A joint model for vehicle type and fuel type

Stephane Hess
Institute for Transport Studies
University of Leeds
s.hess@its.leeds.ac.uk

Mark Fowler
Resource Systems Group
mfowler@rsginc.com

Thomas Adler
Resource Systems Group
tadler@rsginc.com

Aniss Bahreinian
California Energy Commission
ABahrein@energy.state.ca.us

Abstract

In the face of growing concerns about greenhouse gas emissions, there is increasing interest in forecasting the likely demand for alternative fuel vehicles. This paper presents an analysis carried out on stated choice data collected in California, looking at respondent's preferences in a joint vehicle type choice and fuel type choice experiment. Our study recognises the fact that this choice process potentially involves high correlations that an analyst may not be able to adequately represent in the modelled utility components. Importantly, we further hypothesise that the standard Nested Logit model is incapable of capturing the full extent of correlation patterns in such a multi-dimensional choice process, and that a Cross-Nested Logit structure may be more appropriate. Our empirical analysis and a brief forecasting exercise produce evidence to support these suspicions. The findings from this paper are not just of interest in the context of the demand for alternative fuel vehicles but are also relevant for the analysis of multi-dimensional choice processes in general.

Introduction

There is increasing interest in the potential demand for alternative fuel vehicles, given not only growing environmental concerns, but also recent volatility in oil prices. A number of approaches for alternative fuels have been put forward, including using a mix between ethanol and traditional fuel, a greater reliance on natural gas, and various types of partially or fully electric vehicles. The preferences for different types of fuel are difficult to predict, not least because of the strong relationship between fuel type and other attributes such as performance, annual costs as well as incentives (e.g. tax breaks). At the same time, there is a very strong link between fuel type and vehicle type, with certain types of fuel only really being appropriate for specific vehicle types.

With the growing focus on these vehicles as a possible way of reducing greenhouse gas emissions, it should come as no surprise that there has long been an interest in modelling the potential consumer response to the introduction of such vehicles. Examples include Bunch et al. (1993), Train (1993), Golob et al. (1995), Kavalece (1996), Tomkins et al. (1998), Hensher & Greene (2001), Greene (2001), Batley and Toner (2003), Batley et al. (2004), Adler et al. (2004), and Spissu et al. (2009) to name but a few.

In this paper, we discuss work based on the 2008-09 California Vehicle Survey (CVS), aimed at providing input data for the California (light-duty) Conventional and Alternative Fuel Response Simulator (CALCARS) model. The 2008-09 CVS collected both stated and revealed preference data for vehicle ownership and vehicle choice to forecast the penetration and use of both conventional and alternative fuel vehicles.

In the present paper, we focus on the stated choice survey component of this work. This involved the design of a highly complex survey tool, which included seven fuel types, fifteen vehicle types, and up to eleven level-of-service attributes, such as cost, fuel consumption, fuel availability, refuelling time and acceleration. To reduce the survey complexity, each choice situation made use of only four alternatives, where this included a reference vehicle and three other vehicles assigned on the basis of a weighting approach. An internet-based survey was used to collect the data, enabling the collection of a very large sample.

The main contribution of the work described here comes in the use of a Cross-Nested Logit (CNL) structure. Earlier parts of the empirical work revealed the existence of significant levels of correlation between alternatives sharing the same fuel type as well as between alternatives sharing the same vehicle type. As we discuss in detail in the methodology section, the use of multi-level Nested Logit (NL) structures was not appropriate in this context, and the use of the CNL model was shown to lead to significant gains in model performance as well as the realism of forecasts.

The remainder of this paper is organised as follows. We first describe the survey work carried out for this analysis, before discussing modelling methodology. This is followed by the presentation of the empirical results and a brief forecasting example. Finally, we present the conclusions of the work.

Survey work

Survey data were collected using a two-phase, multi-method approach. The first phase involved a revealed preference (RP) recruitment survey that was conducted over the telephone. The second phase included the stated choice survey with eight vehicle choice exercises.

In the RP survey, respondents were asked to indicate the type of vehicle they are most likely to purchase next for their household; including information about the vehicle type, fuel type, expected fuel efficiency, purchase price, vehicle age, and estimated number of miles the vehicle would be driven annually.

After completing the revealed preference (RP) survey over the telephone, the respondents were given the option of completing the stated choice (SC) survey either via a mail-back paper questionnaire or via an online questionnaire. In both cases, the survey was tailored to the specific individual, with data from the RP surveys being used to construct a set of eight stated choice exercises for the SC survey.

Each stated choice exercise presented respondents with four hypothetical vehicles as alternatives. The first vehicle, or the reference vehicle, was presented as the new or used vehicle the respondent planned to purchase next for their household. The attributes that describe the reference vehicle were consistent with what the respondent reported in the RP survey in terms of vehicle type, fuel

type and age, with the remaining attributes varying across choice sets. The next three alternatives were presented as vehicles of different sizes, fuel types and ages. The four vehicles in each exercise were described by a set of ten to twelve attributes, depending on the fuel type presented. Respondents were asked to select the vehicle they would most prefer to purchase based on the attributes presented in each alternative. The values of each attribute varied according to an experimental design (discussed later), requiring respondents to trade off attributes against each other. **Error! Reference source not found.** presents an example of one of the eight stated choice exercises of a hypothetical respondent.

The first two attributes for each alternative were vehicle type and fuel type. A total of fifteen vehicle types and seven fuel types were selected for the exercises. The vehicle type was fixed to the response given in the RP survey for the reference vehicle. For the remaining three alternatives, vehicle type was drawn from one of the following fifteen types:

1. Subcompact car
2. Compact car
3. Mid-size car
4. Large car
5. Sport car or "two door high performance subcompact car"
6. Small cross-utility car or "small wagons with flexible seating"
7. Small cross-utility SUV
8. Mid-size cross-utility SUV
9. Compact SUV
10. Mid-size SUV
11. Large SUV
12. Compact van
13. Large van
14. Compact pick-up truck
15. Standard pick-up truck

While it was possible any vehicle could be selected for the three alternate vehicles, the selection of those vehicles was done using weighted draws based on the respondent's reference vehicle type. Weighted draws were used because it is expected that respondents will have relatively strong preferences for at least a broad category of vehicle (e.g. small or large), and as a result presenting a respondent with a choice between a reference subcompact car and a large van makes little sense. In that situation, vehicle type would dominate the choice process and little or no information could be gained for the sensitivies to other attributes. On the other hand, it was also not seen as appropriate to completely restrict the different combinations of vehicle types presented to a respondent. As a result, a set of weights were developed for each reference vehicle type. With these weights, all vehicle types have a non-zero probability of being included in an exercise, but the probability is higher for those vehicles that are more similar to the reference vehicle type. An especially high weight of approximately 50 percent was used for the reference vehicle type, which ensured that, at least for one pair of alternatives, the relative preference was not influenced by vehicle type. The vehicle types for the three alternative vehicles were drawn without replacement from the list of 15 vehicle types, meaning that, while the reference vehicle was allowed to repeat in one other alternative, allowing respondents to trade off attributes other than vehicle type, no other vehicle types were allowed to repeat across alternatives within a single exercise.

For the reference vehicle, fuel type was fixed to the respondent's RP response. The remaining fuel types were draw from the following list:

1. Standard Gasoline
2. Flex Fuel/E85
3. Clean Diesel
4. Compressed Natural Gas
5. Hybrid-electric
6. Plug-in Hybrid-electric
7. Full Electric

The selection of which fuel type to show for the three alternate vehicles was done entirely randomly, i.e. not using any weights, thus guaranteeing that all possible combinations were represented roughly evenly. As with vehicle types, fuel types were drawn without replacement, meaning that while the reference vehicle fuel type was allowed to repeat in one of the three alternate vehicles, allowing respondents to trade off attributes other than fuel type, no other fuel types were allowed to repeat across alternatives within a single exercise.

The remaining vehicle attributes were dependent on the vehicle and fuel type. While values for vehicle type and fuel type were selected using weighted and random draws as described above, the values for the remaining attributes varied according to an orthogonal experimental design. The orthogonal design is described in more detail later on.

Many of the vehicle attributes vary around a base value. In the case of purchase price, maintenance cost, miles per gallon equivalent, fuel cost per gallon equivalent, and acceleration, a table with base values was used, representing average values for all vehicles of a particular vehicle type, fuel type and vintage.

The remaining attributes included in the survey were as follows:

- The **vehicle age** was automatically set to *new* for plug-in hybrid electric and full electric vehicles, with variations around the reference vehicle age for other vehicles
- The **purchase price** of the vehicle varied around a base value. For the reference vehicle, the base value is the response given in the RP survey. For the three remaining alternatives, the base value was dependent on a "list price" determined from the combination of vehicle type, fuel type, and vintage, where this was adjusted by the ratio between the indicated price of the reference vehicle in the RP survey and the list price for that vehicle, thus accounting for the possibility that a respondent was considering a higher than average or lower than average price for the reference vehicle. Variations across choice sets are then based on the experimental design.
- There were six **purchase incentive** levels shown in the survey, with the exception of gasoline-powered vehicles, where no incentives were used. Incentives included carpool lane access, free parking, tax credits, reduced tolls and reduced purchase price. Variations across choice sets are then based on the experimental design.
- A base **maintenance cost** per mile for each vehicle was assumed based on the vehicle type, fuel type, and vehicle age. The maintenance cost per mile was multiplied by the reported annual VMT to calculate an annual maintenance cost. Variations across choice sets are then based on the experimental design.
- A base value for **miles per gallon equivalent** was assumed based on the vehicle type, vehicle age, and fuel type. Variations across choice sets are then based on the experimental design.
- The **annual fuel cost** was calculated using the fuel cost in gasoline gallon equivalents, which was a design attribute, the vehicle efficiency in miles per gallon equivalent, and the annual miles reported in the RP survey. The variation across choices was based on variations in the fuel cost in gasoline gallon equivalents, as given in the experimental design.

- The **fuel availability**, **refueling time**, and **vehicle range** attributes only applied to full electric and compressed natural gas vehicles, where variations across choice sets for applicable vehicles are then based on the experimental design.
- The **acceleration attribute** was presented as the time it takes to accelerate from zero to 60 miles per hour in seconds. The acceleration of each vehicle was assumed based on the vehicle type, fuel type and vehicle age, and varied according to the experimental design.

The experimental design used for this SC survey was based on an underlying orthogonal design. While several types of designs were considered at the onset, including arguably more advanced efficient designs, it was concluded that, given the complexity of the SC scenarios, an orthogonal design was the most appropriate for this particular application. While efficient designs can be preferable in some situations, the generation of an efficient design requires prior parameter values for all coefficients, as well as a priori decisions in relation to model structure and utility specification, including interactions with socio-demographic variables. While this already causes significant problems in the case of studies looking at the choice between hypothetical options, further problems arise in the present study as it looks at the choice of vehicle types and fuel types. Here, the preferences can be expected to vary across respondents to such an extent – some respondents will strongly prefer compact cars, while others will strongly prefer large SUVs – that it becomes difficult to obtain reliable prior parameter estimates. Additionally, vehicle type and fuel type would have to be included directly in the design, leading to the requirement to generate a very large number of different designs for different combinations of vehicle types and fuel types. This was not necessary with the approach used in this study, where vehicle type and fuel type were added to the design in a second stage, after the generation of the base design.

This base design is an orthogonal design of 144 rows, split into 18 blocks of 8 choices. Orthogonal blocking was used to avoid any correlation between the attributes and the blocks (e.g. avoiding the situation where one respondent gets all the high price options). The design contains the levels for ten attributes (the attributes other than vehicle type and fuel type) and four alternatives. The vehicle types and fuel types drawn according to the approach described above were used as inputs for calculating the base values for the levels in this underlying design. In the actual survey, each respondent was presented with one block of eight choice situations. Care was taken to ensure that the 18 different blocks were presented the same number of times and that there was no correlation between sample subgroups and blocks. The choice situations presented to the respondent were constructed on the basis of the set of vehicle type/fuel type combinations drawn for that respondent, and the block of 8 choice situations used from the experimental design for that respondent. The order in which the 8 choice situations from a given block were presented to a respondent was randomized across respondents.

Modelling methodology

As already alluded to in earlier parts of the paper, the SC data collected in this project was analysed with the help of discrete choice models belonging to the family of random utility models. For a thorough introduction to such models, see Train (2003). This section of the paper covers two main parts. We first discuss the specification of the utility function, which was identical for all models estimated. This is followed by a discussion of model structure. At this stage, it is also worth mentioning that the models used in the present paper exclude respondents from multi-car households.

Utility specification

An extensive specification search was conducted, with the outcome that the following terms were included in the final specification of the utility function:

- Constants for the first three SC alternatives, along with constants for fuel type and vehicle type inertia
- Vehicle type specific constants, with subcompact car being the base
- Fuel type specific constants, with standard gasoline being the base
- A constants for vehicles aged 1 or 2 years, and a constant for vehicles aged 3 years or more, with new vehicles being the base
- Four incentive constants, with *no incentive* being the base
- Marginal utility coefficients associated with vehicle price ($1000s), fuel costs ($1000s), maintenance costs ($1000s)
- A marginal utility coefficient interacting with vehicle price ($1000s) and the income group, where seven equally sized income groups were used, and where this linear relationship was justified on the basis of earlier results using income group specific cost coefficients
- Marginal utility coefficients associated with the miles per gallon equivalent (MPGE), range (miles), acceleration (seconds taken to 60mph)
- Constants associated with the option of plugging in electric vehicles at work and at other locations, and the availability of compressed natural gas at 1 out of 20 stations, with the respective bases being home plug-in only, and availability at 1 in 50 stations
- Constants associated with interaction terms for large households and medium-sized vehicles, large households and large-sized vehicles, alternative fuel vehicles and medium-sized vehicles, and alternative fuel vehicles and large-sized vehicles

This specification led to the use of 44 individual parameters. Note that efforts to include refuelling time in the models were unsuccessful.

Model structure

Although a large number of attributes are used to describe the various alternatives in the SC survey, two of them stand out as main *product* characteristics, namely the vehicle type and the fuel type. Given the nature of the choice scenarios, there are clear grounds to suspect a heightened degree of correlation between two alternatives sharing the same vehicle type or two alternatives sharing the same fuel type. This is even more so the case for two options that are of the same vehicle type and the same fuel type, but vary along some other dimension. To some extent, these correlations can be explained by the inclusion of alternative specific constants, but the degree of variation across respondents in their preference for the different vehicle types and fuel types is potentially so high that a large share of the correlation remains unmodelled. However, the use of a random coefficients approach to model the heterogeneity in the vehicle type and fuel type constants across respondents is not an option, given the high number of random terms this would lead to. On the other hand, not allowing for the effects of this unobserved correlation will likely lead to unrepresentative substitution patterns. Indeed, assume that a respondent is interested in purchasing a compact gasoline car and that for some reason, this vehicle becomes unavailable. This respondent is arguably then more likely to switch to a differently-sized gasoline car (say a sub-compact), then to switch to a vehicle that is of a different fuel type and a different vehicle type. Additionally, there is the possibility that the respondent may more closely evaluate a switch to a differently fuelled compact car (e.g. a hybrid compact) given the similarity in vehicle type. With the most basic modelling approach, a Multinomial Logit (MNL) model, such substitution patterns cannot be represented, and there will be a proportional shift in probability towards all other vehicle type and fuel type combinations.

The typical approach to dealing with such an issue is to estimate a Nested Logit (NL) model. For the sake of illustration, let us assume we're in a situation where a respondent has six vehicles to choose from:

A. Compact gasoline car
B. Compact hybrid-electric car
C. Compact gasoline car
D. Compact gasoline SUV
E. Compact flex fuel SUV
F. Compact hybrid-electric car

In this scenario, vehicles A, B, and C share the same vehicle type, as do vehicles D, E, and F. Vehicles A, C, and D share the same fuel type, as do vehicles B and F. Finally, vehicles A and C share the same vehicle type and the same fuel type. This provides ample source for correlation between alternatives. Various possible NL structures arise, as illustrated in Figure 2. In the first structure, we use a nesting by vehicle type approach, hence accounting for the correlation between options that share the same type of vehicle. As an example, if vehicle A was to become unavailable (or less attractive say due to a price increase), a respondent previously interested in this vehicle may be more likely to shift his/her interest to vehicle B or C, than to vehicles D, E, or F. The second figure shows the corresponding two-level NL structure using nesting by fuel type. Here, we allow for heightened correlation between vehicles A, C, and D, and between vehicles B and F. It is a likely outcome that both structures reveal heightened substitution patterns between alternatives sharing the same vehicle type, respectively vehicles sharing the same fuel type. This leads to a requirement for a structure that can jointly accommodate the two. A possible approach in this context comes in the use of a three-level NL structure, such as shown in the third example in Figure 2. This model nests first by vehicle type, and then by fuel type. It can immediately be seen that another option, not shown here, is to nest first by fuel type, and then by vehicle type. In the model shown in Figure 2, we still allow for correlation between the different compact cars, and for correlation between the different compact SUVs. In addition, we allow for even higher correlation between the two gasoline cars, i.e. options A and B. However, given the ordering of the nesting levels, the model is unable to account for the correlation between two options sharing the same fuel type but being of different vehicle type. As an example, we would expect options B and F to be closer substitutes for one another, but the model treats their errors as completely independent. The issue here is that when using a multi-level NL model for multi-dimensional choice processes, the full correlation can only be accommodated along the highest dimension of nesting in the tree, an issue that was to our knowledge first discussed by Hess & Polak (2006) in an air travel behaviour context.

The solution put forward by Hess & Polak (2006) is to use a CNL structure (cf. Vovsha, 1997), as illustrated for the present scenario in Figure 3. Here, we make use of two separate vehicle type nests and three separate fuel type nests, with each alternative falling into one vehicle type nest and one fuel type nest[1]. In the resulting structure, we have correlation between those alternatives sharing the same vehicle type (i.e. A, B, C, and D, E, F), and vehicles sharing the same fuel type (i.e. A, C, D, and B, F), with even higher correlation for those alternatives sharing the same vehicle type as well as the same fuel type (i.e. A and C).

With the above approach of characterising alternatives along two dimensions, we obtain 105 combinations of vehicle types and fuel types. With the survey making use of four separate SC alternatives, and each alternative potentially taking on one of those 105 combinations, the model

---

[1] On a technical aside, the CNL specification works by allocating an alternative by different proportions into different nests, collapsing back to a NL model when all allocation parameters are equal to 1, i.e. an alternative belongs into one nest one. In the present context, the allocation parameters were all fixed to a value of 1/2, meaning that an alternative belongs to one vehicle type nest and one fuel type nest. The estimation of actual values for the two non-zero allocation parameters for each alternative would have been very difficult due to the high degree of non-linearity and would arguably not have provided any further benefits from an interpretation perspective.

implementation actually made use of four sets of 105 alternatives, i.e. a total of 420 alternatives, of which exactly four were available in each choice situation

Model estimation

All model estimation and forecasting work reported in this paper was carried out using BIOGEME (Bierlaire, 2005), which is easily capable of dealing with such a large CNL structure. One further point needs addressing. The data used in this survey contains multiple observations for each respondent, potentially leading to correlations amongst choices for the same respondent. In the present work, this was not recognised in the modelling work for two reasons. The use of a random coefficients approach was not practical for computational reasons (the estimated models already took several hours without this added complication) and the use of the Jacknife approach (see e.g. Ortúzar, 1997) is not supported by BIOGEME, the only software well versed at estimating the complex model structure used in this paper. From this perspective, some of the estimates with a lower degree of significance should be treated with caution, given the possible underestimation of standard errors resulting from a cross-sectional estimation on repeated choice data.

Empirical results

The estimation results for the different discrete choice models are summarised in Table 1 for the main results and Table 2 for the nesting parameters for the NL and CNL models. The first observation we can make is that the NL model using nesting by fuel type gives us an improvement in model fit over the MNL model by 13.96 units in log-likelihood (LL), which is highly significant, coming at the cost of just six additional parameters. These additional parameters are nesting parameters which we will return to below. Similarly, the NL model using nesting by vehicle type gives us an improvement in LL over the MNL by 20.8 units, which, at the cost of 10 additional parameters, is again highly significant, as is the 37.97 unit improvement for the CNL model, at the cost of 15 additional parameters. Finally, the same is the case for the improvements offered by the CNL model over the NL model using nesting by vehicle type (17.16 units in LL with 5 additional parameters). A likelihood ratio test against the NL model using nesting by fuel type cannot be used due to the extra constraint on one of the nesting parameters in the CNL model, but the adj. ρ^2 measure shows that this improvement is similarly significant at usual levels of confidence.

Looking at the actual estimation results in Table 1, we observe, with a few exceptions, very similar parameter estimates across the four models, where the real differences between the models will become apparent later on, in the forecasting exercise. Going through the various estimates in turn, the values for the three constants suggest some allegiance to the reference alternative, along with a small amount of reading left to right impact. Further evidence of inertia is given by the two following estimates, showing that respondents are highly likely to choose a vehicle of their initially intended vehicle type and to a slightly lesser extent also fuel type. Without attempting to read too much into the various vehicle type and fuel type constants, the large negative values for the CNG and full electric vehicles do stand out, suggesting that additional incentives/improvements are required to increase the attractiveness of such vehicle given the low baseline preference. There is clear evidence of decreasing attractiveness with increasing age, while the various incentives have a positive impact on utility. The different cost components all lead to reductions in utility, though the vehicle price sensitivity is reduced as income increases. Better acceleration, longer range, better fuel efficiency and improved fuel availability all have positive impacts on utility, while large households show the expected preference for larger vehicles, and the attractiveness of alternative fuel vehicles reduces with vehicle size.

We next turn our attention to the nesting parameters, which explain the correlation between alternatives grouped together in a nest, where these parameters, shown in Table 2, are constrained

to be between 0 and 1, with lower values meaning higher correlation. The base value of 1 equates to an absence of correlation (as in a MNL model) and for this reason, the t-ratios are calculated with respect to a base value of 1 rather than 0. Looking first at the model using nesting by fuel type, we observe that the nesting parameter for full electric vehicles has collapsed to a value of 1, indicating no heightened correlation between different full electric vehicles. In addition, the values for the nesting parameters for three other fuel types, namely Flex Fuel/E85, Clean Diesel, and Compressed Natural Gas, are close to 1 and not significantly different from 1, suggesting that only low levels of correlation arise in these contexts. However, high correlation is observed between different gasoline cars, and also between different hybrid-electric cars, suggesting in each case the presence of heightened substitution patterns and greater fuel type *allegiance*. In the model using nesting by vehicle type, a number of nesting parameters once again collapse to a value of 1, while high correlation is for example observed in the Small cross-utility SUV nest and the Compact pick-up truck nest. Overall, the picture in the CNL model is the combination of the NL results, with the exception that we now observe high correlation in the Flex Fuel/E85 nest, and that the nesting parameter for the Compressed Natural Gas, already close to 1 in the NL model, has now collapsed to a value of 1.

Forecasting example

As a final illustration of the differences between the various models estimated in this paper, we now conduct a brief forecasting example. Clearly, rescaling of the model outputs and correction of the constants would be required before undertaking any forecasting for the purposes of guiding policy makers (cf. Louviere et al., 2000), but the aim of this example is purely illustrative.

Our forecasting example uses a single respondent with all 105 combinations of fuel type and vehicle type available. We further assume that this respondent currently owns a subcompact gasoline car, has an annual mileage of 13,500, and comes from a household with average income and four or fewer members. Finally, we assume that this respondent is solely interested in new cars. Our forecasting exercise starts by working out the probabilities for the 105 combinations of vehicle type and fuel type, with the four different models estimated in this paper. Next, we assume that following a government policy intervention, there is a reduction in the cost of Plug-in Hybrid Electric vehicles (fuel type 6) by $4,000, on condition that they also fall into the Subcompact car, Compact car or Mid-size car categories (vehicle types 1, 2, and 3).

Table 3 presents the changes in probabilities that arise as a result of this change in the attribute for these three vehicles. In the MNL model, we observe an equal increase in the probabilities for the three concerned vehicles, where this is drawn proportionally from all remaining vehicles as an effect of the independently distributed errors and the resulting absence of unmodelled correlation in this model. In the model using nesting by fuel type, we observe a bigger increase in the probability for the three vehicles than in the MNL model, but, with more of this being drawn from the remaining Plug-in Hybrid Electric vehicles than other fuel types, the overall increase in the probability for Plug-in Hybrid Electric options, across vehicle type, is far less marked. Both of these effects are consistent with intuition. However, the changes in this model draw proportionally from all other non Plug-in Hybrid Eletric vehicles, independently of the fuel type, which may again not be completely realistic.

In the model using nesting by vehicle type, we observe a bigger draw away from those options with the same vehicle type but with different fuel types (i.e. all non "Plug-in Hybrid Electric" subcompact, compact and mid-size cars). However, we also observe that the actual changes are different across the first three vehicle types. This is an effect of the different levels of correlation in these nests, with higher correlation leading to higher competition. With the nesting parameter (cf. Table 2) being lowest in the mid-size car nest, followed by the compact car nest and the subcompact car nest, the changes are also largest in that order. This model correctly recognises that a larger share of the draw

will come from other options in the same vehicle type categories, such that the increase in the overall share for the three vehicle types is less marked than in the first two models. However, the model doesn't recognise the fact that a bigger draw will also come from other Plug-in Hybrid Electric vehicles. Indeed, for a given vehicle type, all the different fuel types are affected in the same way, including in the case of Plug-in Hybrid Electric. This overstates the overall increase for the probability of that fuel type, especially when compared to the model using nesting by fuel type. Finally, while due to aggregation, the overall changes (i.e. when grossing up across categories) are similar in the CNL and MNL models, the model arguably gives a more intuitively meaningful representation of the changes in probabilities of individual vehicle type and fuel type combinations, with bigger reductions in probabilities for those alternatives that either are of the "Plug-in Hybrid Electric" fuel type or are of the "Subcompact car", "Compact car" or "Mid-size car" vehicle type.

Conclusions

This paper presents the findings of a modelling analysis looking at the choice of vehicle type and fuel type, captured through a stated choice survey conducted in California. In this experiment, respondents were faced with a number of scenarios that presented them with a choice between four vehicles, of varying vehicle type and fuel type.

The main aim of the present paper was to investigate the prevalence of correlation along these two dimensions of choice, i.e. the influence of unmodelled components on the choice process, where these are specific to given vehicle types and fuel types, and potentially lead to heightened substitution patterns between options of the same vehicle type or the same fuel type.

Our theoretical discussions show how the standard approach for dealing with such correlation, namely a Nested Logit model, may not be appropriate, as it is unable to capture the full extent of correlation along both the vehicle type and the fuel type dimension. Instead we suggest the use of a Cross Nested Logit approach, and show how this model offers better performance in estimation as well as producing substitution patterns that are more in line with a priori expectations.

The authors recognise that, especially in the context of large scale forecasting systems, the use of these model structures may not be possible at present, meaning that practitioners should seek to as far as possible incorporate any such correlations in the modelled utility components of more basic models such as MNL. The use of advanced nesting structures in applied work does however remain an important avenue for future work.

In closing, it should also be mentioned that, in the present paper, we have solely investigated the correlation between alternatives sharing the same vehicle type, respectively the same fuel type. It should be said that there are possibly also correlations between vehicles of different but similar type, such as for example compact and sub-compact cars, and the incorporation of such cross-type correlation is an important area for future work.

Acknowledgements

References

Adler, T., Wargelin, L., Kostyniuk, L., Kavalec, C. and Occhiuzzo, G. (2004), Experimental Assessment of Incentives for Alternate Fuel Vehicles, presented at the Transportation Research Board Annual Meeting, Washington, D.C. January 2004.

Batley, R. and Toner, J. (2003) Elimination-by-aspects and advanced logit models of stated choices for alternative-fuel vehicles. Proceedings of the European Transport Conference, Strasbourg, October 2003.

Batley, R., Toner, J. and Knight, M. (2004) A mixed logit model of UK household demand for alternative-fuel vehicles. International Journal of Transport Economics, 31 (1), pp55-77.

Bierlaire, M., 2005. An introduction to BIOGEME Version 1.4. biogeme.ep.ch.

Bunch, D.S., Bradley, M., Golob, T.F., Kitamura, R. and Occhiuzzo, G.P. (1993), Demand for Clean Fueled Vehicles in California: A Discrete Choice, Stated choice Survey, Transportation Research A, Vol 27A, pp237-253.

Golob, T., Brownstone, D., Bunch, D. and Kitamura, R. (1995), Forecasting Electric Vehicle Ownership and Use in the California South Coast Air Basin, Report to the Southern California Edison Company.

Greene, D. L. (2001), TAFV Alternative Fuels and Vehicles Choice Model Documentation, Oak Ridge National Laboratory.

Hensher, D.A. and Greene, W.G. (2001) Choosing between conventional, electric and LPG/CNG vehicles in single-vehicle households in Hensher, D.A. (ed) The Leading Edge of Travel Behaviour Research, Pergamon Press, Oxford, 725-750

Hess, S. & Polak, J.W. (2006b), Exploring the potential for cross-nesting structures in airport-choice analysis: a case-study of the Greater London area, Transportation Research Part E, 42, pp. 63-81.

Kavalec, C. (1996), CALCARS: The California Conventional and Alternative Fuel Response Simulator, A Nested Multinomial Vehicle Choice and Demand Model, Sacramento, CA: California Energy Commission.

Louviere, J.J., Hensher, D.A. and Swait, J.D. (2000), Stated Choice Methods—Analysis and Application, Cambridge University Press, UK.

Ortúzar, J de D., Roncagliolo, D.A., and Velarde, U.C., (1997), Interactions and Independence in Stated choice Modelling, PTRC Perspectives, London.

Spissu, E., Pinjari, A., Pendyala, R. and Bhat, C.R. (2009), A Copula-Based Joint Multinomial Discrete-Continuous Model of Vehicle Type Choice and Miles of Travel, forthcoming TRR, Washington DC.

Tompkins, M., Bunch, D., Santini, D., Bradley, M., Vyas, A., and Poyer, D. (1998), Determinants of Alternative Fuel Vehicle Choice in the Continental United States, TRR 1641.

Train, K. (1983), California Personal Vehicle Energy Demand Model, Sacramento: California Energy Commission

Train, K.E. (2003), Discrete choice methods with simulation, Cambridge University Press, Cambridge, MA.

Vovsha, P. (1997), 'Application of a Cross-Nested Logit model to mode choice in Tel Aviv, Israel, Metropolitan Area', Transportation Research Record 1607, 6–15.

Table 1: Estimation results for different discrete choice models

		MNL		NL (fuel type)		NL (vehicle type)		CNL	
LL		-7,654.51		-7,640.55		-7,633.69		-7,616.54	
par		44		50		54		59	
adj. ρ^2		0.265		0.265		0.266		0.267	
		est.	*t-rat. (0)*	*est.*	*t-rat. (0)*	*est.*	*t-rat. (0)*	*est.*	*t-rat. (0)*
Constants	Vehicle A constant	0.848	13.55	0.506	7.37	0.66	10.57	0.411	5.49
	Vehicle B constant	0.157	3.5	0.155	3.64	0.145	3.57	0.145	3.78
	Vehicle C constant	0.0357	0.78	0.0416	0.95	0.0258	0.63	0.026	0.67
	Vehicle Type Inertia	0.993	21.77	0.935	20.2	1.18	22.76	1.1	18.4
	Fuel Type Inertia	0.217	3.01	0.597	6.83	0.181	2.87	0.47	4.74
Vehicle type	Subcompact car	0	-	0	-	0	-	0	-
	Compact car	-0.131	-1.4	-0.13	-1.46	-0.1	-1.06	-0.106	-1.16
	Mid-size car	0.284	2.98	0.269	2.97	0.28	2.94	0.269	2.87
	Large car	0.0186	0.14	0.048	0.39	-0.0864	-0.66	-0.0353	-0.25
	Sport car	0.0749	0.64	0.11	0.98	0.0115	0.1	0.0357	0.31
	Small cross-utility car	0.449	4.19	0.418	4.13	0.427	4	0.409	3.93
	Small cross-utility SUV	0.44	3.55	0.444	3.83	0.382	3.17	0.399	2.92
	Mid-size cross-utility SUV	0.29	2.1	0.298	2.28	0.207	1.55	0.23	1.62
	Compact SUV	0.64	4.61	0.624	4.77	0.548	4.04	0.567	3.75
	Mid-size SUV	0.588	4.19	0.57	4.32	0.476	3.43	0.493	3.37
	Large SUV	0.825	4.98	0.804	5.22	0.645	3.95	0.67	4.3
	Compact van	0.212	1.46	0.214	1.59	0.134	0.94	0.164	1.17
	Large van	-0.264	-1.41	-0.213	-1.22	-0.315	-1.75	-0.247	-1.39
	Compact pick-up truck	0.144	1.17	0.157	1.33	0.112	0.92	0.127	1.22
	Standard pick-up truck	0.227	1.69	0.245	1.96	0.0953	0.71	0.149	1.01
Fuel type	Standard Gasoline	0	-	0	-	0	-	0	-
	Flex Fuel/E85	0.132	1.53	0.129	1.51	0.108	1.42	0.149	2.24
	Clean Diesel	0.178	1.39	0.141	1.13	0.167	1.48	0.137	1.19
	Compressed Natural Gas	-2.25	-2.16	-2.31	-2.22	-1.77	-1.98	-1.84	-1.99
	Hybrid-electric	0.0881	1.02	0.0752	0.89	0.0879	1.17	0.0843	1.14
	Plug-in Hybrid-electric	0.419	5.86	0.398	5.69	0.358	5.74	0.352	5.81
	Full Electric	-2.78	-3.64	-2.89	-3.79	-2.27	-3.45	-2.39	-3.49
Age	1 or 2 years old	-0.193	-3.06	-0.177	-2.97	-0.171	-3.07	-0.158	-3.04
	3 or more years old	-0.406	-7.59	-0.376	-7.47	-0.385	-8.14	-0.365	-8.77
Incentives	HOV lane use	0.0611	1.02	0.0709	1.22	0.0434	0.82	0.0472	0.87
	Free parking	0.041	0.69	0.0415	0.71	0.0374	0.71	0.0326	0.6
	$1,000 tax credit	0.186	3.11	0.181	3.15	0.162	3.06	0.156	3.02
	$1,000 reduced purchase price	0.0644	1.07	0.0632	1.08	0.0502	0.93	0.0514	0.96
Costs	Vehicle price	-0.0746	-14.27	-0.0704	-14.03	-0.0687	-14.15	-0.0656	-14.83
	Vehicle price * income cat	0.00675	6.47	0.00657	6.73	0.00614	6.46	0.00611	6.99
	Fuel costs	-0.0788	-9	-0.0777	-9.31	-0.0712	-9.08	-0.0702	-9.33
	Maintenance costs	-0.0584	-3.13	-0.0552	-3.16	-0.0523	-3.2	-0.0506	-3.17
Performance, efficiency, and fuel availability	MPGE	0.0169	5.7	0.0174	5.96	0.0142	5.3	0.0149	5.56
	Range	0.279	1.45	0.288	1.5	0.203	1.24	0.21	1.23
	Acceleration (seconds to 60mph)	-0.04	-5.24	-0.0387	-5.42	-0.0366	-5.49	-0.0346	-5.81
	Plug-in at work and other locations (EV)	0.133	1.03	0.13	1.01	0.133	1.21	0.131	1.17
	1 in 20 Stations (CNG)	0.327	2.21	0.322	2.18	0.304	2.36	0.312	2.28
Other	Large HH - Medium vehicles	0.397	2.97	0.401	3.19	0.389	2.94	0.377	2.94
	Large HH - Large vehicles	0.718	4.28	0.674	4.28	0.701	4.17	0.645	3.97
	Alt Fuel - Medium vehicles	-0.0154	-0.21	-0.0418	-0.58	-0.0112	-0.16	-0.053	-0.53
	Alt Fuel - Large vehicles	-0.277	-2.48	-0.296	-2.77	-0.292	-2.8	-0.314	-3.07

Table 2: Estimation results (part 2)

		NL (fuel type)		NL (vehicle type)		CNL	
		est.	t-rat. (1)	est.	t-rat. (1)	est.	t-rat. (1)
Fuel type	Standard Gasoline	0.67	-6.29	-	-	0.39	-3.60
	Flex Fuel/E85	0.76	-0.39	-	-	0.08	-85.01
	Clean Diesel	0.81	-1.10	-	-	0.79	-0.71
	Compressed Natural Gas	0.89	-0.17	-	-	1.00	-
	Hybrid-electric	0.58	-2.51	-	-	0.50	-2.17
	Plug-in Hybrid-electric	0.72	-3.90	-	-	0.58	-3.04
	Full Electric	1.00	-	-	-	1.00	-
Vehicle type	Subcompact car	-	-	0.87	-1.37	0.76	-1.27
	Compact car	-	-	0.71	-5.49	0.54	-6.27
	Mid-size car	-	-	0.70	-6.03	0.46	-5.19
	Large car	-	-	1.00	-	1.00	-
	Sport car	-	-	1.00	-	1.00	-
	Small cross-utility car	-	-	0.76	-1.73	0.65	-1.64
	Small cross-utility SUV	-	-	0.65	-3.67	0.47	-4.18
	Mid-size cross-utility SUV	-	-	0.75	-1.49	0.68	-0.73
	Compact SUV	-	-	0.67	-1.59	0.13	-1.66
	Mid-size SUV	-	-	0.78	-1.75	0.57	-1.59
	Large SUV	-	-	1.00	-	1.00	-
	Compact van	-	-	1.00	-	1.00	-
	Large van	-	-	0.61	-1.88	0.38	-1.37
	Compact pick-up truck	-	-	0.64	-2.58	0.34	-3.14
	Standard pick-up truck	-	-	1.00	-	1.00	-

Table 3: Forecasting example

MNL

Vehicle type	Fuel type 1	2	3	4	5	6	7	Total
1	-1.26%	-1.26%	-1.26%	-1.26%	-1.26%	17.83%	-1.26%	**10.29%**
2	-1.26%	-1.26%	-1.26%	-1.26%	-1.26%	17.83%	-1.26%	**10.29%**
3	-1.26%	-1.26%	-1.26%	-1.26%	-1.26%	17.83%	-1.26%	**10.29%**
4	-1.26%	-1.26%	-1.26%	-1.26%	-1.26%	-1.26%	-1.26%	**-8.80%**
5	-1.26%	-1.26%	-1.26%	-1.26%	-1.26%	-1.26%	-1.26%	**-8.80%**
6	-1.26%	-1.26%	-1.26%	-1.26%	-1.26%	-1.26%	-1.26%	**-8.80%**
7	-1.26%	-1.26%	-1.26%	-1.26%	-1.26%	-1.26%	-1.26%	**-8.80%**
8	-1.26%	-1.26%	-1.26%	-1.26%	-1.26%	-1.26%	-1.26%	**-8.80%**
9	-1.26%	-1.26%	-1.26%	-1.26%	-1.26%	-1.26%	-1.26%	**-8.80%**
10	-1.26%	-1.26%	-1.26%	-1.26%	-1.26%	-1.26%	-1.26%	**-8.80%**
11	-1.26%	-1.26%	-1.26%	-1.26%	-1.26%	-1.26%	-1.26%	**-8.80%**
12	-1.26%	-1.26%	-1.26%	-1.26%	-1.26%	-1.26%	-1.26%	**-8.80%**
13	-1.26%	-1.26%	-1.26%	-1.26%	-1.26%	-1.26%	-1.26%	**-8.80%**
14	-1.26%	-1.26%	-1.26%	-1.26%	-1.26%	-1.26%	-1.26%	**-8.80%**
15	-1.26%	-1.26%	-1.26%	-1.26%	-1.26%	-1.26%	-1.26%	**-8.80%**
Total	**-18.86%**	**-18.86%**	**-18.86%**	**-18.86%**	**-18.86%**	**38.42%**	**-18.86%**	

NL (fuel type)

Vehicle type	Fuel type 1	2	3	4	5	6	7	Total
1	-1.19%	-1.19%	-1.19%	-1.19%	-1.19%	20.63%	-1.19%	**13.51%**
2	-1.19%	-1.19%	-1.19%	-1.19%	-1.19%	20.63%	-1.19%	**13.51%**
3	-1.19%	-1.19%	-1.19%	-1.19%	-1.19%	20.63%	-1.19%	**13.51%**
4	-1.19%	-1.19%	-1.19%	-1.19%	-1.19%	-3.94%	-1.19%	**-11.06%**
5	-1.19%	-1.19%	-1.19%	-1.19%	-1.19%	-3.94%	-1.19%	**-11.06%**
6	-1.19%	-1.19%	-1.19%	-1.19%	-1.19%	-3.94%	-1.19%	**-11.06%**
7	-1.19%	-1.19%	-1.19%	-1.19%	-1.19%	-3.94%	-1.19%	**-11.06%**
8	-1.19%	-1.19%	-1.19%	-1.19%	-1.19%	-3.94%	-1.19%	**-11.06%**
9	-1.19%	-1.19%	-1.19%	-1.19%	-1.19%	-3.94%	-1.19%	**-11.06%**
10	-1.19%	-1.19%	-1.19%	-1.19%	-1.19%	-3.94%	-1.19%	**-11.06%**
11	-1.19%	-1.19%	-1.19%	-1.19%	-1.19%	-3.94%	-1.19%	**-11.06%**
12	-1.19%	-1.19%	-1.19%	-1.19%	-1.19%	-3.94%	-1.19%	**-11.06%**
13	-1.19%	-1.19%	-1.19%	-1.19%	-1.19%	-3.94%	-1.19%	**-11.06%**
14	-1.19%	-1.19%	-1.19%	-1.19%	-1.19%	-3.94%	-1.19%	**-11.06%**
15	-1.19%	-1.19%	-1.19%	-1.19%	-1.19%	-3.94%	-1.19%	**-11.06%**
Total	**-17.80%**	**-17.80%**	**-17.80%**	**-17.80%**	**-17.80%**	**14.58%**	**-17.80%**	

NL (vehicle type)

Vehicle type	Fuel type 1	2	3	4	5	6	7	Total
1	-1.76%	-1.76%	-1.76%	-1.76%	-1.76%	18.60%	-1.76%	**8.02%**
2	-2.43%	-2.43%	-2.43%	-2.43%	-2.43%	22.80%	-2.43%	**8.21%**
3	-2.71%	-2.71%	-2.71%	-2.71%	-2.71%	22.96%	-2.71%	**6.70%**
4	-1.28%	-1.28%	-1.28%	-1.28%	-1.28%	-1.28%	-1.28%	**-8.95%**
5	-1.28%	-1.28%	-1.28%	-1.28%	-1.28%	-1.28%	-1.28%	**-8.95%**
6	-1.28%	-1.28%	-1.28%	-1.28%	-1.28%	-1.28%	-1.28%	**-8.95%**
7	-1.28%	-1.28%	-1.28%	-1.28%	-1.28%	-1.28%	-1.28%	**-8.95%**
8	-1.28%	-1.28%	-1.28%	-1.28%	-1.28%	-1.28%	-1.28%	**-8.95%**
9	-1.28%	-1.28%	-1.28%	-1.28%	-1.28%	-1.28%	-1.28%	**-8.95%**
10	-1.28%	-1.28%	-1.28%	-1.28%	-1.28%	-1.28%	-1.28%	**-8.95%**
11	-1.28%	-1.28%	-1.28%	-1.28%	-1.28%	-1.28%	-1.28%	**-8.95%**
12	-1.28%	-1.28%	-1.28%	-1.28%	-1.28%	-1.28%	-1.28%	**-8.95%**
13	-1.28%	-1.28%	-1.28%	-1.28%	-1.28%	-1.28%	-1.28%	**-8.95%**
14	-1.28%	-1.28%	-1.28%	-1.28%	-1.28%	-1.28%	-1.28%	**-8.95%**
15	-1.28%	-1.28%	-1.28%	-1.28%	-1.28%	-1.28%	-1.28%	**-8.95%**
Total	**-22.25%**	**-22.25%**	**-22.25%**	**-22.25%**	**-22.25%**	**49.00%**	**-22.25%**	

CNL

Vehicle type	Fuel type 1	2	3	4	5	6	7	Total
1	-1.54%	-1.50%	-1.59%	-1.44%	-1.57%	20.08%	-1.47%	**10.97%**
2	-2.82%	-3.25%	-2.28%	-1.40%	-2.54%	25.41%	-1.49%	**11.64%**
3	-3.93%	-4.87%	-2.92%	-1.64%	-3.50%	27.66%	-1.53%	**9.28%**
4	-1.15%	-1.15%	-1.15%	-1.15%	-1.15%	-2.24%	-1.15%	**-9.13%**
5	-1.15%	-1.15%	-1.15%	-1.15%	-1.15%	-2.65%	-1.15%	**-9.53%**
6	-1.15%	-1.15%	-1.15%	-1.15%	-1.15%	-3.61%	-1.15%	**-10.49%**
7	-1.15%	-1.15%	-1.15%	-1.15%	-1.15%	-3.52%	-1.15%	**-10.40%**
8	-1.15%	-1.15%	-1.15%	-1.15%	-1.15%	-2.73%	-1.15%	**-9.62%**
9	-1.15%	-1.15%	-1.15%	-1.15%	-1.15%	-6.77%	-1.15%	**-13.65%**
10	-1.15%	-1.15%	-1.15%	-1.15%	-1.15%	-3.12%	-1.15%	**-10.01%**
11	-1.15%	-1.15%	-1.15%	-1.15%	-1.15%	-2.17%	-1.15%	**-9.06%**
12	-1.15%	-1.15%	-1.15%	-1.15%	-1.15%	-2.23%	-1.15%	**-9.12%**
13	-1.15%	-1.15%	-1.15%	-1.15%	-1.15%	-3.12%	-1.15%	**-10.01%**
14	-1.15%	-1.15%	-1.15%	-1.15%	-1.15%	-3.71%	-1.15%	**-10.60%**
15	-1.15%	-1.15%	-1.15%	-1.15%	-1.15%	-2.09%	-1.15%	**-8.98%**
Total	**-22.07%**	**-23.40%**	**-20.56%**	**-18.25%**	**-21.39%**	**35.21%**	**-18.25%**	

If the following vehicle options were available to you, which would you choose?
Please carefully examine all the attributes of each vehicle and then select the one you will most likely purchase by filling in the circle below your choice.

Vehicle Choice 1	Vehicle A	Vehicle B	Vehicle C	Vehicle D
Vehicle type	Midsize car	Compact SUV	Midsize car	Compact van
Fuel type	Gasoline	Natural Gas (NGV)	Plug-in Hybrid (PHEV)	Clean Diesel
Age of vehicle	New (2009)	New (2009)	New (2009)	New (2009)
Purchase price	$29,400	$36,600	$31,100	$20,900
Incentive	--	--	$1,000 tax credit	--
MPG or equivalent	29 MPG	15 MPG	60 MPG	31 MPG
Fuel cost per year	$1,090	$1,950	$780	$1,170
Fuel availability		1 in 50 stations		
Refueling time		10 Minutes at station, 4 hours at home		
Driving range		300 Miles		
Maintenance cost per year	$460	$370	$350	$550
Acceleration (0-60 mpg)	10.2 seconds	11 seconds	8 seconds	11.8 seconds
Select One:	○	○	○	○

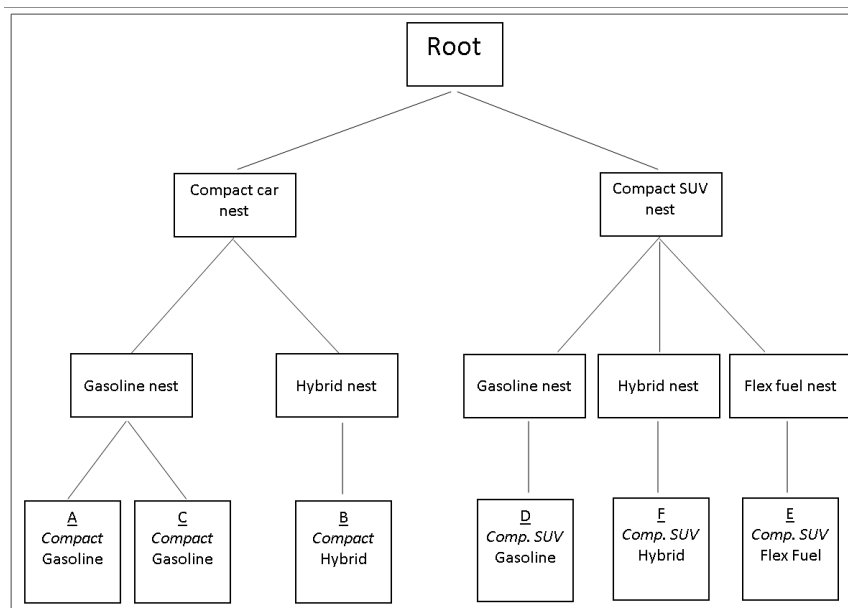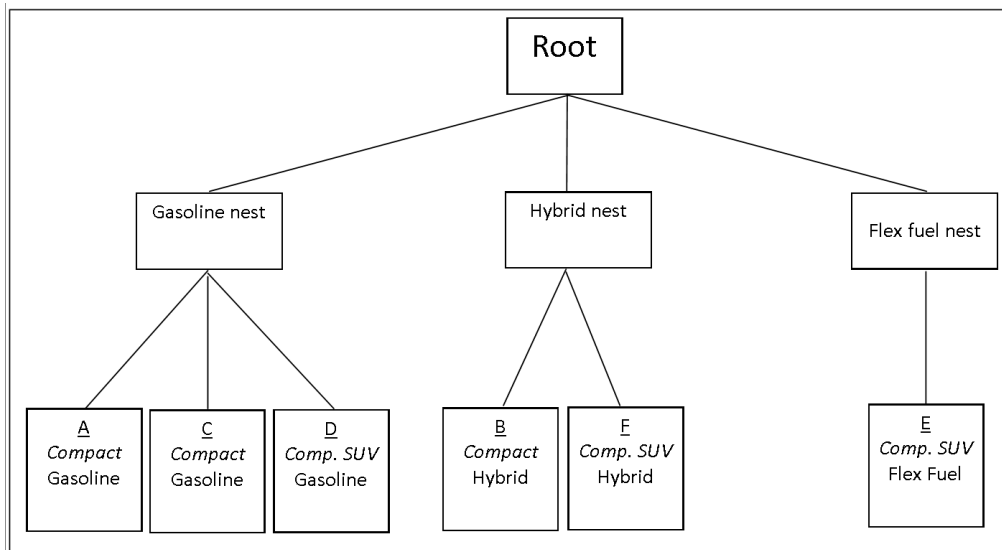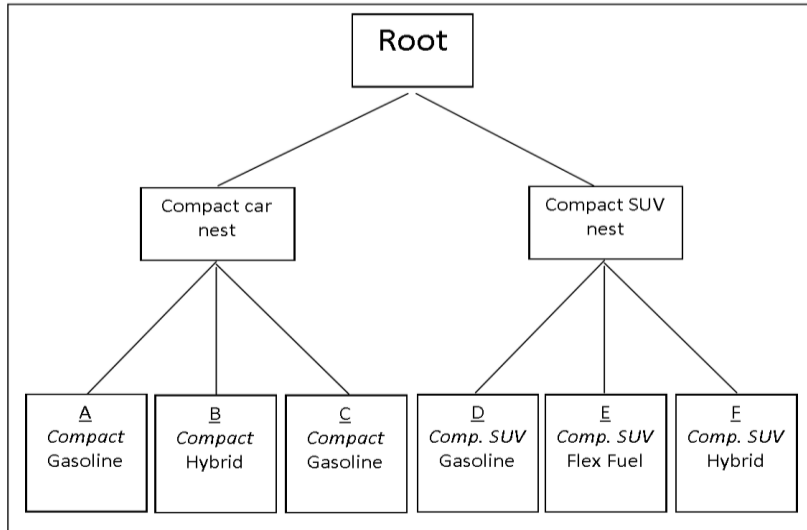Figure 1: Example Stated choice Exercise

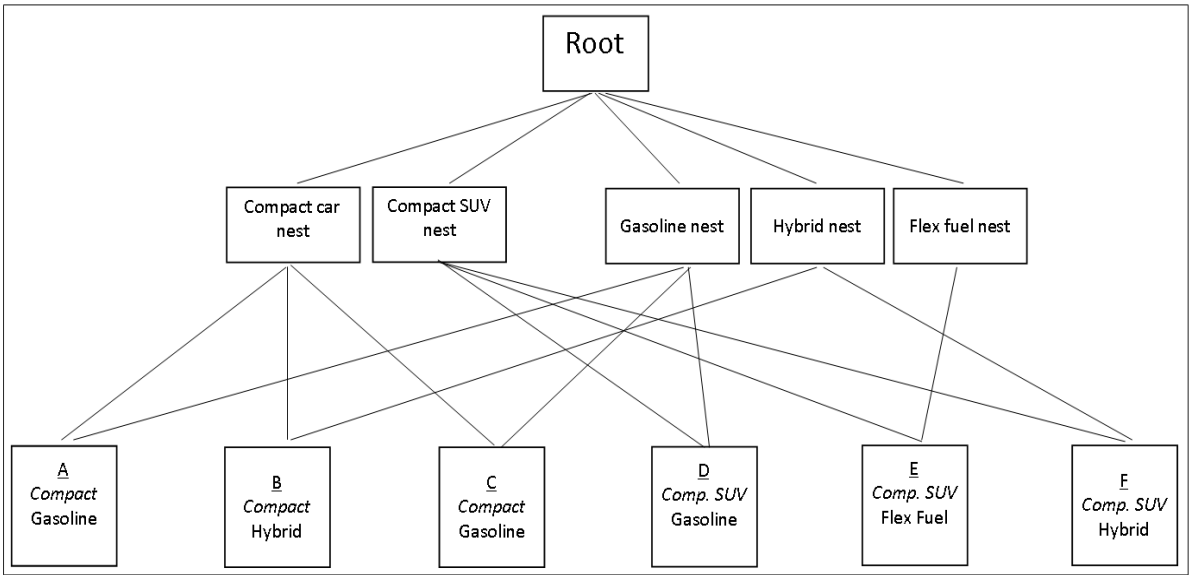15

Figure 2: Different possible NL structures

Figure 3: CNL structure