Which Commuters will Car Share? An Examination of Alternative Approaches to Identifying Market Segments.

Jon Crockett, Gerard Whelan, Caroline Sinclair (MVA Consultancy) and Hugh Gillies (Transport Scotland)

Abstract

Interest in car sharing initiatives, as a tool for improving transport network efficiency in urban areas and on interurban links, has grown in recent years. They have often been proposed as a more cost effective alternative to other modal shift and congestion relief initiatives, such as public transport or highway improvement schemes; however, with little implementation in practice, practitioners have only limited evidence for assessing their likely impacts.

This study reports the findings of a Stated Preference (SP) study aimed at understanding the value that car drivers put on car sharing as opposed to single occupancy trips. Following an initial pilot period, 673 responses were received from a web-based survey conducted in June 2008 amongst a representative sample of car driving commuters in Scotland.

An important methodological aspect of this study was the need to account for differences in behaviour to identify those market segments with the greatest propensity to car share. To this end, we estimated a range of choice model forms and compared the ability of each to consistently identify individual behaviours. More specifically, this included a comparison of:

- Standard market segmentation approaches based on multinomial logit with attribute coefficients estimated by reported characteristics (e.g. age, income, etc);
- A two stage mixed logit approach involving the estimation of random parameters logit models followed by an examination of individual respondent’s choices to arrive at estimates of their parameters, conditional on know distributions across the population (following Revelt and Train, 1999); and
- A latent class model involving the specification of \( C \) classes of respondent, each with their own coefficients, and assigning each individual a probability that they belongs to a given class based upon their observed choices, socio-economic characteristics and their reported attitudes.

As hypothesised, there are significant variations in tastes and preferences across market segments, particularly for household car ownership, gender, age group, interest in car pooling, current journey time, and sharing with a stranger (as opposed to family member/friend). Comparing the sensitivity of demand to a change from a single occupancy to a car sharing trip, it can be seen that the latter imposes a 'penalty' equivalent to 29.85 IVT minutes using the mixed logit structure and 26.68 IVT minutes for the multinomial specification. Segmenting this latter value according to the number of cars owner per household results in ‘penalties’ equivalent to 46.51 and 26.42 IVT minutes for one and two plus car owning households respectively.

Introduction

High Occupancy Vehicle (HOV), or carpool, lanes are reserved for the use of vehicles with two or more occupants (i.e. the driver plus one or more passengers), including buses. Interest in their
deployment has grown in recent years as Central and Local Governments seek initiatives for more efficient use of the highway network. By encouraging car sharing, as opposed to single occupant car trips, they have the potential to reduce the total volume of traffic on the network. Typically they are created through the widening of existing alignments or the use of the hard shoulder on the motorway network. In recent years interest in their deployment from scheme opening has also grown.

Understanding the propensity for using HOV lanes amongst the existing travelling population requires an appreciation not just of traditional attributes such as time and cost, but also the impact sharing with an acquaintance or stranger has on these attributes. Whilst it can be assumed that journey costs can be distributed between the driver and passengers of the car, a decision to car share may lead to an increase in journey time, through diverting to pick up passengers. Additionally, there may be ‘penalties’ to the car driver that can not be captured in the more readily quantifiable and measurable attributes such as monetary cost and time, including issues related to comfort, convenience, reliability and security.

It is common for these more qualitative attributes of travel to be considered as part of the Alternative Specific Constant (ASC), or mode penalty. The ASC is considered to be uniform across all journeys, ie it does not vary by journey time, cost or distance, by a given alternative form of transportation, although it may be segmented by journey purpose and other market segments such as demographic and socio-economic characteristics. Hypothesised components of the ASC construct in the context of this study include:

- Convenience in terms of the additional penalty associated with not being able to perform a ‘door-to-door’ journey or having to share personal space with other passengers;
- Reliability in that additional journey time and detour is likely to add a real, or perceived, additional element of (un)reliability to journey time. Additionally, passengers may have concerns regarding the reliability of their ‘pick-up’;
- Security, particularly in relation to having to share confined space with other travellers with whom the traveller in question may not be familiar;
- Comfort, whereby the ergonomics, cleanliness and privacy of the traveller’s immediate environment is compromised in some way; and
- Autonomy, typically a greater feeling of autonomy is generated by the feeling of being in control over one’s own life. Adherence to a fixed schedule ie a pick up time, may compromise or adversely affect such feelings. However, conversely, some people may seek the social interaction with fellow passengers.

Another means of considering the less tangible factors is via the inclusion of alternative specific In-Vehicle Time (IVT) factors, which quantify the relative (dis)utility of time spent on alternatives.

The objective of this study was to provide quantifiable evidence on the propensity of commuters in Scotland to use a dedicated HOV lane. A Stated Preference (SP) survey of car-borne commuters in Scotland was undertaken, to examine their sensitivity to changes in travel times and costs and to gain a better understanding of the modal penalty (ASC) or equivalent IVT factor, so that the implementation of HOV lanes could be robustly appraised. The primary objectives of the study were to:

- Develop ASCs relative to single occupancy for car sharers;
• Derive ASCs for certain key market segments, according to their relative importance in the model process;
• Examine the potential for deriving IVT factors for car sharing as opposed to single occupancy (normalising the IVT factor of drive alone to one);
• Examine the impacts of sharing with an acquaintance as opposed to a member of a car pool or work colleague; and
• Develop a forecasting tool to assess the relative market share (probability) for car sharing under a number of hypothetical scenarios.

3 Modelling Approaches

The modelling of random taste distributions within a population under the RUM behavioural framework has relied on either continuous distributions or finite ones. Whilst continuous distributions have considerable merit, particularly in terms of their seeming analogy with respect to real world choice data (Train, 2003), their use in modelling choices and estimation presents particular challenges (see Hensher and Greene, 2003). This is especially true when we attempt to specify a functional form (typically normal or lognormal) for what may be considered ‘lumpy’ preference data, where underlying segments, or classes, are the source of the distribution in attribute valuations.

The alternative, finite, approach employing endogenous segmentation of the population, offers considerable benefits in terms of ease of implementation, estimation and interpretation. Such segmentation has traditionally relied on easily collected data relating to the individual or household; however, a growing body of research has identified that preferences are not so easily grouped and additional attitudinal data, non-measured characteristics, and latent classes may be more reliable indicators of differences in taste and preference across the population.

3.1 Multinomial Logit

Standard market segmentation approaches have, for many years, been based on the multinomial logit form with attribute coefficients estimated by reportable, easily measurable, and comparable characteristics (eg age, income, journey purpose etc). Functional forms are well documented and traditionally involve separate valuations of the Alternative Specific Constant (ASC) component of the utility specification. The relative merits of the multinomial logit model, and benefits/disbenefits of prejudging the market segments of interest, do not require wider discussion at this point, except to say that their shortcomings have been a stimulus into research for alternatives for a number of years.

3.2 Mixed Logit

The mixed logit model (see Ben-Akiva and Bolduc, 1996 or McFadden and Train 2001) stands as the most significant extension of the multinomial logit approach. The utility function of an alternative \( j \) for an individual \( i \) is given by:

\[
U_{ij} = \theta_i X_{ij} + \varepsilon_{ij}
\]  

(1)

where:

\( U_{ij} \) = the utility of alternative \( j \) for individual \( i \)
\( \theta_i \) = a vector of unknown coefficients that vary randomly according to individual tastes

\( X_{ij} \) = a vector of observable variables

\( \epsilon_{ij} \) = a random error term which is assumed to follow an IID Gumbel distribution, independent of \( \theta_i \) and \( X_{ij} \)

Among the many attractive features of the mixed logit model is its ability to take account of taste variation among decision-makers by allowing coefficients (\( \theta_i \)s) to follow pre-specified distributions (usually normal or lognormal). Whilst accounting for heterogeneity in the population, simple applications of the technique fail to identify valuable information on differences in preference and behaviour between market segments.

The ‘standard’ approach to overcome this problem when working with the mixed logit model is to identify segments prior to modelling and either specify a set of constant coefficients for each market segment together with an additional error term to ‘mop-up’ any residual variation, or by allowing separate distributions for each market segment.

A ‘two stage’ mixed logit approach involves the estimation of random parameters logit models followed by an examination of individual respondent’s choices to arrive at estimates of their parameters. These latter parameters are conditional on known distributions across the population (see Revelt and Train, 1999).

Following on from equation (1), if \( k_i, k_{i1}, \ldots, k_{it} \) denotes the series of choices made by an individual \( i \) in a choice experiment, then conditional on the individual’s preferences (\( \beta_i \)), the probability that individual \( i \) chooses alternative \( k \) (from alternatives \( j, j+1, \ldots, k \)) in choice scenario \( t \) can be expressed in the logit form:

\[
P_{it}(k_i|\beta_i) = \frac{\exp(\beta X_{i_k})}{\sum_j \exp(\beta X_{j_k})} \tag{2}
\]

The unconditional probability is the integral of the conditional probability over all possible values of \( \beta \):

\[
Q_{it}(k_i|\theta) = \int P_{it}(k_i|\beta) f(\beta|\theta) d\beta \tag{3}
\]

Assuming that the individual’s tastes do not change over choice situations, the conditional probability of individual \( i \)’s sequence of choices is the product of the logits:

\[
S(y_i|\beta) = \prod L_{it}(y_{it}|\beta) \tag{4}
\]

The unconditional probability is:

\[
P(y_i|\theta) = \int S(y_i|\beta) f(\beta|\theta) d\beta \tag{5}
\]

The goal of the first stage of the estimation process is to estimate parameters that describe the distribution of tastes across individuals. Unlike the estimation of standard logit models exact maximum likelihood estimation is not possible since the integral in equation (5) cannot be evaluated analytically. Instead, a simulated likelihood function is specified in which \( P(y_i|\theta) \) is approximated by summation over randomly chosen values of \( \beta \). The process is repeated for \( R \)
random draws of $\beta$ (where $\beta'$ is the $r$-th draw from $f(\beta|\theta)$) and the simulated probability of the individual’s sequence of choices is:

$$SP(y_i|\theta) = \frac{1}{R} \sum_{r=1}^{R} S(y_i|\beta'^r)$$

(6)

So long as the number of random draws is sufficiently large, the simulated probability is an unbiased estimate of the true probability and the simulated likelihood function is constructed as $SLL = \sum \ln(SP(y_i|\theta))$. In recent years, the exploitation of Halton sequences (Train, 1999; Bhat, 2000), and more recently techniques such as the Modified Latin Hypercube Strategy (MLHS – see Hess, Train and Polak, 2006), has improved both the accuracy and speed of estimation.

Although we can estimate the density $f(\beta|\theta)$ in equation (5) describing the distribution of tastes in the population it is also desirable to know where each decision-maker is in this distribution. Following Revelt and Train (1999), let $g(\beta|y_i, \theta)$ denote the density of $\beta$ conditional on the decision-maker’s ($i$) sequence of choices and the population parameters $\theta$. By Bayes’ rule:

$$g(\beta|y_i, \theta) = \frac{P(y_i|\beta) \cdot f(\beta|\theta)}{P(y_i|\theta)}$$

(7)

Equation (7) is then used to calculate the conditional expectation of $\beta$, the individual’s expected tastes $k(\beta)$:

$$E(k|y_i, \theta) = \int k(\beta) \cdot g(\beta|y_i, \theta) d\beta$$

(8)

Substituting the formula for $g$:

$$E(k|y_i, \theta) = \frac{\int k(\beta) \cdot P(y_i|\beta) \cdot f(\beta|\theta) d\beta}{P(y_i|\beta)}$$

(9)

As equation (9) does not have a closed form, the conditional expectation of $\beta$ is approximated by simulation. This procedure involves taking random draws of $\beta$ from the population density $f(\beta|\theta)$ and estimating the weighted average of these draws with the weight of the draw $\beta'$ being proportional to $P(y_i|\beta')$:

$$\tilde{E}(k|y_i, \theta) = \frac{\sum \beta' \cdot P(y_i|\beta')}{\sum P(y_i|\beta')}$$

(10)
Whelan (2003) reports on the results of a two-stage mixed logit approach versus conventional mixed logit from a stated preference experiment on choice of vehicle type. As would be expected, an un-segmented mixed logit structure was found to have a significantly better fit than the standard multinomial logit. Using standard socio-economic characteristics, the process outlined above was found to be unsatisfactory as preferences for different vehicle types were diverse and not easily grouped using conventional market segments. Latent-class models were recommended as an alternative methodology for further analysis (Whelan, 2003).

3.3 Latent Class

Latent-Class (L-C) models are an alternative form of endogenous segmentation. Previous studies have argued its merits as a convenient and intuitive alternative to the mixed logit continuous approach, particularly in terms of computational time and cost (Provencher, Barenklau and Bishop, 2002). It enables preferences amongst a sample to be modelled without making any prior assumptions regarding some observable deterministic explanation for that heterogeneity, i.e. through demographic, socio-economic and travel characteristics or via psychographics1. From a theoretical point of view the latent class structure has much to recommend it, not least in its ability to incorporate respondent attitudes within the analysis.

Use of L-C models has been widespread in other sciences for a number of years, but, with the exception of Greene and Hensher (2002), little applied in transportation research and practice. Examples of L-C models based upon choice data include Provencher et al (2002) and Scarpa and Thiene (2005). In these studies choice data was used to estimate the number of (latent) classes, the probability of class membership, and the preference parameters in each class’s conditional, indirect-utility function. No attitudinal data was used in these studies; Breffle, Morey and Thacher (2004) report on the combination of attitudinal and choice data to produce combined L-CAc models, where the subscripts A and C denote Attitudinal and Choice.

In each preference class it is assumed that individuals respond and behave in a similar manner to one another, i.e. the choice patterns of individuals from the same preference class are more correlated with each other than with individuals in other classes. Each individual is assigned a probability (P) of belonging to a given class (C_i) based upon their observed choices, socio-economic characteristics and their reported attitudes (if provided). Once preference class has been accounted for, it is assumed that all choices are independent across scenarios and individuals. Following Greene and Hensher (2002), for a given class of respondent C, the probability of the discrete choice j among J_i alternatives by individual i in choice scenario t can be specified as:

\[ P_{ijtc} = \frac{\exp(\beta_j X_{ij})}{\sum_j \exp(\beta_j X_{ij})} \]  

where:

- \( P_{ijtc} \) = the probability of individual i of class c choosing alternative j in choice scenario t

---

1 Psychographics are any attributes related to a person’s, or group’s personality, values, attitudes, interests or lifestyles. Attitudinal data relating to these aspects can be incorporated in choice modelling alongside standard variables using techniques such as cluster analysis.
\[ \beta_c = \text{a set of estimated coefficients for class } c \]

\[ X_{ijt} = \text{a set of attributes that describe the choice alternative} \]

This standard logit choice model is specified with its own set of coefficients for \( C \) classes of respondent.

The probability that a respondent belongs to a given class can be based on their observed choices, their observed characteristics and their reported attitudes. This is achieved using another logit model that shows the probability that individual \( i \) belongs to class \( c \) as:

\[
P_{ic} = \frac{\exp(\theta_c Z_i)}{\sum_{c=1}^{C} \exp(\theta_c Z_i)}
\]

where:

\[ P_{ic} = \text{the probability that individual } i \text{ belongs to class } c \]

\[ \theta_c = \text{a set of } c \text{ coefficient vectors} \]

\[ Z_i = \text{a set of observable characteristics and reported attitudes} \]

The models detailed in equations (11) and (12) are calibrated jointly using maximum likelihood estimation. It can be seen that across all classes and alternatives \( P_i \) is equal to one:

\[
P_i = \sum_{c=1}^{C} P_{ij} P_{ic}
\]

Despite these apparent benefits, a significant drawback exists regarding the specification of the number of underlying classes within the data. Indeed, there is no well established statistical test for determining differing hypotheses regarding their number. \( C \) is not a parameter in the interior of a convex parameter space; consequently, comparison of log likelihoods of sequentially smaller models is not an appropriate approach (Greene and Hensher, 2002). Scarpa and Thiene (2005) report on the relative merits of different selection (information) criteria, including the Akaike Information Criteria (AIC), Bayesian Information Criteria (BIC) and the corrected AIC (Hurvich and Tsai, 1989). All of these criteria fail some of the regularity conditions for valid test under the null (Leroux, 1992). The AIC is reported to be biased towards an over-estimate of the number of classes, whilst the BIC is not. Conversely, the BIC tends towards an under-estimate when the sample size is small (McLachlan and Peel, 2000). As the number of classes increases, the significance of the parameter estimates in the utility function decrease. This is especially true in classes with low probability of class membership. The number of classes must therefore account for the significance of parameter estimates and the meaningfulness of their sign and magnitude (Scarpa and Thiene, 2005).

Undoubtedly this potential shortcoming leaves \( L\cdot C \) models vulnerable to ‘attack’ from other parties who could speculate on the ‘arbitrary’ selection (by the practitioner) of the number of classes. Much is therefore dependent on the practitioner’s prior hypothesis. For example, in the
context of car sharing, it would appear intuitive to hypothesise a minimum of three latent classes:

- A class strongly opposed to sharing ‘personal space’ with others, ie those favouring autonomy;
- An indifferent class, who, provided the incentive is strong enough are willing to share; and
- A ‘socialable’ class who would derive utility from the sharing of the space with others.

In addition to the above, we could speculate that there will be classes of respondent primarily motivated by, eg, time or money; however, these attributes are incorporated in the experiment and will not therefore influence latent class membership (under the assumption that the utility parameters for time and cost explicitly account for all variations in taste and preference associated with these attributes).

Greene and Hensher (2002) report on the comparative qualities of a mixed logit and L-C approach on the same stated preference dataset, from an experiment pertaining to route choice. Both forms are preferred to the multinomial logit, with the L-C model offering clear benefits in terms of its lack of distributional assumptions about individual heterogeneity. However, it is noted that the range of functional forms which can be specified within the mixed logit formulation offset this to a great extent. In conclusion, the differences in degree-of-fit are not sufficient to warrant the choice of one model over the other, instead “both models allow the analyst to harvest a rich variety of information about behaviour from a panel, or repeated measures, dataset”.

4 Data

In order to maximise the number of people reached within the target market, a web-based self completion survey was employed. The survey form was designed to be generic, so that responses could be elicited from respondents all across Scotland. This sampling strategy ensured that the valuations and recommendations regarding car sharing are transferable between different contexts and localities. This assumption was tested through the incorporation of geography, in the form of the Scottish Executive’s urban/rural classification, as a potential explanatory variable within the model construct.

The survey was designed in such a way as to ensure that the decision context related to an actual journey likely to be experienced by the respondent, and the ‘response space’ allows the respondents to describe their hypothetical behaviour. Screening questions at the beginning of the surveys ensured that the respondents were able to provide a response relevant to the car sharing context. Users of non-car modes were automatically routed past the SP choice scenarios, and solely provided information regarding their current travel patterns, demographics and socio-economic characteristics.

The study was concerned with the values car users place on car sharing as opposed to driving alone, expressed in terms of the attributes:

- Monetary cost:
  - parking charges; and
  - fuel cost.
- Journey time:
- As a result of HOV lanes;
- walk time;
- waiting time due to pick-up/drop-off of all car passengers; and
- having a designated car parking space at their workplace.

- Car sharing context:
  - Sharing with a friend or family member; and
  - A car pool member or work colleague.

The choice experiment asked the respondent to consider a journey similar to the one they undertook on the day they completed the survey, where the available options were single occupancy car trip, car share (using a HOV lane) and don’t travel/travel elsewhere/use another mode.

To ensure that the SP design was both realistic and relevant to the respondent, the experimental design was expressed as differences from the respondent’s current journey time and costs. The SP design contained five attributes in total. Journey cost and time were assigned three possible levels, with car sharing assigned two levels. The attributes are presented in Table 1.

### Table 1: Attributes and Levels

<table>
<thead>
<tr>
<th></th>
<th>Level 1</th>
<th>Level 2</th>
<th>Level 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Journey Time</strong></td>
<td>15</td>
<td>5</td>
<td>25</td>
</tr>
<tr>
<td><strong>Car Alone</strong></td>
<td>125</td>
<td>50</td>
<td>0</td>
</tr>
<tr>
<td><strong>Cost</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Journey Time</strong></td>
<td>-20</td>
<td>10</td>
<td>-10</td>
</tr>
<tr>
<td><strong>Car Share</strong></td>
<td>-50</td>
<td>-100</td>
<td>25</td>
</tr>
<tr>
<td><strong>Sharing</strong></td>
<td>0</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td><strong>Cost</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes:
(1) Time is shown in minutes difference from current journey
(2) Cost is shown in pence difference from current journey
(3) Sharing is shown as 0 = family member or friend, 1 = colleague or car pool member

The full factorial design of the alternatives presented in Table 1 necessitates nine scenarios in the overall fractional factorial statistical design, summarised for one block of the final design in Table 2.
Table 2: Stated Preference Differences Design (Block 1)

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Journey</th>
<th>Cost</th>
<th>Journey Time</th>
<th>Car Share</th>
<th>Cost</th>
<th>Sharing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scenario 1</td>
<td>15</td>
<td>125</td>
<td>-20</td>
<td>Family/Friend</td>
<td>-50</td>
<td></td>
</tr>
<tr>
<td>Scenario 2</td>
<td>5</td>
<td>50</td>
<td>10</td>
<td>Family/Friend</td>
<td>-100</td>
<td></td>
</tr>
<tr>
<td>Scenario 3</td>
<td>25</td>
<td>125</td>
<td>-10</td>
<td>Stranger</td>
<td>-100</td>
<td></td>
</tr>
<tr>
<td>Scenario 4</td>
<td>15</td>
<td>0</td>
<td>-20</td>
<td>Stranger</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>Scenario 5</td>
<td>25</td>
<td>50</td>
<td>-10</td>
<td>Stranger</td>
<td>-100</td>
<td></td>
</tr>
<tr>
<td>Scenario 6</td>
<td>5</td>
<td>0</td>
<td>10</td>
<td>Family/Friend</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>Scenario 7</td>
<td>5</td>
<td>125</td>
<td>10</td>
<td>Stranger</td>
<td>-50</td>
<td></td>
</tr>
<tr>
<td>Scenario 8</td>
<td>15</td>
<td>50</td>
<td>-20</td>
<td>Family/Friend</td>
<td>-100</td>
<td></td>
</tr>
<tr>
<td>Scenario 9</td>
<td>25</td>
<td>0</td>
<td>-10</td>
<td>Family/Friend</td>
<td>25</td>
<td></td>
</tr>
</tbody>
</table>

4.1 Sample

A total of 673 surveys were received from respondents whose most common mode of travel to work was to drive a car or van. The survey collected data about the regular commute to work and demographic and socio-economic characteristics of the respondents.

Tables 3 presents the gender and age profile of the survey respondents against corresponding data for the total Scottish population (General Register Office for Scotland), and shows that the survey respondents were typical of the general population. As the survey was targeted at commuters, the percentage of respondents over the age of 60 will be lower than the general population, but overall, a similar profile in age groups is apparent.

Table 3: Profile of Survey Respondents and Scotland

<table>
<thead>
<tr>
<th>Gender</th>
<th>Survey (%)</th>
<th>Scotland</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>52</td>
<td>48</td>
<td>+4</td>
</tr>
<tr>
<td>Female</td>
<td>48</td>
<td>52</td>
<td>-4</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17 – 24 years</td>
<td>6</td>
<td>16</td>
<td>-10</td>
</tr>
<tr>
<td>25 – 34 years</td>
<td>27</td>
<td>19</td>
<td>+8</td>
</tr>
<tr>
<td>35 – 44 years</td>
<td>35</td>
<td>24</td>
<td>+11</td>
</tr>
<tr>
<td>45 – 54 years</td>
<td>25</td>
<td>22</td>
<td>+3</td>
</tr>
<tr>
<td>55 – 59 years</td>
<td>4</td>
<td>10</td>
<td>-6</td>
</tr>
<tr>
<td>60 years and over</td>
<td>3</td>
<td>8</td>
<td>-5</td>
</tr>
</tbody>
</table>
Table 4 shows the car ownership of the respondents, who under the initial screening questions have indicated that their normal commute was as a car driver (and are therefore likely to own a car). The majority of respondent’s households own more than two cars.

<table>
<thead>
<tr>
<th>Car Ownership</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>0.6</td>
</tr>
<tr>
<td>One</td>
<td>38.7</td>
</tr>
<tr>
<td>Two</td>
<td>52.5</td>
</tr>
<tr>
<td>Three plus</td>
<td>8.2</td>
</tr>
</tbody>
</table>

Table 5 shows the propensity of the survey respondents to travel with other people, by gender. It shows that the majority of people travel alone, but that females are more likely to travel with companions, and their companions are more likely to be children. Existing journeys with companions such as children could reduce the flexibility of respondents choosing to join car-pool schemes.

<table>
<thead>
<tr>
<th>Travelling Companions</th>
<th>Percentage by Gender (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male</td>
</tr>
<tr>
<td>Alone</td>
<td>82</td>
</tr>
<tr>
<td>Self + Child</td>
<td>4</td>
</tr>
<tr>
<td>Self + Adult</td>
<td>10</td>
</tr>
<tr>
<td>Self + Two others</td>
<td>4</td>
</tr>
</tbody>
</table>

The flexibility of respondents with respect to potential for car-pooling was explored further by considering ‘trip-chaining’ or performing journeys with multiple purposes eg shopping on the way home from work. Table 6 illustrates the other activities undertaken by survey respondents on their journey to work. The majority of respondents do not undertake additional activities, but shopping and performing the school run are popular activities. The survey also revealed that women and respondents aged 25-54 were more likely to undertake additional activities.

<table>
<thead>
<tr>
<th>Additional Activities</th>
<th>Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nothing</td>
<td>50.5</td>
</tr>
<tr>
<td>Take spouse/ partner</td>
<td>7.7</td>
</tr>
<tr>
<td>Take children to school</td>
<td>14.9</td>
</tr>
</tbody>
</table>
Respondents were also asked if they would consider joining a car-pool initiative. 36% were interested in car sharing while 3% indicated they were already a member of such a scheme. Of the 415 respondents who provided a response to the potential use of car pooling should there be HOV lanes, 20% indicated that they would be interested in joining a car share scheme.

5 Results

The data from the Stated Preference (SP) experiments was analysed under a discrete choice modelling framework. This framework is based on the principle that a decision-maker (the traveller) chooses the choice alternative (mode) that yields greatest satisfaction or 'utility', where utility is taken to be related to the ‘attributes’ of the choice alternative (e.g., monetary cost, journey time, and who the vehicle was shared with). The choice context is composed of a finite set of alternatives.

The aim was to develop a model that shows the probability that a decision-maker will choose a choice alternative and to quantify how this choice probability is influenced by changes in the attribute of the alternatives. By comparing the relative influence of one attribute against another, it is possible to infer its relative value. For example, by comparing the influence that changes in journey time have on choice compared with the influence that changes in journey cost have on choice, it is possible to estimate the implied Value of Time (VoT). The objective of the study was to quantify the additional value that car drivers and passengers put on their time spent sharing their vehicle with others (friends/family or car pool member/colleague). This can be expressed as an absolute value in money or equivalent journey time units (an ASC or modal penalty), or can be related to the length of the journey by expressing the value proportionate to journey time (an IVT factor).

Following conventional modelling practice, we started the data analysis using a Multinomial Logit (MNL) model, followed by more complex model forms which take better account of correlation between choice alternatives and variations in tastes and preferences across travellers.

5.1 Multinomial Logit Models

The main models were estimated with a non-nested simple linear in the parameters form. The estimation models underwent a thorough set of econometric diagnostic tests to ensure that they were statistically robust. This included an examination of their overall level-of-fit, precision in parameter estimation and accommodation of correlation between attributes and alternatives. Two main models were developed as follows:

- **Model 1** has a logit structure with the value of sharing (as opposed being alone) expressed as an absolute value; and
Model 2 has a logit structure with the value of time spent sharing (as opposed to alone) expressed as a factor of in-vehicle time.

In the first model the utility of each mode is specified as a function of monetary cost (fuel and parking), in-vehicle time, and Alternative Specific Constants (ASCs: $\delta_1$, $\delta_2$, and $\delta_3$). The ASCs capture the preferences for travel by each alternative after taking account of the other variables (attributes) in the model, and as such represent the value of car sharing as opposed to car alone. The resulting utility specifications take the form:

$$V_{\text{Car Alone}} = \beta_1 \text{Cost}_{\text{Car Alone}} + \beta_2 \text{IVT}_{\text{Car Alone}} + \delta_1$$
$$V_{\text{Car Share}} = \beta_1 \text{Cost}_{\text{Car Share}} + \beta_2 \text{IVT}_{\text{Car Share}} + \delta_2$$
$$V_{\text{NT}} = \delta_3$$

(14)

An alternative specification is in terms of a factor for In-Vehicle Time (IVT). Initial analysis was undertaken with no ASCs; however, this provided a positive coefficient on the IVT for single occupancy car trip, which meant a (credible) IVT factor could not be computed.

$$V_{\text{Car Alone}} = \beta_1 \text{Cost}_{\text{Car Alone}} + \beta_2 \text{IVT}_{\text{Car Alone}} + \delta_1$$
$$V_{\text{Car Share}} = \beta_1 \text{Cost}_{\text{Car Share}} + \beta_4 \text{IVT}_{\text{Car Share}} + \delta_2$$
$$V_{\text{NT}} = \delta_3$$

(15)

In addition, to the above a dummy variable for sharing with a stranger (car pool member/colleague) was also included in an earlier specification. However, within this model specification, the parameter was found to be insignificant and of the intuitively wrong sign. It was therefore excluded from further analysis in the aggregate model. Table 7 presents the results of the model estimations.

The adjusted Rho squared for Model 1 is relatively high (0.187) showing a good degree-of-fit to the data. T-statistics are highly significant whilst parameter estimates have intuitively correct signs. The Value of Time (VoT) for commuters is marginally lower than that recommended by the Scottish Government (£5.68/hour for car drivers and £5.02/hour for other occupants). The implied value for car sharing as opposed to drive alone is 26.68 minutes per trip.

Although the adjusted Rho squared for Model 2 suggests a good degree-of-fit to the data at 0.189 and all t-statistics are significant at the 5% level, the parameter for sharing IVT is of a lower magnitude than single occupancy. This is counter intuitive to the general hypothesis that car sharing imposes a greater penalty upon the traveller than travelling by car alone, and contradicts the longitudinal trend of reducing car occupancy rates across Scotland. The utility construct under this form is also unnecessarily complicated by the consideration of both IVT factors and ASCs, with uncertainty regarding the ‘true’ penalty for car sharing for the practitioner. The combination of the two terms results in a distorted, and unrepresentative, distribution of sharing valuations when combined with the ‘pure’ IVT. The IVT factor results do not therefore form an appropriate specification technique for aggregate demand models such as Transport Model for Scotland (TMfS). Subsequent analysis focuses on the specification of models which consider additional terms, market segments, and specifications.
### Table 7: Car Sharing Valuation (Models 1 and 2)

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Model 1 Coefficient</th>
<th>T-Stat</th>
<th>Model 2 Coefficient</th>
<th>T-Stat</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASC_CarAlone ($\delta_1$)</td>
<td>0</td>
<td>--fixed--</td>
<td>0</td>
<td>--fixed--</td>
</tr>
<tr>
<td>ASC_CarShare ($\delta_2$)</td>
<td>-1.23</td>
<td>-23.45**</td>
<td>-1.53</td>
<td>-20.61**</td>
</tr>
<tr>
<td>ASC_NT ($\delta_3$)</td>
<td>-2.3</td>
<td>-42.99**</td>
<td>-2.68</td>
<td>-30.66**</td>
</tr>
<tr>
<td>Cost ($\beta_1$)</td>
<td>-0.00569</td>
<td>-15.64**</td>
<td>-0.00585</td>
<td>-15.91**</td>
</tr>
<tr>
<td>Time ($\beta_2$)</td>
<td>-0.0461</td>
<td>-25.15**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IVT_CarAlone ($\beta_3$)</td>
<td>-0.0708</td>
<td>-14.93**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IVT_CarShare ($\beta_4$)</td>
<td>-0.0251</td>
<td>-6.14**</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Null log-likelihood -6555.42
Log-likelihood -5327.797
Adjusted Rho Sq. 0.187
Respondents 673
Observations 5967

Notes: t-statistics are shown relative to zero. ** indicates significance at the 5% level. The models were estimated using BIOGEMEv1.5 (Bierlaire, 2003).

### Market Segment Analysis

Building on the base models, a series of additional models were estimated to examine how tastes and preferences vary across the individuals in the sample. This was achieved through a combination of approaches involving the estimation of separate utility parameters for key segments, including: income ($Inc$), age group, car ownership (1 or 2+), gender (Female), employment status (FT or PT), geography (Scottish Executive Sixfold urban/rural Classification), whether the person they were sharing was a friend/family member or a stranger (ShareStranger), current journey time ($JT$), interest in partaking in a car pool (CarPoolInterest), and current propensity to currently car share.

In Model 3, analysis of differences and values by market segment was undertaken on the absolute values of the ASCs. This builds on Model 1 by provided a disaggregated valuation of $\delta_2$ according to key market segments. Piecewise estimation of the ASCs showed a strong relationship between choice, car ownership level, age group, gender, current journey time (continuous variable), and interest in car pooling. The preferred specification is shown in Table 8. Please note that there are 90 fewer observations in this model as 10 respondents did not report an estimate of their current journey time.
Table 8: Market Segment Analysis (Model 3)

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Coefficient</th>
<th>T-Stat</th>
<th>Implied</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASC_{CarAlone} (\delta^1_1)</td>
<td>0</td>
<td>--fixed</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>ASC_{NT} (\delta^3_3)</td>
<td>-1.92</td>
<td>-38.61**</td>
<td>58.01</td>
<td>Min</td>
</tr>
<tr>
<td>ASC_{CarShare} * CarOwn2+ (\delta^4_1)</td>
<td>-0.396</td>
<td>-7.65**</td>
<td>11.96</td>
<td>Min</td>
</tr>
<tr>
<td>ASC_{CarShare} * Female (\delta^2_2)</td>
<td>-0.0951</td>
<td>-1.76</td>
<td>2.87</td>
<td>Min</td>
</tr>
<tr>
<td>ASC_{CarShare} * Age 25-44 (\delta^3_1)</td>
<td>-0.167</td>
<td>-3.15**</td>
<td>5.05</td>
<td>Min</td>
</tr>
<tr>
<td>ASC_{CarShare} * CarPoolInterest (\delta^4_3)</td>
<td>1.85</td>
<td>8.14**</td>
<td>-55.89</td>
<td>Min</td>
</tr>
<tr>
<td>ASC_{CarShare} * JT (\delta^5_5)</td>
<td>-0.00827</td>
<td>-6.59**</td>
<td>0.25</td>
<td>Min</td>
</tr>
<tr>
<td>ASC_{CarShare} * Share (\delta^6_6)</td>
<td>-0.241</td>
<td>-4.27**</td>
<td>7.28</td>
<td>Min</td>
</tr>
<tr>
<td>Cost (\beta_1)</td>
<td>-0.00391</td>
<td>-11.31**</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Time (\beta_2)</td>
<td>-0.0331</td>
<td>-20.07**</td>
<td>8.47</td>
<td>p/min</td>
</tr>
</tbody>
</table>

£5.08 £/hour

Null log-likelihood -6456.544
Log-likelihood -5373.867
Adjusted Rho Sq. 0.166
Respondents 663
Observations 5877

Notes: t-statistics are shown relative to zero. ** indicates significant at the 5% level.

The model fit is good, the parameters are generally statistically significant and the implied value of time is just marginally lower than that recommended by the Scottish Government. In general, higher car ownership, whether they are female, being aged between 25 and 44, lengthier journey times, and having to share with a stranger means that respondents are more likely to choose a single occupancy car trip. On average, car sharing showed an 18 minute penalty over driving alone with the distribution of values shown in Figure 1a below.

5.2 Mixed Logit Specification – Panel Data Models

A random parameters or mixed logit specification offers two principal advantages over traditional multinomial logit specifications. Firstly, it accommodates taste variation across the sample by specifying the model coefficients as distributions (eg a normal distribution) rather than fixed values. Secondly, the model takes account of potential correlations in the data introduced because of the repeat observation nature of the data. In SP experiments, each respondent typically provides choice information for many scenarios (in this experiment nine) and because these choices come from the same individual they can not be treated as independent. The random parameters model can be specified to account for such panel data effects by allowing the choices of an individual to be correlated.
In Model 4, analysis was undertaken using an absolute specification of the disbenefit of car sharing (as in Model 1) with all parameters specified to be normally distributed across individuals. Table 9 shows estimates of both the mean and standard deviation of the parameters.

### Table 9: Mixed Logit Results (Model 4)

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Mean of Random Parameter</th>
<th>Std Deviation of Random Parameter</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient</td>
<td>T-Stat</td>
</tr>
<tr>
<td>ASC_{Car Alone} (δ₁)</td>
<td>0</td>
<td>--fixed---</td>
</tr>
<tr>
<td>ASC_{Car Share} (δ₂)</td>
<td>-4.2782</td>
<td>-14.80**</td>
</tr>
<tr>
<td>ASC_{NT} (δ₃)</td>
<td>-7.8466</td>
<td>-18.88**</td>
</tr>
<tr>
<td>Cost (β₁)</td>
<td>-0.0176</td>
<td>-14.87**</td>
</tr>
<tr>
<td>Time (β₂)</td>
<td>-0.1433</td>
<td>-19.94**</td>
</tr>
<tr>
<td>Share Family/Friend</td>
<td>-0.0948</td>
<td>-0.76</td>
</tr>
</tbody>
</table>

Null log-likelihood       -6555.42  
Log-likelihood            -3153.61  
Adjusted Rho Sq.          0.506     
Respondents               673       
Observations              5967      

Notes: t-statistics are shown relative to zero. ** indicates significant at the 5% level. Mixed logit model estimated with a panel specification using 5000 Halton draws.

As is typically experienced with mixed logit specification, the overall level of fit is substantially improved compared with the non-random parameters specification – an adjusted Rho squared of 0.506 compared with 0.187. The estimated parameters have the anticipated signs and magnitude and are generally estimated with a high degree of precision. The standard deviations on the random parameters indicate a considerable degree of taste variation across the sample, although the mean estimates of the value of time are broadly similar to those from Model 1. Although the mean value of sharing with family/friends is still insignificant, there is statistically significant taste variation across the sample.

Given that the estimated parameters describe distributions rather than fixed points, the estimation of the distribution of the relative attribute value (e.g. the value of time) is not straightforward. Conventional practice is to estimate the properties of the relative attribute distribution by Monte-Carlo simulation. In this instance 10,000 draws were made from the ASC distribution and the time coefficient distribution to generate a Cauchy distributed disbenefit of car sharing. Although the ratio of the mean coefficients is 29.85 minutes, the mean of the 10,000 simulated values is 56.39 minutes (median 27.51). Where the extreme values are excluded (i.e., values greater than or less than 200 minutes) the mean value estimated by Monte-Carlo simulation is 30.66 minutes. A histogram of this simulation is shown in Figure 1b. From this
analysis it can be seen that although most respondents see a disbenefit to car sharing, around 20% of respondents show a benefit from sharing.

An alternative to simulating the distribution using Monte-Carlo methods is to employ the two stage mixed logit approach outlined in Section 3.2 involving the estimation of random parameters logit models followed by an examination of individual respondent’s choices to arrive at estimates of individual specific values. The distribution of individual specific values (Figure 1c) has a mean value of 36.96 minutes (Model 5).

5.3 Latent Class Results

The latent class models proved relatively difficult to estimate with only the two class specification able to be estimated. Although deterministic, the latent specification is a relatively difficult optimisation problem with Greene and Hensher (2003) reporting that the choice of a good starting point is crucial. The estimated model is shown in Table 10 with the probability of an individual belonging to Class 1 being equal to 71%. The estimated coefficients show those in Class 1 having a low sensitivity to cost and a correspondingly high value of time, and a dislike of sharing. Individuals in Class 2 have a lower value of time (5.6 pence per minute) and are relatively indifferent to car sharing. On average, respondents show a disbenefit of car sharing equal to 35.4 minutes.

Table 10: Latent Class Analysis (Model 6)

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Latent Class 1</th>
<th>Latent Class 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASC\textsubscript{Car Alone}</td>
<td>0 --fixed---</td>
<td>0 --fixed---</td>
</tr>
<tr>
<td>ASC\textsubscript{Car Share}</td>
<td>-4.17310, -6.29</td>
<td>-0.44277, -2.64</td>
</tr>
<tr>
<td>ASC\textsubscript{NT}</td>
<td>-2.84931, -9.85</td>
<td>-3.02839, -10.26</td>
</tr>
<tr>
<td>Cost</td>
<td>-0.08706, -7.44</td>
<td>-0.09509, -8.97</td>
</tr>
<tr>
<td>Time</td>
<td>-0.00209, -1.61</td>
<td>-0.01694, -9.34</td>
</tr>
<tr>
<td>Share Family/Friend</td>
<td>0.20366, 0.67</td>
<td>-0.47712, -2.44</td>
</tr>
<tr>
<td>ASC Class</td>
<td>0</td>
<td>0.10308, 0.58</td>
</tr>
</tbody>
</table>

Null log-likelihood -6456.544
Log-likelihood -5274.52
Respondents 673
Observations 5967

Notes: t-statistics are shown relative to zero. ** indicates significant at the 5% level. Model estimated using GAUSS. Please note that this model is directly comparable with Models 1, 2 and 4 in terms of null log-likelihood and number of observations. Model 3 has a reduced number of observations, due to some respondents not recording their journey time.
Figure 1a: MNL Model Values (Model 3)

Figure 1b: MMNL Simulated Distribution (Model 4)

Figure 1c: Individual Specific Values (Model 5)
5.4 Discussion

A variety of models have been produced based upon the SP data in order to derive valuations of car sharing in terms of adjustments to the Alternative Specific Constants (ASCs) and in equivalent alternative-specific IVT factors (relative to car alone).

For each of these, further analysis has been undertaken according to key market segments, namely: income; car ownership; age group; gender; employment status; geography (Scottish Executive Sixfold Classification); sharing with a stranger; current journey time; interest in partaking in a car pool; and current propensity to car share.

All of the base models produce significant results across the modelled parameters. The accompanying recommended values and factors therefore represent a solid foundation for the modelling and appraisal of High Occupancy Vehicle (HOV) lanes in Scotland. The corresponding ASC is 26.68 minutes from the multinomial specification and approximately 30 minutes from the mixed logit specification with normal distributions on time, cost and the ASCs. Further analysis has shown that significant parameters can also be developed for market segments, including: car ownership (two cars plus households); gender; age group; current journey time; sharing with a stranger; and interest in partaking in a car pool.

The alternative model specifications however generate different distributions of the value of the disbenefit of car sharing. The mean values include 26.7 minutes for the MNL (Model 1), 18 minutes for the MNL (Model 3), 30 minutes for the MMNL (Model 4), an average of 37 minutes across individual specific values and 35.4 minutes for the latent class model. The latent class model shows the least variation across individuals and the MMNL the greatest variation and there is little correlation between the individual specific values derived from the MMNL and the MNL (Model 3).

6 Conclusions

This paper has examined respondents’ willingness-to-pay for car sharing initiatives in Scotland. A range of different functional forms and specifications were estimated to the data, each producing models with a good degree-of-fit to the dataset and significant parameters. It was hypothesised that underneath the aggregate results there would be significant variations in taste and preference, which were dependent upon underlying characteristics, or the psychographic profile, of the respondent.

Examination of market segments using three distinct techniques added invaluable insights into the data, and highlighted a number of issues associated with distribution assumptions, simulation and estimation. Perhaps most pertinently, the analysis emphasises that parameter estimates can differ significantly and are highly dependent upon model specification and any a priori assumptions made by the modeller.

Commencing with the MNL model, whilst its degree-of-fit was as expected the lowest, estimation and specification remain the most well established. The resulting valuations are neither implausible nor difficult to apply in conventional transport models. This does not though deflect from two fundamental concerns, namely

- that finite distributions do not accurately reflect taste and preference distributions amongst the population; and
• the specification of endogenous segments may miss the true drivers of variations in taste and preference.

We also see (in Figure 1a) that the finite segmented distribution method results in an unnaturally ‘lumpy’ distribution of the disbenefit of car sharing, which does not appear analogous to ‘real world’ tastes. The incorporation of psychographic profiling can begin to offset the reliance on well established market segments, eg income, but we would conclude that the MNL structure should only represent the starting point in the model estimation process.

Whilst the two mixed logit approaches described here show clear benefits in terms of their resulting distributions of valuations amongst the population, their estimation is not without difficulty. The use of Monte-Carlo simulation techniques is shown to clearly skew the mean value of the parameter unless extreme values are excluded. Defining the boundaries represents an initial source of possible error into the model estimation process. In this respect we believe that the two-stage approach offers clear benefits. Whilst its distribution is not as attractive in terms of its analogy with established functional forms, it does benefit from being readily explainable (and justifiable) to non-technical audiences and requires no assumptions and interventions on the part of the modeller. In addition, its distribution is broadly in line with that produced using simulation techniques. It is noted that the individual specific values of the disbenefit of car sharing derived from the MNL and two stage MMNL show very little correlation and a particularly pertinent advancement of existing techniques would be to examine alternative functional forms of the MMNL in more detail.

Although we were only able to estimate the two class $L-C$ model, we believe that if the techniques and associated software can be developed further it offers a substantial incremental improvement on the MNL. As Greene and Hensher (2003) have previously discussed, the differential between the $L-C$ and MMNL is less defined, and was hampered here by difficulties in estimation for $c = 3+$ classes. We believed that this is related to the requirement to commence estimation from a good starting point(s); two drawbacks then emerge, namely:

• the requirement for the ‘good starting point’ and the extent to which this prejudges valuations or effects model estimation; and

• difficulties when the attributes being investigated are more novel in nature and such starting points may not be intuitively apparent or available (such as this study).

By contrast, a particular benefit is the lack of a requirement to prejudge the characteristics and attitudes of interest. Specification of a single normally distributed population would not appear analogous to our experience of variation in taste and preference across the population, and $L-C$ models would appear well placed to capture this. In conclusion, we believe that $L-C$ models could offer significant benefits and that if the current advancements we believe to be taking place can overcome the starting point drawback, then they will be placed to advance our understanding of variations in taste and preference.

7 References


