for new cars can be approximated). We do not know when the car was purchased. So we can only relate the type choice, which took place at an unknown previous point in time, with the current household and person attributes. It also does not contain attributes of the previous car, which can be helpful in predicting the type of the next car.
• The OVG does not give information on transactions, so it cannot be used to estimate a disaggregate duration or Markov panel model for vehicle transactions.
• The OVG does not contain SP data, so it cannot be used to get SP observations on large cost changes, future car types (including electric cars and hybrids) and accelerated scrappage schemes.

In fall-back option 2, the OVG would be used to estimate a fuel type-weight-vintage-ownership choice model. FACTS would be retained to produce the number of cars in the household. For car use, FACTS would be enhanced or the LMS would be used.

7.4 Development of the preferred model and fall-back models

Data collection

In the Netherlands there are no recent data, which can link number of vehicles owned and brand-model-vintage of the vehicle to household types. The OVG does not
contain brand and model and the PAP has been abandoned. New data would need to be collected both for the preferred and the fall-back model 1².

The DVTM was estimated on a new mail-back questionnaire, developed for this project, which was distributed to members of the existing panel of car drivers of the Consumentenbond and ANWB. The original duration, type choice, car use and energy use models were estimated on the 1992 wave which contained retrospective questions on the last car transaction (about 4,000 car drivers). Later, the duration models and type choice were re-estimated using information on actual transactions (not retrospectively) from three waves of panel data: 1992, 1993 and 1994. There were no SP experiments in this data set.

The data used by Brownstone et al. (2000) came from two waves of a panel survey in California. First, respondents were recruited by telephone (also to customise the SP), then they had to complete a mail-back SP interview. The RP data on actual vehicle purchases came from comparing wave 1 and wave 2, which were 15 months apart, for the same respondents. Wave 1 had almost 5,000 households and wave 2 almost 3,000.

Sample sizes of several thousands respondents are needed to estimate the type choice models. For the transaction models similar sample sizes are required. This makes SP interviews with an interviewer and a laptop too costly. Two waves is better than retrospective questions, because the long run memory of respondents might not always be very reliable: the question is whether people accurately recall their last car transaction. Given that expensive cars need to be included as well, the Consumentenbond/ANWB car panel is not sufficient.

One option would be to carry out a new survey that starts by recruiting respondents (by phone). Two waves (with net sample sizes of 5,000 and 3,000 respectively) mail-back interviews or computer-assisted telephone interviews (CATI) will then cost 100,00-150,00 Euro in total. This includes SP experiments on new car types (electric, hybrid, other fuels), large cost changes and on accelerated scrappage schemes, but not much can be saved by only asking RP questions. If the first wave contains retrospective data on the last vehicle transaction, a preliminary duration model of car ownership can be estimated on this data, as was done in the DVTM project, which can later be replaced by a duration model on observed transactions between two waves of data.

Another option would be to use the Consumentenbond car panel again, with an additional survey among owners of expensive cars (including many typical company and lease-cars). The Consumentenbond car panel still exists and AVV is considering using it for a new annual questionnaire. The fieldwork cost per wave for the questionnaire that was used for the DVTM in 1992-1994 was about 18,000 Euro. The questionnaire for the new model will probably be longer (SP experiments need to be added). Whether this option is attractive depends to a large extent on the possibilities for selecting owners of expensive cars. If the car registration data can be used for this, this selection can be done at low cost. However, if the registration data could not be

² Unless the CBS could be convinced to add vehicle brand and model to the OVG questionnaire.
used, owners of expensive cars would have to be selected by methods like randomly calling households (possibly in wealthier areas) and asking for the car brand-model combination. This would be an inefficient procedure and would reduce the savings that could be expected from using the existing Consumentenbond car panel and a relatively small new survey instead of a full new survey. The field work cost estimate, that follows later in this chapter, is only an indication, roughly based on the cost in 1992-1994. We have not asked the Consumentenbond for a new cost estimate.

The optimal time between the two waves in the Netherlands depends on the amount of changes in the household location and in the car ownership status. Both of these change probably at a lower rate than in the US and therefore the optimal interval will be between 18 and 24 months.

**Model structure and estimation**

The preferred model system consists of:

- Duration models for the time between vehicle transactions (and the type of transaction: disposal, replacement, acquisition, also scrappage) to explain the total number of cars. An alternative option for this would be a Markov-type panel model. Duration models predict the time until the next vehicle transaction in continuous time, which can be simulated by looking at any possible time interval. Markov models use a pre-determined time interval and predict changes in vehicle holdings between two such time periods.

- Vehicle type choice models for the choice of a brand-model-vintage alternative for all vehicle transactions that involve purchasing a(nother) car. These choice alternatives can be aggregated to get the composition of the fleet in terms of most of the required distinctions. Some less important distinctions need to be made by a post-processing procedure.

- Regression equations for the use of every car in the household, measured in terms of annual kilometrage, or through a logsum linkage with the LMS.

- A micro-simulator for ‘birth’ and ‘death’ of households and transitions between households types over time; a simpler but less consistent (in terms of dynamics) alternative would be to reweigh a given sample of households in each time period.

- Possibly a model for the number of business cars (company-owned and lease cars), depending on (sectoral) economic development, which need to be allocated to households. Private car ownership could be made conditional on the outcome of this.

- An allocation procedure to the 1308 LMS zones (also post processing).

Such models have been developed before, particularly as components of the Dutch Dynamic Vehicle Transactions Model (DVTM) and/or the model for the likely penetration of electric and hybrid cars for California.
Vehicle type choice models and panel models for the number of vehicles can be estimated with ALOGIT, developed by Hague Consulting Group. This includes the possibility of mixed logit models, as have been used for the California model for type choice and as would be needed for panel models (to account for the correlation over time of the same respondents). Probit models, as have been used in some Markov models, can also be estimated by ALOGIT (by using simulated maximum Likelihood methods). Duration models can be estimated with the LIMDEP software, as was done in the DVTM project. This also goes for regression equations for annual kilometrage (instrumental variables estimation).

**Indication of the required time and budget**

Below we provide indicative money and time budgets for the preferred model and the fall-back option. However, this is not a project proposal. RAND Europe is interested in doing research on this, and if AVV would request it, we would be happy to submit a proposal with a more detailed cost estimate.

**Preferred model:**
- Survey design: 50,000 Euro
- Fieldwork:
  - full new survey: 125,000 Euro, or
  - Consumentenbond car panel plus new survey among owners of expensive cars using registration data: 85,000 Euro, or
  - Consumentenbond car panel plus new survey among owners of expensive cars selected without using registration data: 100,000 Euro
- Model specification and estimation: 125,000 Euro
- Testing and implementation: 60,000 Euro
- TOTAL: 320,000-360,000 Euro.

**Fall-back option 1:**
- Survey design: 40,000 Euro
- Fieldwork:
  - full new survey: 100,000 Euro, or
  - Consumentenbond car panel plus new survey among owners of expensive cars using registration data: 75,000 Euro, or
  - Consumentenbond car panel plus new survey among owners of expensive cars selected without using registration data: 90,000 Euro
- Model specification and estimation: 50,000 Euro
- Changing FACTS and integration of models: 40,000 Euro
- Testing and implementation: 40,000 Euro
- TOTAL: 245,000-270,000 Euro.

**Fall-back option 2:**
- Acquiring OVG for several recent years: PM
- Model specification and estimation: 50,000 Euro
- Changing FACTS and integration of models: 40,000 Euro
- Testing and implementation: 40,000 Euro
- TOTAL: 130,000 Euro.
Time required for data collection and model development:
- Preferred model: 3 years
- Fall-back model 1: 2.5 years.
- Fall-back model 2: 1 year.