Choosing Public Transport – Incorporating Richer Behavioural Elements in Modal Choice Models

David A. Hensher  
Waiyan Leong  
Alejandro Tirachini  
Zheng Li  
John M. Rose  
Institute of Transport and Logistics Studies  
The University of Sydney Business School  
The University of Sydney  
NSW 2006 Australia  
David.Hensher@sydney.edu.au  
Waiyan.Leong@sydney.edu.au  
Alejandro.Tirachini@sydney.edu.au  
Zheng.li@sydney.edu.au  
John.Rose@sydney.edu.au

23 August 2012(revised 13 November 2012)  
To appear in Transport Reviews

Abstract

The development of behaviourally richer representations of the role of well established and increasingly important influences on mode choice, such as trip time reliability and accounting for risk attitude and process rules, has moved forward at a fast pace in the context of automobile travel. In the public transport setting, such contributions have, with rare exception, not been considered. In this paper, we discuss and empirically illustrate the merits of advanced modelling developments aimed at improving our understanding of public transport choice, namely the inclusion of reliability in extended expected utility theoretic forms, to recognise risk attitude and perceptual conditioning, the consideration of passenger crowding and its inclusion in linear additive models; and the role of multiple heuristics in representing attribute processing as a way of conditioning modal choice. We illustrate the mechanics of introducing these behaviourally appealing extensions using a modal choice data set collected in Sydney.

Keywords: public transport, behaviour, mode choice, risk, decision weights, expected utility, heuristics, preference heterogeneity

Acknowledgments. This paper is a contribution to the research program of the Volvo Research and Education Foundation Bus Rapid Transit Centre of Excellence. We want to thank the many colleagues who have contributed to our joint research efforts in the areas of activity presented in this paper. We especially want to thank Bill Greene, Andrew Collins, David Layton, Caspar Chorus, Stephane Hess, Riccardo Scarpa, Corinne Mulley and Matthew Beck. An earlier substantially different version was presented at the Twelfth International Conference on Advanced Systems for Public Transport (www.caspt.org), Santiago, Chile, July 23-27, 2012. The comments of three referees are appreciated and have been most helpful in improving the paper.
Introduction

Behavioural advances in the study of traveller choice have outstripped the state of practice in specifying modal choice models that are used in the majority of applied public transport studies that focus on the optimisation of public transport service provision and network design. There may be good practical reasons for limiting the functional specification of the utility expressions, in the forms of generalised time or generalised cost with just travel time and out of pocket costs, typically as a linear additive form, or with each attributes separately influencing utility. However, recent developments in the study of traveller behaviour suggest a number of improvements that can make a difference to both preference revelation and consequent implications on choice response.

Recent developments of particular interest are the inclusion of reliability and crowding as extended expected utility theoretic forms, to recognise risk attitude and perceptual conditioning; and the role of multiple heuristics in representing attribute processing, as a way of conditioning modal choice. The development of these behaviourally richer representations has moved forward at a fast pace in the context of automobile travel. In the public transport setting, such contributions have, with rare exception, not been considered.

This paper presents the main elements of these contributions to highlight how they might be integrated into the next generation of utility expressions that drive the choice between available modes of transport, especially public transport alternatives, including conventional bus in mixed traffic, bus rapid transit, heavy rail and light rail. The following sections set out the essential features of each of the themes, including an empirical illustration using a single data set, followed by a set of conclusions.

Public Transport Reliability and Crowding

A number of features of existing public transport networks have come under increasingly strong criticism in recent years, especially the record of reliability of services, bus and rail, and the amount of crowding at rail stations, as well as on trains and on buses (Cantwell et al. 2009). Overcrowding is becoming an increasingly important issue for the rail industry in particular. The demand for urban passenger rail travel in Australia, for example, is at a 50 year high, and is growing at rates of between one percent and three percent per annum. In many countries, the network is at, or close to, capacity and there is limited scope to price-off demand as fares are usually regulated. For example, the average load factor (ratio between number of passengers and number of seats inside trains) measured at the Sydney central business district (CBD) station in the AM peak has increased between September 2009 and 2010 from 80-130 percent to 100-170 percent. Historically, the economic evaluation of competing investments has focussed on a narrow set of user benefits - predominantly travel time, out-of-pocket cost, and safety.

Two noticeable omissions in many transport studies are the reliability of travel time and the amount of crowding associated with public transport. Recent research in valuing reliability has found high users’ valuation for reductions in travel time variability, relative to reductions

---

\(^1\) Even though it has been recognised that travel time reliability is important to public transport users (see Rietveld et al. 2001, Cantwell et al. 2009), and hence should be given greater emphasis in transport policy, and included in patronage forecasting and appraisal studies.
in mean travel time. For example, Bates et al. (2001) found reliability ratios\(^2\) of around 1.3 for cars and no higher than 2 for public transport, whereas de Jong et al. (2009) suggest reliability ratios of 0.8 for cars and 1.4 for public transport, as agreed values in Europe. (see Li et al. 2010 for a review). On the other hand, there is growing evidence from stated preference studies that travellers are willing to pay a non-marginal sum to reduce exposure to crowding on buses and trains (see Li and Hensher 2011a for a review).

Despite good progress in measuring reliability in a car context, and rail crowding modelled as a deterministic phenomenon (see Wardman and Whelan 2011), there remains a gap in the research associated with including these two attributes, together with traditional influences on modal choice and timing of trip choice, in a trip prediction model that allows for (i) risk attitude for both reliability and crowding, (ii) probability weighting of the occurrence of varying levels of travel time and crowding over repeated commuting (and non-commuting) activity (referred to as perceptual conditioning), and (iii) preference heterogeneity across the population of travellers to capture the differing role of reliability and crowding in the population of travellers. Research by Hensher, Greene and Li (2011) in the context of trip time variability for car travel has shown how important it is to account for these behavioural drivers of travel choice, which show up through different mean and standard deviation estimates of willingness to pay (WTP). There is a need to capture these dimensions in the public transport context for reliability and also crowding.

Kahneman and Tversky (1979), developed a framework (prospect theory) within which to capture the notion of risk in decision making, through the non-linear specification of a value function \(v(x)\) to represent the weighted value of an attribute \(x\), and attribute-associated decision weights \(w(p)\), where \(w\) represents the impact of a relevant probability \(p\) on the utility of a prospect. When combined separably, \(w(p) \cdot v(x)\) defines the value \(V(x,p)\) of a prospect and results in an outcome with a probability \(p\).

Prospect theory has offered an alternative decision paradigm to (random) utility maximisation (RUM) and expected utility maximisation; however it is not without its critics (see van de Kaa 2008 and Li and Hensher 2011 for reviews), and there is appeal in staying within the Expected Utility (EUT) framework, while drawing selectively on the contributions from prospect theory that incorporate the underlying probabilities of outcomes for an attribute as a probability weighting function. Unlike EUT that uses the induced probabilities directly, prospect theory recognises that a respondent’s perception of ‘objective’ probabilities often translates into an over- or under-weighting of such probabilities. Incorporating perceptual conditioning (through decision weights) into a EUT specification of particular attributes, but staying within an overall RUM framework, offers a new variant on EU, which Hensher, Greene and Li. (2011) call Attribute-Specific Extended EUT (AS_EEUT). This framework has also been used in Hensher et al. (in press) and Li et al. (2012) in the context of car commuter choice behaviour with travel time variability.

The AS_EEUT functional form embedded within a RUM model allows for non-linearity in an attribute-specific value specification (\(\alpha\)) conditioned on probability weighting (\(w(p)\)), with the attribute of interest entering non-linearly. The specification of the attribute has an associated chance of each level occurring over R occasions (\(r=1,\ldots,R\)). The overall utility

\(^2\) The reliability ratio is defined as the ratio of the value of saving one minute of the standard deviation of travel time, to the value of reducing one minute of average travel time.
function for this representation of one attribute expressed as extended EUT, is given in equations (1) and (2).

\[
\text{AS}_\alpha \text{EEUT}(U) = \beta \{[W(P_1)x_1^{1-\alpha} + W(P_2)x_2^{1-\alpha} + \ldots + W(P_R)x_R^{1-\alpha}] / (1-\alpha)\}
\]

\[U = \text{AS}_\alpha \text{EEUT}(U) + \sum_{z=1}^{Z} \beta_z S_z\]

\(W(P)\) is a non-linear probability weighting function which converts raw probabilities \((P)\) associated with attribute \(x_1, x_2, \ldots x_R\) with \(R\) levels over \(R\) occurrences, as shown typically in a stated choice experiment; and \(\alpha\) has to be estimated, where \((1-\alpha)\) indicates the attitude towards risk. There are also a number of other variables \((S)\) in the utility expression that are not specified this way, and are added in as linear in parameters.

A number of parametric functional forms for such decision weights have been developed in the published literature together with alternative treatments of risk embedded in the value function (see Stott 2006 for a review). Until recently, traveller behaviour research has not considered these behavioural enhancements, and as far as we are aware, there is not one study in the context of public transport reliability and crowding. What has emerged from the literature is a behaviourally appealing parsimonious candidate set of functional forms. The popular non-linear decision weighting functional forms are based on evidence that individuals tend to overweight outcomes with lower probabilities, and to underweight outcomes with higher probabilities (see e.g., Tversky and Kahneman 1992; Camerer and Ho 1994). Four ‘popular’ parametric forms are given in equation (3) (Hensher, Greene and Li, 2011):

\[
w(p_o) = \frac{p_o^\gamma}{[p_o^\gamma + (1-p_o)^\gamma]^\gamma}\]

\[
w(p_o) = \frac{\tau p_o^\gamma}{[\tau p_o^\gamma + (1-p_o)^\gamma]}\]

\[
w(p_o) = \exp(-\tau(-\ln p_o)^\gamma)\]

\[
w(p_o) = \exp(-(\ln p_o)^\gamma)\]

\(w(p_m)\) is the non-linear probability weighting function; \(p_m\) is the probability associated with the \(m\)th outcome for an alternative with multiple outcomes (e.g., travel times or crowding levels over repeated trips for the same origin and destination, and start time); \(\gamma\) and \(\tau\) are the decision weighting parameters. \(\tau>0\) identifies the elevation of the weighting function, and \(\gamma>0\) represents the degree of curvature. If \(\gamma=1\), \(w(p)=p\) which implies an linear probability weighting of EUT; if \(\gamma\neq1\), then \(w(p)\neq p\) which implies the existence of non-linearity in probability weighting. Non-linear probability weighting helps to explain the Allais paradox (Allais 1953), which revealed the violation of normative EUT (i.e., the independence axiom). The four different non-linear probability weighting functions (see Equations (3a-3d)) can be applied in an EU framework in a separable manner. The presence of \(\alpha\), \(\gamma\) and \(\tau\) in equations (1) and (3) results in an embedded attribute-specific treatment in the overall utility expression associated with each alternative, that is non-linear in a number of parameters. Only if \((1-\alpha) = 1\), and \(\gamma\) and \(\tau=1\) does equation (2) collapse to a linear utility function.
Equation (3a) is an inverse S-shaped single-parameter weighting function with overweighing of low probabilities, and under-weighting of medium to high probabilities for values of $\gamma < 1$. Equation (3b) is a two-parameter form, originally proposed by Lattimore et al. (1992), which assumes a linear in a log-odds mapping between $w$ and $p$. As $\tau$ increases, we observe less overall risk aversion for gains and more overall risk aversion for losses; and as $\gamma < 1$ decreases we see greater diminishing sensitivity to probabilities around the boundaries of 0 and 1. Equations (3c) and (3d) were proposed by Prelec (1998). These forms allow for overweighing of low probabilities and underweighting of high probabilities for $\gamma < 1$, as well as the principles of sub-proportionality and sub-additivity of decision weights, two phenomena that are violated under expected utility theory, linked to the well known Allais paradox.

For the non-linear utility specification, we need to include a value function. The popular forms are the constant relative risk aversion (CRRA) model form and CARA (i.e., exponential specification). CRRA postulates a power specification (e.g., $U = x^{\alpha}$), widely used in behavioural economics and psychology (e.g., Tversky and Kahneman 1992; Holt and Laury 2002; Harrison and Rutström 2009) where $\alpha$ indicates an individual’s risk or uncertainty attitude. CRRA often delivers “a better fit than alternative families” (Wakker 2008, p.1329). In estimating the CRRA model form, a general power specification has appeal (i.e., $U = x^{1-\alpha}/(1-\alpha))$ (Andersen et al. 2009).

An example of a revised specification of the overall utility expression associated with a public transport alternative, that includes trip time reliability and crowding, can be summarised in equation (4). The equation is a representation (in the first box) of the travel time for a trip with the associated chance of that travel time occurring (distinguishing an average ($A$) time, a ‘slowest’ ($S$) time, and a ‘quickest’ time ($Q$)) as well as the levels of crowding defined as low levels ($Lcr$), average or common ($Acr$) levels, and high levels ($Hcr$) (in the second box). The other attributes might include trip cost, socioeconomic characteristics of the traveller, and other service level attributes. The representation of reliability and crowding given in equation (4) allows for non-linearity in the utility specification ($\alpha$ for risk) and probability weighting ($w(p)$). Estimation requires a non-linear generalised mixed logit form.

$$U = \beta_1[w(p_{A})A_{T}^{1-\alpha}/(1-\alpha)] + \beta_2[w(p_{Q})Q_{T}^{1-\alpha}/(1-\alpha)] + \beta_3[w(p_{S})S_{T}^{1-\alpha}/(1-\alpha)] + \beta_4[w(p_{Lcr})Lcr^{1-\alpha}/(1-\alpha)] + \beta_5[w(p_{Acr})Acr^{1-\alpha}/(1-\alpha)] + \beta_6[w(p_{Hcr})Hcr^{1-\alpha}/(1-\alpha)] + \beta_{Cost}Cost + \beta_{Socia}Socia + \ldots..$$

Given that the construct of crowding is difficult to conceptualise and measure in stated preference studies, qualitative research can provide insights. Key issues that should be tested through qualitative research include (i) defining the base level of crowding, (ii) identifying the smallest incremental change in crowding respondents can distinguish, and (iii) establishing the impact that the rolling stock/bus configuration has on perceptions of crowding. Examples of candidate specifications for standing and seating attributes are (i) the probability of getting a seat, assumed to be one minus the proportion of seats occupied (Hensher, Rose and Collins, 2011) and (ii) the density of standees per square metre (Whelan and Crockett, 2009).
Interactions between the various influencing attributes can be tested in line with suggestions that we find from the qualitative research of the type suggested above. An example of candidate interactions is set out in (5).

\begin{equation}
\text{Utility} = a(fare*D^*I^*) + a_3(D_{\text{trip}}*D^*I^*) + a_4(D_{\text{trip}}*D^*\text{CROWD}_{\text{load}})*D^*\text{CROWD}_{\text{load}}) + a_5(D_{\text{trip}}*D^*\text{CROWD}_{\text{load}})*D^*\text{CROWD}_{\text{load}})
\end{equation}

\(CROWD_{\text{load}} = \max(\text{loadfactor} - z_{\text{treatment}}, 0), \text{sx} = 1 = \text{sit}; \text{sx} = 2 = \text{standees by capacity}\)

IVT=in-vehicle time, D=distance, I=income, REL=trip reliability.

To illustrate the way the model might be estimated, within the limitations of available data, we have drawn on a study undertaken in 2009 that investigated the patronage potential of a new Metro rail system for Sydney (Hensher, Rose and Collins 2011). As part of this study, a new modal choice data set was collected to establish the role of traditional attributes such as travel times and costs, as well as trip time reliability and crowding, where the latter has a critical role in the calculation of service frequency and capacity needs of vehicles and stations. We have used a sub-set of the commuter data set where the choice was (i) between the current bus and the proposed metro, or (ii) between car and metro. Full details are given in Hensher, Rose and Collins (2011), but for clarity we present, in Figure 1, an example choice set screen, and in Figure 2 examples of public transport graphical layouts shown to respondents (in line with the combinations of attribute levels identified in a D-optimal design – see Rose et al. 2008) that depict the amount of sitting and standing in a particular mode.

The number of standees, by itself, is not a good representation of the crowding phenomenon because it does no inform about the confinement conditions of the travellers involved. The number of standees must be aligned with a measure of capacity of vehicles and carriages in order to have an indication of the level of crowding. In Figure 2, the amount of sitting and standing are set to scale to align with the capacity of each carriage or vehicle. Thus, when estimating modal choice we use the density of standees in passengers per square metre, as the attribute associated with crowding when there are passengers standing (Whelan and Crockett, 2009; Wardman and Whelan, 2011). Assuming that 4.4 pax/m² is the maximum standing capacity for public transport passengers in Sydney (based on Tirachini et al. 2012), and that 27, 120 and 125 are the maximum number of standees inside a bus, train car and metro car, respectively, the density of standees is obtained from the number of standees (NS) indicated to respondents as NS*4.4/27 for bus, NS*4.4/120 for train and NS*4.4/125 for metro.

Ideally, a non-linear treatment of crowding should also be tested in the utility function, the same as travel time variability as shown in equation (4). However, in the choice experiment that we use as an example (see Figure 1), only one level of crowding was provided, unlike travel time variability where there were three possible scenarios per alternative within a choice task (i.e., quickest, average and slowest) for car and bus. Given this empirical constraint, crowding is limited to a linear (in utility and probability weighting) form in the utility specification when modelling, albeit with an interaction with travel time. Choice experiments, where crowding is treated the same as travel time variability, should in future stated choice studies, include a range of levels such as low, average and high levels of crowding per respondent alternative with associated occurrence probabilities for each trip. These levels should be respondent-specific so as to reflect real experiences.
Figure 1: Example Choice Scenario Screen
Source: Hensher, Rose and Collins 2011
The final mixed multinomial logit (MMNL) model is summarised in Table 1. We focus the discussion on the Extended EUT elements for trip time variability. Unconstrained normal distributions are applied to the Expected Travel Time parameter. Given that the distributions for $\alpha$ and $\gamma$ are likely to be asymmetrical, skewed normal distributions are used for these two parameters. The skewed normal distribution is given as $\beta_{k,i} = \beta_k + \sigma_k V_{k,i} + \theta_k |W_{k,i}|$ where both $V_{k,i}$ and $W_{k,i}$ are distributed as standard normal. The third term is the absolute value. $\theta_k$ may be positive or negative, so the skewness can go in either direction. The range of this parameter is infinite in both directions, but since the distribution is skewed, it is therefore asymmetric.
Table 1 Application of the AS_EEUT Model Form

<table>
<thead>
<tr>
<th>Variable</th>
<th>Modes</th>
<th>Parameter</th>
<th>t-ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Non random parameters:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mode-specific constant</td>
<td>Bus only</td>
<td>-0.3808</td>
<td>-0.69</td>
</tr>
<tr>
<td>Mode-specific constant</td>
<td>Train only</td>
<td>-2.9056</td>
<td>-4.11</td>
</tr>
<tr>
<td>Mode-specific constant</td>
<td>Metro only</td>
<td>-2.7289</td>
<td>-3.91</td>
</tr>
<tr>
<td>Fare ($)</td>
<td>Bus only</td>
<td>-0.1159</td>
<td>-1.96</td>
</tr>
<tr>
<td>Fare ($)</td>
<td>Train and Metro</td>
<td>-0.2878</td>
<td>-3.91</td>
</tr>
<tr>
<td>Cost</td>
<td>Car only</td>
<td>-0.1195</td>
<td>-8.94</td>
</tr>
<tr>
<td>Travel time (T: mins)</td>
<td>Train and Metro</td>
<td>-0.0086</td>
<td>-2.41</td>
</tr>
<tr>
<td>Headway (mins)</td>
<td>All PT</td>
<td>-0.0147</td>
<td>-5.55</td>
</tr>
<tr>
<td>Crowding1 (% seated × travel time)</td>
<td>All PT</td>
<td>-0.0121</td>
<td>-1.83</td>
</tr>
<tr>
<td>Crowding2 (density of standees × travel time)</td>
<td>All PT</td>
<td>-0.0040</td>
<td>-4.92</td>
</tr>
<tr>
<td>Number of Transfers</td>
<td>All PT</td>
<td>-0.1578</td>
<td>-3.11</td>
</tr>
<tr>
<td><strong>Means for random parameters associated with trip time variability:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alpha (α)</td>
<td>Bus and car</td>
<td>0.6928</td>
<td>4.95</td>
</tr>
<tr>
<td>Gamma (γ)</td>
<td>Bus and car</td>
<td>1.3108</td>
<td>2.31</td>
</tr>
<tr>
<td>Expected Travel Time (E(T): mins)</td>
<td>Bus and car</td>
<td>-0.1571</td>
<td>-2.28</td>
</tr>
<tr>
<td><strong>Standard deviations for random parameters associated with trip time variability:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alpha (α)</td>
<td>Bus and car</td>
<td>0.9071</td>
<td>2.47</td>
</tr>
<tr>
<td>Gamma (γ)</td>
<td>Bus and car</td>
<td>0.6236</td>
<td>1.40</td>
</tr>
<tr>
<td>Expected Travel Time (E(T): mins)</td>
<td>Bus and car</td>
<td>0.1239</td>
<td>2.46</td>
</tr>
<tr>
<td>Skew normal θ for Alpha</td>
<td>Bus and car</td>
<td>0.4928</td>
<td>2.50</td>
</tr>
<tr>
<td>Skew normal θ for Gamma</td>
<td>Bus and car</td>
<td>-0.0053</td>
<td>-3.50</td>
</tr>
<tr>
<td><strong>Model Fits</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log-likelihood (β)</td>
<td></td>
<td>-1931.75</td>
<td></td>
</tr>
<tr>
<td>Log-likelihood (0)</td>
<td></td>
<td>-4358.50</td>
<td></td>
</tr>
<tr>
<td>Rho2 (relative to zero parameters)</td>
<td></td>
<td>0.557</td>
<td></td>
</tr>
<tr>
<td>Number of respondents</td>
<td></td>
<td>524</td>
<td></td>
</tr>
<tr>
<td>Number of observations</td>
<td></td>
<td>3,144</td>
<td></td>
</tr>
</tbody>
</table>

Note: For w(p), we use equation 3a

The risk attitude random parameter $\alpha$ is statistically significant, and the mean value of $(1-\alpha)$, distributed as shown in Figure 3, is less than one (with 20 of the 524 observations exceeding 1.0). For decision making where risk is associated with travel time, a value $(1-\alpha)$ less than one suggests risk-taking attitudes; and a value $(1-\alpha)$ greater than one suggests risk-averse attitudes. The empirical findings of risk attitude for the majority of commuters suggest risk-taking attitudes for the sampled respondents in the face of risky mode choices. Hence there is a high degree of risk-taking attitude present in the way that each individual evaluates the travel times on offer. As an example, given an expected travel time of 20 minutes, commuters would prefer a trip time which has a 50 percent chance of being 10 minutes and a 50 percent chance of being 30 minutes, in contrast to a 100 percent (sure) chance of the trip time being 20 minutes. This is a potentially important finding that needs to investigated further using other data sets, but it indicates that commuters feel good when they get a very good trip experience and this more than outweighs the risk of a very bad experience.
Figure 3: Distribution of the risk attitude

The parameter associated with the perceptual condition decision weights, $\gamma$, does have an impact on the shape of the weighting function (see Figure 4) in which outcomes with lower probabilities tend to be under-weighted (e.g., $w(p=0.2) = 0.14$), while outcomes with high probabilities tend to be over-weighted (e.g., $w(p=0.8) = 0.83$). The differences are however quite small in this empirical example suggesting that a linear perceptual conditioning outcome is not unreasonable.

Figure 4: Individual probability weighting function curves (MMNL)
The role of multiple heuristics in representing attribute processing as a way of conditioning modal choices

Arising from research in multiple disciplines, there is growing recognition in the choice modelling literature that contrary to the usual assumption of fixed, well-defined and context independent preferences, individuals are likely to approach a choice task through a process of preference construction, in other words, using rules and heuristics that are dependent on the choice environment. More specifically, heuristics that are defined by the local choice context, such as the gains or losses of an attribute value relative to the other attributes, seem to matter significantly.

Leong and Hensher (2011) review some of the key findings on heuristics and decision rules across the psychology, marketing, transport and environmental disciplines. Using experimental data in the context of a proposed toll road, they found that for certain components of the time and cost attributes, allowing for non-linear referencing to the least desired attribute level in the local choice set offers improvement over the standard linear-in-the-attributes and linear-in-the-parameters specification. Other heuristics can also be embedded into the model. These include the majority of confirming dimensions heuristic, which in the local choice set can be modelled as the number of ‘best’ attributes that an alternative possesses, and a heuristic linked to reference point revision, which occurs when a non status-quo alternative was chosen in the preceding choice set. An alternative approach to identifying and weighting multiple heuristics in a utility function by means of a logit-type specification for the weights of the heuristics is also appealing. We focus on this mixed heuristics model in this section, illustrating its merit in the context of the same empirical application used earlier, which is the potential patronage of a new Metro rail system for Sydney.

If it is believed that there is heterogeneity in decision processes, i.e., different respondents use different heuristics (and the possibility that the same individual uses different heuristics for different attributes within and between alternatives and choice sets), one popular approach is to appeal to probabilistic decision process models (which are essentially latent class structures) where the functional form of the heuristic under consideration is expressed through the utility expressions in each class (Hensher and Collins, 2011, McNair et al., 2011, 2012, Hess et al., 2011). Typically, each class represents one heuristic, which means that each respondent is assumed to be relying only on one heuristic. However, what that heuristic (i.e., class membership) might be for each individual can only be known up to a probability.

An alternative to the latent class model approach is to weight each heuristic directly in the utility function. Within the utility function, this approach allocates the proportional contribution of each heuristic to overall utility, with the possibility of linking this share outcome to the characteristics of respondents and other possible contextual influences. In a model with a total of \( H \) heuristics, the weights of each heuristic, denoted by \( W_h, \ h=1,2,...,H \) can be given by means of a logistic function shown in Equation (6).

\[
W_h = \frac{\exp(\sum_l \gamma_{lh}Z_l)}{\sum_{h=1}^H \exp(\sum_l \gamma_{lh}Z_l)} \tag{6}
\]
$z_l$ denotes the value of variable $l$ which is typically a socio-economic or context characteristic. $\gamma_{l_h}$ is a parameter weight that is allowed to vary according to each of the $l$ variables and each of the $m$ heuristics. To ensure identification of the model, it will be necessary to normalise, for every variable $l$, one $\gamma_{l_h}$.

As an illustration of this approach, we explore a ‘mixture’ of the linear in the parameters and linear in the attributes (LPLA) standard fully compensatory decision rule and the “non-linear worst level referencing” (NLWLR)\(^3\) heuristic. This example is very much in the spirit of Tversky and Simonson’s (1993) componential contextual model, where utility comprises a context independent effect (in this case LPLA) and a context dependent effect (in this case NLWLR). For this model, define the LPLA and NLWLR specifications (respectively as $H_1$ and $H_2$) as illustrated in Equation (7). For ease of illustrating this multiple heuristics approach, the utility function for each alternative is defined by only two attributes, which are the trip cost ($\text{cost}$), defined as the fare in the case of a public transport option, and the sum of running cost, toll cost and parking cost in the case of the car option and the travel time ($\text{TT}$).

\[
H_1 = -\beta_{\text{cost}} \cdot \text{cost}_j - \beta_{\text{TT}} \cdot \text{TT}_j
\]

\[
H_2 = \left( \beta_{\text{cost}} \cdot \text{cost}_{\text{max}} - \beta_{\text{cost}} \cdot \text{cost}_j \right)^{\phi} + \left( \beta_{\text{TT}} \cdot \text{TT}_{\text{max}} - \beta_{\text{TT}} \cdot \text{TT}_j \right)^{\phi}
\]

(7)

In the NLWLR model, respondents are assumed to make reference to the worst attribute level of each choice set. This reference may be defined as the maximum of each of the $\text{cost}$ and $\text{TT}$ attributes in the choice set, since higher levels of $\text{cost}$ and $\text{TT}$ give rise to greater disutility. Moreover, as $\text{cost}_{\text{max}}$ and $\text{TT}_{\text{max}}$ precede the minus sign, the prior expectation is for $\beta_1$ to be positive. If the NLWLR model is a better representation of choice behaviour, then the power parameter $\phi_2$ is expected to satisfy the inequality $0 < \phi_2 < 1$. This arises from one of the predictions of prospect theory, which suggests that gains in utility, relative to the reference, are best represented by a concave function.

For this example, the full dataset is used. A choice set in the data may comprise up to five alternatives. The utility functions for these alternatives can be written in the form of Equation (8).

\[
U_{\text{bus}} = \beta_{0,\text{bus}} + H_2 + \epsilon_0
\]

\[
U_{\text{train}} = \beta_{0,\text{train}} + W_1 \cdot H_1 + W_2 \cdot H_2 + \epsilon_1
\]

\[
U_{\text{metro}} = \beta_{0,\text{metro}} + W_1 \cdot H_1 + W_2 \cdot H_2 + \epsilon_2
\]

\[
U_{\text{other}} = H_2 + \epsilon_3
\]

\[
U_{\text{taxi}} = \beta_{0,\text{taxi}} + H_2 + \epsilon_4
\]

(8)

\(^3\) This model was first introduced as a contextual concavity model by Kivetz et al. (2004) who use it to model a specific phenomenon known as extremeness aversion. They make the prior assumption that relative to the worst performing attribute, utility is concave in the gains. This assumption is empirically testable and we find that it does not always hold (see Leong and Hensher, 2011). Hence, it may be more useful to label such a functional specification as a “non-linear worst level referencing” (NLWLR) model instead.
In a labelled experiment, there is potential for allowing some heterogeneity in decision rules across alternatives. In Equation (8) therefore, we have allowed the train and the metro equations to be some combination of the LPLA and NLWLR rules. The justification for this assumption might be to appeal to the observation that the train and the proposed metro share similar characteristics, hence perhaps similar decision rules might be applied for these two alternatives. At the same time, the decision rules might differ from other modes – perhaps the metro alternative might be more thoroughly evaluated since explicit attention had been drawn to the metro in the introductory screens.

We condition the heuristic weights $W_1$ and $W_2$ on two variables: the age of the respondent and the income of the respondent. In a two-heuristic model, $W_1$ and $W_2$ are given by Equation (9).

$$W_1 = \frac{\exp(\gamma_0^{H_1} + \gamma_{age}^{H_1} * \text{age} + \gamma_{inc}^{H_1} * \text{income})}{\exp(\gamma_0^{H_1} + \gamma_{age}^{H_1} * \text{age} + \gamma_{inc}^{H_1} * \text{income}) + \exp(\gamma_0^{H_2} + \gamma_{age}^{H_2} * \text{age} + \gamma_{inc}^{H_2} * \text{income})}$$

$$W_2 = \frac{\exp(\gamma_0^{H_2} + \gamma_{age}^{H_2} * \text{age} + \gamma_{inc}^{H_2} * \text{income})}{\exp(\gamma_0^{H_1} + \gamma_{age}^{H_1} * \text{age} + \gamma_{inc}^{H_1} * \text{income}) + \exp(\gamma_0^{H_2} + \gamma_{age}^{H_2} * \text{age} + \gamma_{inc}^{H_2} * \text{income})}$$

(9)

The restrictions $\gamma_0^{H_2} = -\gamma_0^{H_1}$, $\gamma_{age}^{H_2} = -\gamma_{age}^{H_1}$ and $\gamma_{inc}^{H_2} = -\gamma_{inc}^{H_1}$ were imposed for identification.

In a model where $\gamma_{age}^{H_1}$ and $\gamma_{inc}^{H_1}$ are assumed homogenous in the sample, the heuristic weights $W_1$ and $W_2$ will still differ across respondents following the variations in the socio-economic characteristics. Table 2 reports the results of the estimation for a fixed parameters model. This model, which combines the LPLA and NLWLR rules, at the cost of estimating three additional parameters ($\varphi_1$, $\varphi_2$ and $\gamma_0^{H_1}$) shows a substantial improvement in fit compared to the typical MNL specification with the socio economic characteristics entered in the usual way. The parameters for the $TT$ and $cost$ attributes are significant at the one percent level and of the correct sign.
Table 2: Estimation of a Weighted LPLA and NLWLR Decision Rules in Utility

<table>
<thead>
<tr>
<th></th>
<th>$\hat{\beta}$ (z-ratio)</th>
<th>$\hat{\phi}$ (z-ratio)</th>
<th>$\hat{\gamma}$ (z-ratio)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Travel time (TT) (minutes)</td>
<td>0.0217 (9.40)</td>
<td>0.424 (6.21)</td>
<td></td>
</tr>
<tr>
<td>Cost ($)</td>
<td>0.0418 (3.56)</td>
<td>0.695 (8.10)</td>
<td></td>
</tr>
<tr>
<td>Alternative Specific Constants</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- bus</td>
<td>-0.644 (-9.73)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- train</td>
<td>-0.447 (-9.89)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- taxi</td>
<td>-1.405 (-8.85)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heuristic constant</td>
<td></td>
<td>-0.871 (-4.28)</td>
<td></td>
</tr>
<tr>
<td>Age (years)</td>
<td></td>
<td>0.0168 (4.67)</td>
<td></td>
</tr>
<tr>
<td>Income ($'000)</td>
<td></td>
<td>-0.00567 (-3.51)</td>
<td></td>
</tr>
<tr>
<td>No. of observations</td>
<td></td>
<td>6,138</td>
<td></td>
</tr>
<tr>
<td>LL</td>
<td></td>
<td>5120.95</td>
<td></td>
</tr>
<tr>
<td>LL(0)</td>
<td></td>
<td>-9878.73</td>
<td></td>
</tr>
<tr>
<td>LL (MNL)</td>
<td></td>
<td>-5163.46</td>
<td></td>
</tr>
</tbody>
</table>

Turning to the heuristic weights, the partial derivatives of $W_m$ with respect to each of its $l$ arguments are functions that take the same sign as $\gamma_{lm}$. Hence, the estimation results show that $W_1$, which is the weight of the LPLA heuristic, is, all else equal, lower than $W_2$, the weight of the NLWLR heuristic. $W_1$ is increasing in age and decreasing in income, with the opposite effect for $W_2$. These are interesting findings as they demonstrate a relationship between the use of a heuristic and the socio economic characteristics of a respondent. A multiple heuristics approach suggested herein may in fact be a preferred way of accounting for the impact of socio-economic characteristics on decision making.

Conclusions

This paper presents a number of themes that have been found in contexts outside of public transport modelling, planning and operations research, to add important behavioural relevance to how we explain choice making. Although we can always include additional attributes into the commonly used linear-in-parameters preference expressions associated with public transport modal alternatives in choice models (such as measures of reliability and crowding), new evidence from the extensive literature on choice modelling suggests that we must do more than simply add, in a linear fashion, additional attributes.

This paper has highlighted some developments that appear to make a difference in the overall fit of mode choice models, and which are expected to increase predictive performance in a behaviourally appropriate way, in contrast to calibrating base year models which might reproduce known market activity but may not be particularly good at predicting behavioural responses when we introduce policy shocks to the public transport system. In ongoing
research we recommend a fuller integration of the functional forms in equation (4), (8) and (9).

The challenge facing the research community in public transport demand and choice modelling is to collect appropriate data from a sample of travellers that can test for, and hence account for, the various behavioural propositions presented in this paper. We also remind analysts that these behaviourally appealing enhancements on the demand side require suitable supply side data to enable the application of these enhanced models to real planning and evaluation problems.

References


