# AN ACCUMULATION OF PREFERENCE: TWO ALTERNATIVE DYNAMIC MODELS FOR UNDERSTANDING TRANSPORT CHOICES

**Thomas O. Hancock (Corresponding Author)**
Choice Modelling Centre and Institute for Transport Studies
34-40 University Road
University of Leeds
Leeds
LS2 9JT
United Kingdom
tstoh@leeds.ac.uk

**Stephane Hess**
Choice Modelling Centre and Institute for Transport Studies
University of Leeds
S.Hess@leeds.ac.uk

**Charisma F. Choudhury**
Choice Modelling Centre and Institute for Transport Studies
University of Leeds
C.F.Choudhury@leeds.ac.uk

## ABSTRACT

Interest in behavioural realism has gradually led to the introduction of alternatives to random utility maximisation (RUM) as a paradigm for discrete choice models, with notable interest for example in random regret minimisation (RRM). These models have however continued to rely on a framework where a single *value function* of some form is calculated once for each alternative in each choice setting, and the choice probabilities are calculated by comparing these value functions across alternatives. In contrast, research in mathematical psychology has used a more dynamic approach, where the preference value of each alternative updates over time within a single choice process while the decision maker is deliberating about the choice to make. These *accumulator* models are well suited to accommodating a variety of context effects, and have been shown to give good performance for data collected in laboratory-based settings. The present paper considers two such accumulator models, namely decision field theory (DFT) and the multi-attribute linear ballistic accumulator (MLBA), and makes a number of methodological improvements to address limitations that have thus far prevented their use in travel behaviour research. This includes the ability to capture the influence of socio-demographics, the presence of underlying preferences for specific alternatives, or dealing with attributes that have opposite effects on choice probabilities. We offer what we believe to be the first in-depth simultaneous comparison of DFT and MLBA with typical discrete choice models, and also for the first time test both DFT and MLBA on a revealed preference dataset. We find that both models outperform typical RUM and RRM implementations for both estimation and out-of-sample prediction across our datasets, including in a large scale simulation experiment.

## 1. INTRODUCTION

Whilst mainstream choice modelling has been grounded in firm economic foundations [McFadden, 1974], attempts to understand decision-making behaviour in other fields has been implemented with very different aims and objectives. Since work in the 1970s [Tversky, 1972, Tversky and Kahneman, 1973, Tversky, 1977], the field of behavioural economics has considered choice from an economic viewpoint whilst simultaneously demonstrating that decision-makers are subject to biases, heuristics and context effects that result in choices being made that are not the most likely under traditional choice models. Choice modellers have long had an interest in increasing the behavioural realism of their models, with recent methodological advances aimed at incorporating alternative behavioural ideas such as random regret minimisation (RRM) [Chorus et al., 2008, Chorus, 2010], the incorporation of heuristics [Swait, 2001] and satisficing [González-Valdés and Ortúzar, 2017].

Moving away from the traditional random utility maximisation (RUM) framework however entails a number of disadvantages, notably an inability to perform welfare analysis. This means that careful consideration is required before we move to alternative models. In this context, the question then arises whether, if we are willing to move away from RUM, we should move to models that are substantially different from it, rather than still staying within a logit framework as is the case for random regret minimisation [Hess et al., 2018]. This observation leads us to look further afield, and in particular at the work in mathematical psychology, where researchers have tended to try and build models to mathematically represent context effects such as the attraction, compromise and similarity effects [Roe et al., 2001, Trueblood et al., 2013b, Noguchi and Stewart, 2014] as well as decision-making under time pressure [Busemeyer and Townsend, 1993].

It is notable that very few papers as of yet have tested whether models developed in mathematical psychology can be used for predicting choices in general (i.e. outside laboratory settings). Some notable exceptions include Hawkins et al. [2014] who applied the linear ballistic accumulator (LBA, Brown and Heathcote 2008) to consumer attitudes and patient preferences and Berkowitsch et al. [2014], who applied decision field theory (DFT) to consumer choices for products such as computers, cameras and racing bicycles. In key comparisons against *traditional* choice models, DFT in particular has been found to outperform random utility and random regret based models (e.g. Berkowitsch et al. [2014], Hancock et al. [2018]). Other models from mathematical psychology are yet to be put to the test in such a rigorous manner.

DFT and the similarly popular (in mathematical psychology) multi-attribute linear ballistic accumulator (MLBA) [Trueblood et al., 2013a, 2014] differ from more traditional discrete choice models in one specific dimension. RUM and RRM models are characterised by their utility and regret functions respectively, which are used to calculate a single *value function* for each alternative, where comparison of this across alternatives then leads to probabilities of a given alternative being chosen. This value function is calculated once per choice situation. On the other hand, DFT and MLBA are members of a broad family of *accumulator* models, where the preference values for an alternative in a single choice context are not static but are updated over time. It is important to note that this is different from work looking at preferences evolving over a sequence of choices, such as models incorporating value learning [McNair et al., 2012], state dependence [Bruno et al., 2015] or dynamic discrete choice models [Liu and Cirillo, 2018]. *Accumulator* models are structures for internal preference accumulation at the level of every single choice, not models that accumulates evidence over a sequence of choices. The accumulation models thus capture the mental deliberation from the time a particular choice is faced (or stated choice scenario presented) to the point where the choice is made. The preferences are reset after that, so the accumulation effect is not carried over to the next choice task, i.e. the accumulation made for choice $t$ does not affect choice $t+1$ although such extensions are possible too.

Under DFT, the decision maker updates his/her preference for given alternatives by repeated comparisons between them where the attribute values of the alternatives in that situation remain constant across these comparisons. Under MLBA, a 'drift rate' is generated for each alternative allowing the preference values to update within a single choice context. Thus far, there has, to the best of our knowledge, not been any application of MLBA to transport data and only a few, mainly theoretical, applications of DFT. The way in which preferences evolve over time and their inherent ability to accommodate a range of what economists might call behavioural anomalies however make these models at first hand very appealing for studying travel behaviour. For example, DFT conceptually should be an appropriate model for dealing with a variety of travel situation effects including situational dynamics, types of travel, cultural habits and societal norms [Stern and Richardson, 2005]. Additionally, DFT has been combined with the Queuing Network-Model Human Processor to model a driver's speed control [Zhao et al., 2011]. It has also been demonstrated that DFT accurately predicts the share of participants who choose park and ride, car, bus or subway [Qin et al., 2013], although this study only considered a single choice set.

The large and rich datasets typically found in transport have meant that computational limitations have until now limited the use of DFT in transport applications [Otter et al., 2008]. Our previous work on DFT has focused on methodological improvements that have made it possible to rigorously test DFT against typical choice models [Hancock et al., 2018]. This motivates us to investigate the suitability of MLBA in modelling travel behaviour as well, as it has been found to outperform DFT in mainstream mathematical psychology literature [Trueblood et al., 2014, Cohen et al., 2017, Turner et al., 2017][1].

Beyond simply comparing the two structures, we make a number of methodological improvements to both DFT and MLBA to facilitate their use on rich multi-alternative multi-attribute datasets. The key contribution relates to allow analysts to use DFT with attributes that have opposite effects on choice probabilities, and where this directionality is not known a priori. Previously, DFT models included 'attention weights' which could be used to capture the relative importance of attributes. As these weights must be positive (and sum to one), a priori knowledge is required as to whether an attribute has a positive (e.g. comfort of journey) or negative (e.g. cost) impact on the likelihood of an alternative being chosen. This is particularly an issue for consumer attributes which some decision-makers may like and others dislike, such as the size of a car. We propose the use of attribute-specific scaling coefficients, meaning that such a priori knowledge is no longer

[1] Furthermore, there has been increasing attention in transport on best-worst datasets [Giergiczny et al., 2013, Rose, 2014] and research in mathematical psychology has shown that the linear ballistic accumulator (a simpler form of the model, where each alternative has a mean drift rate simply equal to an alternative-specific constant), performs well for these datasets [Hawkins et al., 2014].

required. We show that these coefficients can also be added to MLBA to capture the relative importance of different attributes, a feature not typically accounted for in standard MLBA implementations. Further improvements include the ability to capture the influence of socio-demographics and the presence of underlying preferences for specific alternatives, in a manner equivalent to alternative specific constants in typical discrete choice models. We also look in detail at identification issues for both models, with a number of empirical tests to help inform future applications.

In our empirical work, we offer what we believe to be the first in-depth simultaneous comparison of DFT and MLBA with typical discrete choice models, and also for the first time test both DFT and MLBA on a revealed preference dataset. We find that both models outperform typical RUM and RRM implementations for both estimation and out-of-sample prediction across our datasets, including in a large scale simulation experiment.

The remainder of this paper is organised as follows. In the next section, we first provide an overview of the two models in their current form before presenting our various methodological improvements. This is followed by our empirical work on stated choice and revealed preference data, before some further tests on simulated data. The final section summarises the findings and presents some directions for future research.

## 2. METHODOLOGY: CONTRASTING AND IMPROVING MODELS FROM MATHEMATICAL PSYCHOLOGY

In this section, we first provide an introduction to accumulator models and the state-of-the-art implementations of DFT and MLBA[2]. This is followed by our various methodological improvements.

### 2.1. Introduction to accumulator models

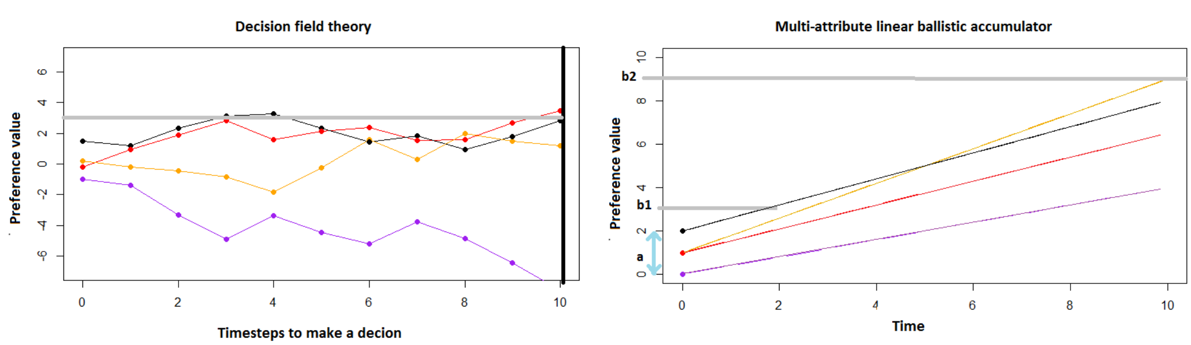
Since the introduction of the drift diffusion model [Ratcliff, 1978], many different variations of sequential sampling models (or accumulation models) have been developed by mathematical psychologists [Busemeyer and Townsend, 1992, Usher and McClelland, 2001, Krajbich et al., 2012]. The idea of a sequential sampling model is that preferences for alternatives update over time depending on what information is being considered. An individual may consider, for example, cost, before then considering travel time. They might make comparisons across alternatives sequentially or randomly. By contrast, mainstream choice models such as random utility or random regret models estimate just a single preference or utility value for each alternative given a set of attribute values and then use that value to calculate choice probabilities. Critically, accumulation models instead assume that these preferences change over the course of the deliberation process whilst the decision-maker is choosing an alternative (even if the attributes of the alternatives stay the same). As already highlighted in the introduction, this preference accumulation is internal and happens at the level of every single choice, i.e. it is not an accumulation over a sequence of choices. This thus allows us to contrast the models to typical discrete choice structures.

[2]For further details and a more comprehensive description of decision field theory, please refer to Section 2 in Hancock et al. [2018].

These models aim to 'understand the motivational and cognitive mechanisms that guide the deliberation process involved in decisions' [Busemeyer and Townsend, 1993]. Accumulator models have subsequently been shown to resemble neural activity. For example, Gold and Shadlen [2000] found that during a motion perception task, there was an accumulation of sensory evidence in the neural circuits of a monkey's brain, creating a behavioural response when the appropriate amount of information had been received. Furthermore, accumulator models have been demonstrated to predict contextual effects [Hotaling et al., 2010, Trueblood et al., 2014], capture risky choice behaviour [Busemeyer and Townsend, 1993, Stewart and Simpson, 2008] and can predict preference reversals [Diederich, 2003]. Additionally, dynamic models provide a naturalistic method for the modelling of decision making in dynamic choice settings [Holmes et al., 2016].

One popular model from mathematical psychology that can easily be compared to traditional choice models is decision field theory (DFT), first introduced by Busemeyer and Townsend [1992, 1993] and first operationalised in the context of travel behaviour by Hancock et al. [2018]. In a DFT model, preference values for the alternatives update stochastically over time. At each moment, an attribute is compared across alternatives and a valence (momentary preference) is added to the preference value for each alternative. At some point, the decision-maker comes to a conclusion, either as one of the alternatives reaches some threshold (similar to satisficing [Kaufman, 1990, Schwartz et al., 2002, González-Valdés and Ortúzar, 2017]) or as an external cue forces the decision-maker to make a choice, in which case the decision-maker chooses the alternative with the highest preference value at that moment. As an example, the left panel in Figure 1 demonstrates that different alternatives may be chosen depending on which threshold is used. The first alternative to reach the internal threshold value is outperformed by another alternative if the decision is not made until the time threshold applies. Here, it should be noted that the value that evolves over comparison is a *preference value*, rather than a probability, where the latter is calculated from the expectation of the former. The horizontal axis is measured in *timesteps*, which relate to the number of comparisons between alternatives, each time using one attribute.

**FIGURE 1** : An example decision process under both accumulation models




The linear ballistic accumulator model [Brown and Heathcote, 2008] and its multi-attribute version MLBA [Trueblood et al., 2013a, 2014] have a similar accumulation process for the preference of alternatives, but, in contrast with DFT, the updating is not stochastic. Instead, decision-makers start with some random amount of initial 'evidence' for each alternative, that then 'drifts' until one of the alternatives reaches a threshold. These preference values grow linearly at some drift rate dependent on the attributes of the alternative. Depending on the level of the threshold, different alternatives may be chosen. This is demonstrated in the right panel of Figure 1, in which the alternatives start with some random initial value, which we show as an interval $a$, and different alternatives are chosen if the threshold value is $b_1$ or $b_2$. The linear drift rates imply that, once the alternative with the largest drift rate value 'gains the lead', unlike in DFT, there is no way for another alternative to recover and be chosen. Whilst this would not be the case with a non-linear specification, the current model is specifically linear to allow for simple calculation of the probabilities of alternatives. Of course, different alternatives can be chosen depending on the length of the deliberation process. As with DFT, the value that evolves over time is a preference value, while the horizontal axis in Figure 1 now relates to actual time, given that no additional comparisons are made.

The mathematics underlying MLBA and DFT is vastly different. LBA was specifically designed such that it is 'simple' [Brown and Heathcote, 2008] and mathematically tractable, with MLBA subsequently developed such that it can also accurately capture and predict context effects. The simpler mathematical nature means that the probabilities of alternatives can easily be calculated from a combination of normal and uniform cumulative density functions (see Section 2.2.2 for a full description of MLBA).

It may be noted that there are numerous other accumulation models from mathematical psychology that are able to explain choice processes and predict choices. However, not all are currently suitable for transitioning into applied choice modelling. Given the complex nature of revealed preference datasets with many alternatives, even implementations without random coefficients will impose large computational costs. This is then further increased if analysts wish to add random heterogeneity in preferences, and models from mathematical psychology thus need to be efficient to run at a basic level if they are to compete. This means that models that do not have analytical solutions for calculating the probability of alternatives will likely not be suitable options. For example, the leaky competing accumulator model (LCA, Usher and McClelland 2001), would require two levels of simulation if we wished to calculate the likelihood for an LCA model with random parameters. Additionally, difficulties in parameter recovery for the LCA model [Miletić et al., 2017] make it unlikely for this model to provide a viable alternative to typical choice models. Requirements of computer intensive simulation are also issues for the associative accumulation model [Bhatia, 2013], the attentional drift diffusion model [Krajbich et al., 2012] and also a version of DFT where the consideration process stops upon one of the alternatives reaching a 'preference' threshold. We avoid this situation by instead using 'external' thresholds, for which the probability with which each alternative is chosen can be analytically calculated.

## 2.2. State-of-the-art implementations of decision field theory (DFT) and the multi-attribute linear ballistic accumulator model (MLBA)

### *2.2.1. Decision field theory*

#### 2.2.1.1 Theory

Decision field theory, as an accumulator model, has preference values for each alternative that update over time[3] (see left panel in Figure 1).

$$P_t = S \cdot P_{t-1} + V_t \tag{1}$$

The previous values, $P_{t-1}$, are multiplied by a feedback matrix, $S$, and a valence vector $V_t$ is added. The feedback matrix has two parameters that control for the impact of attraction, similarity and compromise effects [Roe et al., 2001, Hotaling et al., 2010, Noguchi and Stewart, 2014], and is defined as:

$$S = I - \phi_2 \times \exp(-\phi_1 \times D^2) \tag{2}$$

where $\phi_1$ is a sensitivity parameter, $\phi_2$ is a memory parameter and $D$ is the distance between the attributes of the alternatives. Whilst the relative importance of the different attributes can be taken into account with a psychological distance function [Hotaling et al., 2010] and work on new distance functions is possible [e.g. Berkowitsch et al., 2015], the Euclidean distance between the attributes can also be used for simplicity [Qin et al., 2013]. The sensitivity parameter, $\phi_1$, affects how much the alternatives compete with each other. Values very close to zero results in the distance between the attributes of alternatives becoming less important, whereas higher values result in more competition between similar alternatives. The memory parameter (also known as the decay parameter) determines the relative importance of attributes considered towards the end of the decision process relative to those considered at the start. A value of one results in zeros on the diagonals of the feedback matrix, which results in the preference already accumulated becoming irrelevant. As this value tends towards zero, the importance of the already accumulated preference increases.

At each timestep, a DFT model assumes that the decision-maker compares a single attribute across all of the alternatives. This results in a random valence vector at time t, $V_t$, which can be calculated as:

$$V_t = C \cdot M \cdot W_t + \varepsilon_t \tag{3}$$

where $C$ is a contrast matrix used to rescale the values such that they total zero, $M$ is the matrix of attribute values and $W_t = [0..1..0]'$ with the $k^{th}$ entry, i.e. $W_{t,k} = 1$ if and only if attribute $x_k$ is the attribute being attended to by the decision-maker at timestep $t$. A DFT model thus estimates a weight, $w_k$, for the likelihood of attribute $x_k$ being the single attribute attended to a given timestep, where $\sum_k w_k = 1$. There is also a random error vector, $\varepsilon_t = [\varepsilon_1..\varepsilon_n]'$, with $\varepsilon_i \sim N(0,s)$, identically and independently distributed across alternatives, time and individuals. This allows for flexibility in the range of probability values that DFT predicts. This is in essence an error or noise parameter [Roe et al., 2001], for which higher values would be expected for more complex decision-making tasks [Hotaling et al., 2010].

#### 2.2.1.2 Estimation

It has been demonstrated that the probabilities with which each alternative is chosen after $t$ preference accumulation steps can be calculated with the expected value and the covariance of

[3] Note that whilst it is possible that separate choice tasks could be linked through parameters controlling for learning effects, all of the DFT models in this paper assume that all choice tasks are entirely independent of each other to make them comparable with the RUM and RRM models without state-dependence.

the preference values ($\xi_t$ and $\Omega_t$) [Roe et al., 2001]. This results in the stochastic variation being averaged out such that probabilities of the alternatives can be calculated without the requirement of computationally heavy simulation.

To calculate the expected value of the preference values, we must first expand Equation 1, which results in:

$$P_t = \sum_{r=0}^{t-1} S^r \cdot V_{t-r} + S^t \cdot P_0 \tag{4}$$

where $P_0$ is the initial preference vector. This is often assumed to be a vector of zeros [Busemeyer and Diederich, 2002] but can also be used to capture underlying preferences for different alternatives [Hancock et al., 2018].

The attribute weights $w_k$ are stationary, therefore $W_t$ can be considered a stationary stochastic process. This means that $V_t$ is also a stationary stochastic process with mean $E[V_t]$ and a variance covariance matrix given by $Cov[V_t]$. We let $\varepsilon_t$ vary according to a normal distribution with mean zero and variance $\sigma_\varepsilon^2$. Thus, given that $\mu = E[V_t]$, it can be calculated as $\mu = C \cdot M \cdot w$, where $w$ is a vector containing the attribute attention weights, $w_k$, which corresponds to the probability of each of the attributes being considered. We also have $Cov[V_t] = \Phi = C \cdot M \cdot \Psi \cdot M' \cdot C' + \sigma_\varepsilon^2$, where $\Psi = Cov[W_t]$ (C and M are matrices of constants). We can then calculate the expected value and the covariance of $P_t$. With S being a constant, $E[P_t]$ reduces to:

$$E[P_t] = \xi_t = \sum_{r=0}^{t-1} S^r \cdot \mu + S^t \cdot P_0 \tag{5a}$$

$$= (I-S)^{-1}(I-S^t) \cdot \mu + S^t \cdot P_0 \tag{5b}$$

We can also now calculate the covariance of the preference values:

$$Cov[P_t] = \Omega_t = Cov\left[\sum_{r=0}^{t-1} S^r \cdot V_{t-r} + S^t \cdot P_0\right] \tag{6a}$$

$$= \sum_{r=0}^{t-1}\left[S^r \cdot \Phi \cdot S^{r'}\right] \tag{6b}$$

The resulting calculations are complex, but as shown in our earlier work, we can further simplify $Cov[P_t]$ such that we can avoid requiring a summation [Hancock et al., 2018]. We replace the feedback matrices with a matrix Z of size $J^2 \times J^2$ (where J is the number of alternatives) and reshape $\Phi$ (with entries $p_{i,j}$) into a column matrix, $\overline{\Phi}$:

$$Z = \begin{bmatrix} z_{1,1} & z_{1,2} & \cdots & z_{1,J^2} \\ z_{2,1} & z_{2,2} & \cdots & z_{2,J^2} \\ \vdots & \vdots & \ddots & \vdots \\ z_{J^2,1} & z_{J^2,2} & \cdots & z_{J^2,J^2} \end{bmatrix}, \overline{\Phi} = \begin{bmatrix} p_{1,1} \\ p_{2,1} \\ \vdots \\ p_{J,1} \\ p_{2,1} \\ \vdots \\ p_{J,J} \end{bmatrix} \tag{7}$$

As S is a symmetric matrix, we can then define Z by setting it as the Kronecker product of S with itself: $z_{i,j} = s_{(i \bmod J, j \bmod J)} \cdot s_{(\lceil i/J \rceil, \lceil j/J \rceil)}$. The calculation of the covariance of $P_t$ now simplifies to:

$$Cov[P_t] = \Omega_t = \sum_{r=0}^{t-1} \left[ S^r \cdot \Phi \cdot S^{r'} \right] \tag{8a}$$

$$= \sum_{r=0}^{t-1} \left[ Z^r \cdot \overline{\Phi} \right] \tag{8b}$$

$$= (I - Z)^{-1} (I - Z^t) \overline{\Phi} \tag{8c}$$

These succinct forms for $\xi_t$ and $\Omega_t$ mean that we can now calculate the probability with which each alternative is chosen. On the basis of the multivariate central limit theorem, $P_t$ converges to the multivariate normal distribution [Roe et al., 2001]. This means that if a time threshold is reached, the probability of choosing alternative $A$ from a set of $n$ alternatives at time $t$ is:

$$Prob\left[ \max_{i \in n} P_t[i] = P_t[A] \right] = \int_{X>0} exp\left[ -(X-\Gamma)'\Lambda^{-1}(X-\Gamma)/2 \right] / (2\pi|\Lambda|^{0.5}) dX \tag{9}$$

with $X = [P_t[A] - P_t[B], \ldots, P_t[A] - P_t[n]]'$, $\Gamma = L\xi_t$, $\Lambda = L\Omega_t L'$ where

$$L = \begin{bmatrix} 1 & -1 & 0 & \ldots & \ldots & 0 \\ 1 & 0 & -1 & \ddots & & \vdots \\ 1 & \vdots & \ddots & \ddots & \ddots & \vdots \\ 1 & \vdots & & \ddots & -1 & 0 \\ 1 & 0 & \ldots & \ldots & 0 & -1 \end{bmatrix} \tag{10}$$

with L being a matrix constructed with a column vector of 1s and a negative identity matrix of size $n-1$ where $n$ is the number of alternatives. The column vector of 1s is placed in the $i^{th}$ column where $i$ is the chosen option.

### *2.2.2. Multi-attribute linear ballistic accumulator*

#### 2.2.2.1 Theory

Under MLBA, each alternative has a value that linearly grows towards a threshold (see right panel in Figure 1). The chosen alternative in an MLBA model is the first alternative to pass a threshold value, $\chi$. There are two components in this process; the start points and the drift rates.

Start points for each of the alternatives are drawn separately from a uniform distribution $U[0,A]$ where A is estimated. For example, Figure 1 demonstrates what a decision might look like if the start points are drawn from a distribution $U[0,2]$. A different value $A_j$ could be estimated for each alternative $j$, although it is common practice [Trueblood et al., 2014] to assume that all alternatives have starting values that are drawn using the same estimate $A$.

Trueblood et al. [2013a, 2014] demonstrate that there are different methods for specifying drift rates for an MLBA model such that they explain context effects. In this application, however, we choose to fit versions similar to the mainstream version of MLBA [Trueblood et al., 2014] as this outperforms the first version (described by Trueblood et al. [2013a]) for our two basic route choice datasets (see appendix A). Under MLBA, we define the drift rates for the different alternatives as independent draws from normal distributions (truncated above zero), where, for alternative $j$, we have the drift rate $D_j$ given as:

$$D_j \sim TN\left(d_j, s_j\right) \tag{11}$$

with mean drift rate $d_j$ and standard deviation $s_j$. Typically, the standard deviation is set to be the same value for all alternatives, i.e. $s_j = s, \forall j$, but a different value could be estimated for each drift rate [Trueblood et al., 2013b].

In the current version of MLBA, mean drift rates follow the specification used by Trueblood et al. [2014]:

$$d_j = v_j + I_0 \tag{12}$$

where $I_0$ is a positive constant (which can be specified such that all drift rates have a positive mean) and $v_j$ is a value function, similar to random regret minimisation in that it compares an alternative $j$ against all other alternatives $i$ across each attribute $x$. Specifically, with $K$ attributes, we have that:

$$v_j = \sum_{j \neq i} \sum_{k=1}^{K} \left(w_{x_{k,i,j}} \cdot (x_{k,j} - x_{k,i})\right). \tag{13}$$

In this notation, $x_{k,i}$ is the value for the $k^{th}$ attribute for alternative $i$, and $w_{x_{k,i,j}}$ is a weight for attribute $x_k$ and alternative pairing $i$ and $j$, which relates to the similarity between them[4]. In particular, it is defined such that it is an exponential decaying function of distance, with:

$$w_{x_{k,i,j}} = exp(-\lambda \cdot |x_{k,i} - x_{k,j}|) \tag{14}$$

Two different values of $\lambda$ are used depending on whether the difference between $x_{k,i}$ and $x_{k,j}$ is positive or negative:

$$\lambda = \begin{cases} \lambda_1, & \text{if } x_{k,j} \geq x_{k,i}. \\ \lambda_2, & \text{if } x_{k,j} < x_{k,i}. \end{cases} \tag{15}$$

This feature can capture differences between the subjective similarity between A and B and the subjective similarity between B and A, which may not be equal [Tversky, 1977], with gains and losses regularly having been shown to be treated differently in a transport context [Hess et al., 2008, Masiero and Hensher, 2010, Stathopoulos and Hess, 2012].

It is worth noting that MLBA [Trueblood et al., 2014] was adapted from the original version [Trueblood et al., 2013a] to additionally translate attribute values into ‘subjective values’. In their example, they had two similar attributes: testimony strength of eyewitness P and testimony strength of eyewitness Q. A parameter was then introduced such that an ‘indifference curve’ could be calculated to avoid issues of extremeness aversion [Chernev, 2004], where, for example, values of 50-50 might be preferred to 70-30. Given that indifference curves cannot be so simply constructed

[4] Note that Equation 13 is equivalent to Equation 3 in Trueblood et al. [2014], but for multiple alternatives and multiple attributes.

in typical transport choice tasks, we do not detail the additional parameters used to translate the attribute values into subjective values here, noting that we instead require some measure to translate the different attributes appropriately such that the relative importance of the attributes is accounted for. We discuss this further in Section 2.3.3.

#### 2.2.2.2 Estimation

If we have values (either estimated or fixed) for the drift rates of the alternatives and for the start and end points ($A$ and $\chi$ respectively), we can calculate the probability of each alternative's accumulator being the first to finish, i.e. for its value function to exceed the threshold $\chi$ before any others do [Brown and Heathcote, 2008].

The amount of evidence that needs to be accumulated for an alternative to reach the threshold is U[$\chi$-A,$\chi$] (assuming $\chi > A$). Given an alternative's drift rate distribution, $D_j$, the cumulative distribution function for the time taken for the accumulator associated with alternative $j$ is given by:

$$F_j(t) = Prob\left(\frac{U[\chi - A, \chi]}{D_j} < t\right) \tag{16}$$

Brown and Heathcote [2008] demonstrate that for a mean drift rate following a normal distribution[5], this reduces to:

$$\begin{aligned} F_j(t) = 1 + \frac{\chi - A - t\cdot D_j}{A}\cdot\Phi\left(\frac{\chi - A - t\cdot D_j}{t\cdot s}\right) - \frac{\chi - t\cdot D_j}{A}\cdot\Phi\left(\frac{\chi - t\cdot D_j}{t\cdot s}\right) \\ + \frac{t\cdot s}{A}\cdot\phi\left(\frac{\chi - A - t\cdot D_j}{t\cdot s}\right) - \frac{t\cdot s}{A}\cdot\phi\left(\frac{\chi - t\cdot D_j}{t\cdot s}\right) \end{aligned} \tag{17}$$

where $\phi$ and $\Phi$ are the standardised normal distribution's density and cumulative density functions, respectively. The associated probability density function is then:

$$\begin{aligned} f_j(t) = \frac{1}{A}\Big[ -D_j\cdot\Phi\left(\frac{\chi - A - t\cdot D_j}{t\cdot s}\right) + D_j\cdot\Phi\left(\frac{\chi - t\cdot D_j}{t\cdot s}\right) \\ + s\cdot\phi\left(\frac{\chi - A - t\cdot D_j}{t\cdot s}\right) - s\cdot\phi\left(\frac{\chi - t\cdot D_j}{t\cdot s}\right)\Big] \end{aligned} \tag{18}$$

To then calculate the probability of a given alternative $j$ being chosen[6], we need to calculate the probability density function of alternative $j$ reaching the threshold $\chi$ before all other alternatives $i \neq j$:

$$PDF_j(t) = f_j(t)\prod_{i\neq j}(1 - F_i(t)) \tag{19}$$

Thus we have:

$$Prob(j) = \int_0^\infty PDF_j(t)dt. \tag{20}$$

[5] We follow the first adjustment made by Heathcote and Love [2012] to translate this for truncated normals.

[6] For full derivations of equations 17, 18 and 19, refer to appendix A of Brown and Heathcote (2008).

### 2.3. Methodological developments

In this section we detail a number of methodological improvements to both MLBA and DFT to make both models more applicable to non-laboratory based choice contexts. Our key aim here is to make both models suitable for capturing the influence of socio-demographics, the presence of underlying preferences for alternatives and to capture the relative importance of attributes, without having to know the directionality of the effect of the attribute on the choice probability of the alternative. We additionally detail considerations required for optimising the estimation of both DFT and MLBA.

#### *2.3.1. Scaling of DFT*

In a typical linear additive RUM or RRM model, changing the units of a single attribute only affects the parameter for that attribute. For example, changing the unit of travel time from minutes to hours results in the corresponding marginal utility component being multiplied by 60, with no impact on other parameters.

DFT on the other hand is scale-variant [Busemeyer and Diederich, 2002, Trueblood et al., 2013a], which means that a change in the scale for one attribute will have an impact on the relative importance weights for all attributes rather than just its own relative importance weight. Indeed, the attention weight parameters for the different attributes are linked and adjusting one will also shift the weights for the other attributes, given the requirements of summation to one ($\sum_k w_k = 1$). An illustration of this is given in Table 1. If we originally have weights of 0.6 and 0.4 for cost and time respectively, then the new importance for time with be multiplied by a value of 60 if we change from minutes to hours. These then need to be rescaled to ensure that summation to 1, leading to new weights of 0.024 for cost and 0.976 for time. Whilst this adjustment can be easily made for only two attributes, estimation is simpler if this can be avoided.

**TABLE 1** : Impact of changing the unit of time on attribute importance estimates

| | | Cost | Time |
|---|---|---|---|
| DFT coefficients | original weight | 0.600 | 0.400 |
| | new importance | 0.600 | 24.000 |
| | new weight | 0.024 | 0.976 |

We therefore define a new scaling method which attempts to translate attribute values into subjective values. This is achieved by multiplying the values by a set of attribute-specific scaling coefficients[7], $\beta_{DFT}$. As the estimation of these parameters would be confounded with estimates for the attention weights (for choice-only datasets where no additional information about the choice process is known), we instead set the weights equal to $1/K$ where $K$ is the number of attributes. This results in a different function for the random valence vector at time t, $V_t$, which can now be

[7]We use the term $\beta_{DFT}$ here as these values correspond to the marginal utility components, $\beta$, of RUM, but they cannot be used equivalently in, for example, value of travel time calculations.

calculated as:

$$V_t = C \cdot M^* \cdot W_t + \varepsilon_t \tag{21}$$

where $M^*$ is the original attribute matrix $M$ but with each attribute multiplied by its corresponding element from the vector $\beta_{DFT}$, which has different (estimated) scaling values for each attribute $x$. With this specification, a decision-maker still attends to a given attribute at random in a given evaluation - we simply no longer estimate separate weights $w_j$, as they are all set to be equal. In $W_t$, as before, one element is equal to 1 with all others 0, but the probabilities of this are now constant across attributes.

This change to DFT results in a number of important benefits. Firstly, the revised version of DFT is no longer scale-variant. Changing the unit of travel time from minutes to hours will now impact the estimate for the travel time scaling coefficient only. This means that for each marginal utility coefficient in a RUM model (or a marginal regret coefficient in a RRM model), there is a corresponding attribute scaling coefficient in the DFT model. This allows us to make comparisons across the different models in terms of relative importance of different attributes. Additionally, the attributes are now adjusted accordingly for their relative importance before they enter the feedback matrix, meaning that we can now calculate an appropriate psychological distance by simply taking Euclidean distances in the calculation of the feedback matrix. Consequently, we do not need a separate parameter $w$, as defined by Berkowitsch et al. [2015] to take the relative importance of the attributes into account.

An even more important benefit of the proposed scaling approach relates to the possibility of attributes having opposite impacts on probabilities, i.e. some attributes being desirable and others being undesirable. In the traditional DFT model, an analyst needs to make a priori assumptions about this directionality, and failing to correct for the *sign* of attributes can have undesired consequences, as illustrated in Table 3 of Hancock et al. [2018]. With our new approach, we no longer require a priori knowledge or assumptions on whether an attribute has a positive or negative impact on the likelihood of an alternative being chosen, as the attribute scaling parameters can be estimated to be either positive or negative. This not only results in it being possible to take all attributes into account without any initial adjustments, but would also, in a random coefficients DFT model, allow for the possibility of different signs for a given parameter across different individuals.

Finally, this new scaling method allows for the impact of attributes to be adjusted by incorporating socio-demographic interactions such as income effects parameter or alternative specific coefficients for given attributes (as demonstrated in our empirical applications section).

It is worth noting here that the new scaling method results in an additional parameter, which means that there is an overspecification until at least one parameter is fixed, a point we return to in Section 2.3.4.

#### *2.3.2. Incorporating baseline preferences in MLBA*

A key feature of discrete choice models belonging to the RUM family is the concept of alternative specific constants that capture baseline preferences for specific alternatives. Hancock et al. [2018] discusses in detail how this can be implemented in a DFT model. Here, we extend this to a MLBA model too.

In particular, we rewrite Equation 12 as

$$d_j = \delta_j + v_j + I_0, \tag{22}$$

where $\delta_j$ is an additional alternative-specific estimated constant capturing a baseline preference for alternative $j$. Unlike in RUM models, where the same differences between constants $\delta_j$ results in the same probabilities, each mean drift rate can have a separately identified constant, as the greater the rates, the less deterministic the choice is. However, if we additionally estimate $I_0$, then one of the constants $d_j$ must be fixed to ensure identification.

### *2.3.3. Incorporating attribute specific weights in MLBA*

An additional limitation of the current implementation of MLBA is in the treatment of the different attributes. Firstly, this applies in terms of directionality, noting that $\lambda_1$ is used for a positive difference between $x_{k,i}$ and $x_{k,j}$ independently of whether attribute $x_k$ is a desirable attribute or not. This limitation is analogous to the issue with using weight parameters in DFT and would require an analyst to a priori change the sign on undesirable attributes. Secondly, the actual impact of differences between alternatives in a given attribute $x_k$ is constant across attributes. Whilst one possibility is to use different valuation and weighting functions [Cohen et al., 2017], Trueblood et al. [2014] suggest that attribute biases can be dealt with by including attribute-specific 'bias parameters', $\beta_k$ (an approach analogous to the attribute-specific scaling coefficients that we defined for DFT) in Equation 14, which becomes:

$$w_{x_{k,i,j}} = exp(-\lambda \cdot \beta_k \cdot |x_{k,i} - x_{k,j}|) \tag{23}$$

However, we can relax the limitations of attribute bias and directionality simultaneously by also making an adjustment to the value function (Equation 13), which now takes the same form as that of the original specification with the exception that we add in attribute-specific scaling coefficients, $\beta_k$. This results in the value function from Equation 13 being redefined as:

$$v_j = \sum_{j \neq i} \sum_{k=1}^{K} (w_{x_{k,i,j}} \cdot \beta_k \cdot (x_{k,j} - x_{k,i})). \tag{24}$$

As with the scaling applied to DFT, this change allows us to also make inferences about the relative importance of different attributes in MLBA, as well as incorporate interactions with socio-demographics at the level of individual attributes.

### *2.3.4. Improving the estimation of DFT and MLBA*

Under a DFT model, at the conclusion of the deliberation process, the alternative that is chosen is the one with the greatest preference value, regardless of whether the individual stopped deliberating due to a time threshold or due to the preference value for one of the alternatives reaching a preference threshold. Given that most choices do not have a strict time threshold, some applications of DFT calculate the probability for each alternative's preference value reaching a particular threshold first (see examples in Hotaling et al. [2010] and Turner et al. 2017). However, as this probability has no closed-form solution for more than two alternatives, we rely on Roe et al. [2001]'s method to calculate the probability for each alternative after a particular number of deliberation timesteps. Whilst the choices that we investigate also do not have a strictly imposed time threshold, the number of deliberation timesteps is an estimated parameter, meaning that we do not impose a strict time threshold. From the previous sections, we can see that in order to estimate the probability of alternatives under a DFT model, we require estimates for $K$ attribute scaling parameters (where K is the number of attributes) and estimates for four 'process parameters'[8], which are exclusive to DFT and inform the process by which alternatives accumulate preference ($\phi_1$ and $\phi_2$, the sensitivity and memory parameters respectively, the number of timesteps, $t$, and the variance of the error term, $\sigma_\varepsilon^2$). Correspondingly, for the probability of alternatives in a MLBA model, we require estimates for $n$ scaling parameters and estimates for six process parameters ($A$ and $\chi$, the start and threshold parameters respectively, a drift rate constant, $I_0$, a parameter for the standard deviation of the drift rates, $s$ and similarity parameters $\lambda_1$ and $\lambda_2$).

The process parameters in DFT and MLBA have important behavioural roles. However, both models are routinely estimated on data where the only observed outcome is the choice itself, with little information about the process by which that choice was reached. If such process information was available, analysts could use it as additional indicators (i.e. additional dependent variables) in a joint estimation of process and outcome, and this would help inform the values of these parameters. In the absence of such data however, some of the parameters may become partially confounded. This results in restrictions that need to be considered to improve the stability of DFT and MLBA.

The various restrictions are detailed in Table 2, and are also used in all of our empirical applications. For DFT, the noise that is added on at each timestep to the valence (see Equation 3) is drawn from a normal distribution with mean 0 and variance $\sigma_\varepsilon^2$. Consequently, as $\sigma_\varepsilon^2 \geq 0$, we instead estimate the standard deviation, $\sigma_\varepsilon$. Additionally, the number of timesteps must exceed a value of one. Furthermore, the sensitivity parameter should be positive, as this ensures that alternatives that are more similar to each other compete more than alternatives that are less similar. Finally, the use of the new scaling method (detailed in Section 2.3.1) results in an overspecification. This is a result of the same probabilities being generated if all attribute scaling factors and the standard deviation of the error are multiplied by some factor $f$ and the sensitivity parameter, $\phi_1$ is divided by $f^2$. Consequently, we always fix one of the attribute scaling coefficients in our empirical applications.

For MLBA, the similarity parameters are also positive so as to ensure that similarity is a function of distance with more similar alternatives competing more, relative to less similar alternatives. It is these features which allow for these two models to predict the similarity effect. The drift rate constant in MLBA must also be positive, as should the start parameter $A$, from which the initial preference for each alternative is determined (from a uniform distribution U[0,A]). Addi-

[8]Henceforth, if we refer to 'process parameters' of either DFT or MLBA, we mean parameters which have no equivalent measure in a traditional model such as a RUM or RRM model.

**TABLE 2** : Restrictions on the parameters within DFT and MLBA

| Model | Parameter | Description | Restrictions | Estimated parameter | Relation |
|---|---|---|---|---|---|
| DFT | $t$ | timesteps | $> 1$ | $t^*$ | $t = 1 + exp(t^*)$ |
| DFT | $\phi_1$ | sensitivity | $> 0$ | $\phi_1^*$ | $\phi_1 = exp(\phi_1^*)$ |
| MLBA | $\lambda_1$ | similarity | $\geq 0$ | $\lambda_1^*$ | $\lambda_1 = exp(\lambda_1^*)$ |
| MLBA | $\lambda_2$ | similarity | $\geq 0$ | $\lambda_2^*$ | $\lambda_2 = exp(\lambda_2^*)$ |
| MLBA | $I_0$ | drift rate constant | $\geq 0$ | $I_0^*$ | $I_0 = exp(I_0^*)$ |
| MLBA | A | start | fixed | not estimated | n/a |
| MLBA | $\chi$ | threshold | $\geq$ A | $\chi^*$ | $\chi = A * (1 + exp(\chi^*))$ |
| MLBA | s | drift rate standard deviation | fixed | not estimated | n/a |

tionally, for choice-only data (i.e. where no additional process information is available), the start and threshold parameters $A$ and $\chi$ are perfectly confounded. Indeed, multiplying $A$ and $\chi$ by some factor $f$ results in no change in the probabilities with which each alternative is chosen, but it simply changes the time that alternative $j$ finishes in from $t_j$ to $f \cdot t_j$. As all alternatives are impacted in the same way, we consequently do not need to estimate both $A$ and $\chi$. The threshold $\chi$ also must be specified such that it is at least the same value as the start parameter (to avoid the possibility that more than one alternative reaches the threshold before any deliberation has taken place). We therefore fix $A$ and estimate $\chi = (1 + exp(\chi^*) \cdot A)$. Furthermore, we fix the variance of the drift rate, $s$, to avoid the possibility that the probabilities with which the alternatives are chosen remain exactly the same if all mean drift rates $d_j$ and the standard deviation $s$ are doubled[9]. Finally, an MLBA model which does not find a similarity effect will result in $\lambda$ parameters approaching zero. In this case, these parameters are fixed to zero to avoid overspecification, as small changes in these parameters when they are arbitrarily small will result in no change in the probabilities of each alternative being chosen. This issue resulted in previous applications of MLBA resorting to different valuation and weighting functions [Cohen et al., 2017].

Some of the above constraints are necessary to avoid identification issues, while others simply avoid sign issues. For the latter, free estimation may in theory be possible, but we have found the constraints to be helpful in our work.

The estimation of DFT and MLBA remains a non-trivial computational task even with the above constraints, and efficient implementation as well as good starting values are essential. In our work, we use the R packages maxLik [Henningsen and Toomet, 2011] and *Apollo* [Hess and Palma, 2019] for estimation of the likelihood function and the RCPP package together with the Armadillo C++ linear algebra library for fast calculation of the matrices required for finding the probability under which each alternative is chosen under a DFT model [Eddelbuettel et al., 2011, Sanderson and Curtin, 2016]. Additionally, we use an initial parameter search algorithm based on the heuristic for non-linear global optimisation developed by Bierlaire et al. [2010] in an attempt to reduce the risk of convergence to poor local optima as well as an excessively long estimation process.

## 3. EMPIRICAL APPLICATIONS ON REVEALED AND STATED CHOICE DATA

In this section, we present empirical results using DFT and MLBA on three different datasets, two from stated choice (SC) surveys and one from a revealed preference (RP) survey, where the latter is the first DFT/MLBA application to RP data. We provide a detailed investigation as to empirical identification of DFT and MLBA. This is crucial, as in the context of choice-only data, some of these parameters will be confounded, and it has not yet been established what normalisation should be applied. We also compare the estimation results to typical MNL and RRM models. We finally present an empirical comparison between the different existing specifications of DFT and our proposed new scaling approach.

[9] This is possible, for example, if the value for each $\beta_x$ is doubled, the positive constant $I_0$ is doubled, and the value of each similarity parameter $\lambda_1$ and $\lambda_2$ is halved.

### 3.1. First stated choice survey

Our first dataset is a subset from the Danish value of time dataset [Fosgerau, 2006]. This dataset comes from a typical stated choice survey, where 545 participants faced a total of 4,214 choices between them. The choices were for car drivers and specifically the choice between two different routes, described only by travel cost and travel time, where one route would be cheaper, but the other would be faster. The aim of such a setup is to understand trade-offs between time and money, leading to estimates of the value of travel time (VTT). While very simplistic in nature, this type of datasets is a useful first step in moving from the abstract settings in mathematical psychology towards more complex choices in a transport setting. In all models, we only focussed on the time and cost attributes after earlier results confirmed there was no left-right bias that would require the inclusion of alternative specific constants.

Table 3 shows the results for the first SC dataset. Where appropriate, we used the constraints from Table 2 but then report the actual transformed estimates in Table 3, along with the transformed standard errors, obtained using the Delta method [cf. Daly et al., 2012].

We first have two MNL models, one using a purely linear specification while the second additionally estimates parameters for the logarithm of time and cost. This latter model offers a significant improvement in fit over the first model, and all four coefficients remain negative, where the significant estimates for the log-time ($\beta_{LTT}$) and log-fare ($\beta_{LF}$) parameters indicate non-linear sensitivities.

Whilst a number of different parameters within DFT could be fixed to solve the overspecification issue identified in Section 2.3.4, we choose to fix the first attribute scaling coefficient, thus focussing on relative sensitivities, where we use the value from the MNL model to aid comparison. We then trial two different DFT models to test the impact of removing the effect of the feedback matrix (model 2 compared to model 1).

As there are only two alternatives in the Danish dataset, $\phi_1$ is of little meaning, given that it is a parameter for the level of competition between alternatives dependent on the distance between the alternatives, and its estimate tends to zero in DFT model 1. Additionally, $\phi_2$, the memory parameter, has little meaning when the sequence of attribute attendance is not known. It however also contributes to the level of competition between alternatives as a value of $\phi_2 = 0$ results in the value of $\phi_1$ having no impact (see Equation 2). As $\phi_2 = 0$ in model 1, we can thus also remove $\phi_1$ also. For this Danish dataset, we can thus use an identity matrix for the feedback matrix, leading to no loss in fit for model 2 compared to model 1, showing that with binary data, the estimation of the feedback matrix does not seem to apply.

Similar to decision field theory, the multi-attribute linear ballistic accumulator has many parameters that have little interpretable output if an analyst only has access to the choice data and no additional psychometric or process data. For example, a decision-maker could make a choice quickly because there is a small difference between the start and threshold parameters or because they have a higher deviation in the drift rates. Consequently, if we only have choice data and no information about the process in which the choice was made, then some of the MLBA parameters may become confounded. As with DFT, we initially test MLBA using a full specification, which

**TABLE 3** : Estimation results and identifications tests on the first SC dataset

| Model | | MNL | | DFT | | MLBA | | |
|---|---|---|---|---|---|---|---|---|
| Version | | 1 | 2 | 1 | **2** | 1 | 2 | 3 |
| Free Pars. | | 2 | 4 | 5 | **3** | 6 | 5 | 5 |
| Log-likelihood | | -2,301.53 | -2,212.10 | -2,018.73 | **-2,018.73** | -2,007.88 | -2,037.24 | -2,008.55 |
| BIC | | 4,619.75 | 4,457.58 | 4,079.19 | **4,062.50** | 4,065.83 | 4,116.21 | 4,058.83 |
| $\beta_{TT}$ | est. | -0.1939 | -0.1590 | -0.1939 | -0.1939 | -2.8676 | -3.0671 | -3.2293 |
| | r. t-rat. | -13.54 | -8.86 | **fixed** | **fixed** | -63.67 | -10.77 | -7.34 |
| $\beta_F$ | est. | -2.4087 | -1.7643 | -3.0988 | -3.0987 | -45.3590 | -45.6207 | -50.9441 |
| | r. t-rat. | -13.52 | -9.21 | -23.36 | -23.35 | -34.66 | -9.91 | -6.34 |
| $\beta_{LTT}$ | est. | | -1.0355 | | | | | |
| | r. t-rat. | | -2.71 | | | | | |
| $\beta_{LF}$ | est. | | -1.9005 | | | | | |
| | r. t-rat. | | -5.38 | | | | | |
| $\phi_1$ | est. | | | 0.0688 | 0.0000 | | | |
| | r. t-rat. | | | 0.00 | **fixed** | | | |
| $\phi_2$ | est. | | | 0.0000 | 0.0000 | | | |
| | r. t-rat. | | | 0.00 | **fixed** | | | |
| $\sigma_\varepsilon$ | est. | | | 0.1587 | 0.1587 | | | |
| | r. t-rat. | | | 0.88 | 0.88 | | | |
| $t$ | est. | | | 6.6097 | 6.6097 | | | |
| | r. t-rat. (*vs* 1) | | | 12.12 | 12.11 | | | |
| $\chi$ | est. | | | | | 1.1938 | 2.0000 | 1.1634 |
| | r. t-rat. (*vs* 1) | | | | | 55.87 | **fixed** | 6.43 |
| $I_0$ | est. | | | | | 2.1548 | 39.6940 | 1.8380 |
| | r. t-rat. | | | | | 104.17 | 30.96 | 2.66 |
| $\lambda_1$ | est. | | | | | 0.0008 | 0.0025 | 0.0000 |
| | r. t-rat. | | | | | 12.90 | 10.36 | **fixed** |
| $\lambda_2$ | est. | | | | | 0.2068 | 0.0369 | 0.1962 |
| | r. t-rat. | | | | | 16.47 | 11.88 | 6.54 |

implies only fixing the start parameter $A$ and the drift rate standard deviation $s$ to values of 1.

In model 2, we set fix the threshold parameter $\chi$, which is a common approach in mathematical psychology [Trueblood et al., 2014, Cohen et al., 2017, Cataldo and Cohen, 2018] (fixing it to a value of 2 as done in the original MLBA paper Trueblood et al. 2014), but find that this is not appropriate in this case, leading to a substantial loss of fit. On the other hand, our initial estimate for $\lambda_1$ is so close to zero that a constraint does not lead to any significant loss of fit. This is however an empirical issue rather than a theoretical identification requirement. A value of $\lambda_1 = 0$ results in weights, $w_{x_{k,i,j}} = 1$, thus resulting in a simplified calculation of the mean drift rates with linear contributions from positive differences of attributes $x_{k,j} - x_{k,i}$.

In terms of model performance, we see that DFT and MLBA both outperform MNL. The difference in fit between DFT and MLBA is much smaller than between these two models and MNL, with a slight advantage for MLBA.

### 3.2. Second stated choice survey

The second stated choice dataset we consider has a total of 368 participants, each completing 10 choice tasks resulting in 3,680 choices. The participants are all public transport commuters living in the UK. Each task involves the choice between an invariant reference trip and two hypothetical alternatives, where each of the three alternatives is described by travel time, cost, rate of crowded trips, rate of delays (both out of 10 trips), the average length of delays (entered into models both as the average extent of delays, RA, and as the expected delay, RB, by multiplying the length of delays by the rate of delays) and the provision of a delay information service (none used as the base, with parameters for a charged, ICH, and free, IFR, service). Following earlier results by Hess and Stathopoulos [2013], we applied a log-transform to the fare attribute (described as LF).

Table 4 shows the results for the second SC dataset. In the presence of three alternatives, we can now include a RRM model alongside MNL, where we see fairly similar performance for these two models, with a slight advantage for MNL. All parameters have the expected sign in these models, and we are also able to include two alternative specific constants (ASCs)[10].

For DFT, we follow the same specification tests as on our first SC dataset. However, this time we are now able to estimate a significant memory parameter $\phi_2$, suggesting that initial comparisons matter more than current or more recent ones. The estimate for the off-diagonal term $\phi_1$, which is more meaningful in data with three alternatives, remains only weakly significant. However, constraining the feedback matrix to be an identity matrix (as in model 2) now clearly leads to a significant drop in model fit.

For MLBA, we again show that constraining $\chi = 2$ is not appropriate, leading to a loss of fit for model 2. We are however able to constrain it to $\chi = 1$ and in addition can constrain $I_0 = 0$ (model 3) without any loss of fit, where this is potentially due to the fact that we are now able to estimate significant ASCs. Higher values are observed for $\lambda_2$ compared to $\lambda_1$, meaning that a greater importance weight is given to positive attribute differences $x_{k,j} \geq x_{k,i}$ compared to negative ones $x_{k,j} < x_{k,i}$ in the estimation of the mean drift rates.

In terms of model performance, we see that DFT and MLBA again both outperform MNL (and also RRM), where, with the present data, DFT offers better performance than MLBA, potentially as it is better able to deal with the differential competition between the three alternatives than MLBA (in contrast to the earlier binary dataset).

### 3.3. RP data

Whilst both DFT and MLBA have been used extensively on experimental data and have been demonstrated to accurately explain choices in stated preference surveys, as far as we are aware,

[10]Which results in improvements in log-likelihood of 46 and 26 units respectively for DFT and MLBA

**TABLE 4** : Estimation results and identifications tests on the second SC dataset

| Model | | MNL | RRM | DFT | | MLBA | | |
|---|---|---|---|---|---|---|---|---|
| Version | | 1 | 1 | **1** | 2 | 1 | 2 | **3** |
| Free Pars. | | 10 | 10 | **13** | 11 | 14 | 13 | **12** |
| Log-likelihood | | -3,360.43 | -3,363.91 | **-3,299.82** | -3,327.28 | -3,321.75 | -3,333.66 | **-3,322.36** |
| BIC | | 6,802.97 | 6,809.92 | **6,706.41** | 6,744.88 | 6,758.45 | 6,774.06 | **6,743.26** |
| $\beta_{TT}$ | est. | -0.0471 | -0.0320 | -0.0471 | -0.0471 | -0.0586 | -0.0230 | -0.0592 |
| | r. t-rat. | -9.50 | -9.58 | **fixed** | **fixed** | -1.53 | -5.71 | -5.07 |
| $\beta_{LF}$ | est. | -5.9990 | -4.1090 | -6.5220 | -6.2709 | -11.0745 | -4.8500 | -11.3073 |
| | r. t-rat. | -18.87 | -17.66 | -9.67 | -9.52 | -1.56 | -8.55 | -3.38 |
| $\beta_{CR}$ | est. | -0.2230 | -0.1212 | -0.2137 | -0.2360 | -0.2842 | -0.1174 | -0.2902 |
| | r. t-rat. | -8.58 | -5.82 | -7.10 | -6.95 | -1.02 | -5.69 | -4.09 |
| $\beta_{RA}$ | est. | -0.1870 | -0.0441 | -0.1523 | -0.1957 | -0.1499 | -0.0687 | -0.1519 |
| | r. t-rat. | -5.96 | -2.71 | -4.54 | -3.62 | -0.71 | -1.90 | -2.88 |
| $\beta_{RE}$ | est. | -0.0619 | -0.1457 | -0.0912 | -0.0624 | -0.1341 | -0.0412 | -0.1349 |
| | r. t-rat. | -2.64 | -8.59 | -2.58 | -1.16 | -0.69 | -1.43 | -5.37 |
| $\beta_{RB}$ | est. | -0.0293 | -0.0186 | -0.0138 | -0.0244 | -0.0205 | -0.0100 | -0.0219 |
| | r. t-rat. | -3.25 | -3.06 | -1.85 | -1.86 | -0.17 | -3.84 | -1.31 |
| $\beta_{ICH}$ | est. | -0.0910 | -0.0510 | -0.0013 | -0.0651 | -0.0424 | -0.0194 | -0.0464 |
| | r. t-rat. | -1.13 | -0.95 | -0.03 | -1.00 | -0.20 | -1.28 | -0.47 |
| $\beta_{IFR}$ | est. | 0.3305 | 0.2179 | 0.2270 | 0.2772 | 0.3224 | 0.1413 | 0.3220 |
| | r. t-rat. | 4.95 | 4.85 | 4.28 | 4.33 | 0.35 | 7.52 | 3.99 |
| $asc_1$ | est. | 0.3902 | -0.2730 | 0.2841 | 0.7258 | 1.0941 | 0.4588 | 1.1213 |
| | r. t-rat. | 5.85 | -4.17 | 4.30 | 6.20 | 0.40 | 7.81 | 3.72 |
| $asc_2$ | est. | 0.1633 | -0.1656 | 0.1354 | 0.2178 | 0.3356 | 0.1714 | 0.3560 |
| | r. t-rat. | 3.30 | -3.38 | 3.24 | 2.83 | 0.20 | 6.09 | 2.45 |
| $\phi_1$ | est. | | | 0.0205 | 0.0000 | | | |
| | r. t-rat. | | | 1.17 | **fixed** | | | |
| $\phi_2$ | est. | | | -0.5586 | 0.0000 | | | |
| | r. t-rat. | | | -4.90 | **fixed** | | | |
| $\sigma_\varepsilon$ | est. | | | 0.1276 | 0.3566 | | | |
| | r. t-rat. | | | 3.46 | 5.20 | | | |
| $t$ | est. | | | 5.2245 | 7.5332 | | | |
| | r. t-rat. (*vs* 1) | | | 8.34 | 6.37 | | | |
| $\chi$ | est. | | | | | 1.0004 | 2.0000 | 1.0000 |
| | r. t-rat. (*vs* 1) | | | | | 1.83 | **fixed** | **fixed** |
| $I_0$ | est. | | | | | 0.0167 | 1.6128 | 0.0000 |
| | r. t-rat. | | | | | 0.39 | 7.53 | **fixed** |
| $\lambda_1$ | est. | | | | | 0.1188 | 0.3490 | 0.1191 |
| | r. t-rat. | | | | | 0.33 | 7.12 | 2.03 |
| $\lambda_2$ | est. | | | | | 1.0805 | 2.0307 | 1.0556 |
| | r. t-rat. | | | | | 0.39 | 16.42 | 4.33 |

neither model has been fitted to revealed preference (RP) data. In this section, we first fit MNL, RRM, DFT and MLBA models to our full RP dataset. We then provide elasticities as well as additionally testing out-of-sample prediction for all four models.

Our RP data comes from the national UK value of travel time study [Arup, ITS Leeds and Accent, 2015]. Questionnaires were completed by 2,646 individuals travelling by train from Birmingham, Stoke or Peterborough to London. After extensive data cleaning (see page 164 of Arup, ITS Leeds and Accent 2015), 725 observations were left, with either one or two observations for each of the 578 individuals[11]. For every decision recorded, the available alternatives are one or two of Chiltern railways, Northern rail and Midlands railways as well as one of Virgin Trains and East Coast. Travel time, travel cost and headway were used to describe the alternatives.

We run a basic MNL model with a specification based on the model used by Arup, ITS Leeds and Accent [2015], with four different travel time coefficients for different groups. Individuals are first segmented by travel purpose (employees' business, commute ($TT_C$ in Table 5) or 'other non-work' ($TT_O$)). Individuals on employees' business where further segmented into those who were very sure ($TT_{EB1}$) and those who were quite sure ($TT_{EB2}$) about the attributes of the unchosen alternatives. Two further attribute parameters are estimated (travel cost, $TC$, and headway, $HW$). For all three attributes, log values are used [Arup, ITS Leeds and Accent, 2015]. Additionally, Arup, ITS Leeds and Accent [2015] use three alternative specific constants for train services run by Chiltern railways ($ASC_C$), Midlands railways ($ASC_M$) and Northern rail ($ASC_N$). Finally, two parameters are incorporated to capture income effects. Value of travel time coefficients ($\beta_{TT_n}$) are calculated for each individual $n$:

$$\beta_{TT_n} = \beta_{TT_{i,n}} \cdot \left( {RI_n}^{\lambda_{inc}} \cdot (1 - z_{miss,n}) + \lambda_{miss} \cdot z_{miss,n} \right) \tag{25}$$

where $\beta_{TT_{i,n}}$ is a travel time coefficient depending on the individual's trip purpose, $RI_n$ is the relative income of the individual, $\lambda_{inc}$ is an income elasticity on the time sensitivity and $\lambda_{miss}$ is a multiplier on the time sensitivity used only if the individual did not provide their income in the questionnaire (in which case the dummy variable $z_{miss,n} = 1$). Table 5 provides model estimates for these parameters under MNL, RRM, DFT and MLBA.

For DFT, we again use the MNL value for the cost coefficient (model 1). With 118 out of 725 observations having three alternatives available and the rest having only two alternatives available, it is unsurprising that, in line with the results from the 1st SC dataset, the effect of the feedback parameters being removed (DFT model 2 relative to DFT model 1) has little impact on the log-likelihood.

For MLBA, we see that fixing $\lambda_2$ to infinity (which results in the corresponding weight, $w_{x_{k,i,j}} = 0$, when $x_{k,j} < x_{k,i}$) has no impact on model fit (model 2 compared to model 1). Additionally fixing both $I_0$ and $\chi$ results in an insignificant impact on model fit with a lower BIC value obtained for model 3 compared to model 2.

[11] Note that this means we have data that is panel data for some individuals, but not others. This does not impact the results of the models as the only impact this has is in the calculation of standard errors.

**TABLE 5** : Results, estimates and robust t-ratios from MNL, RRM, DFT and MLBA models on the RP dataset

| Model | | MNL | RRM | DFT | | MLBA | | |
|---|---|---|---|---|---|---|---|---|
| Version | | 1 | 1 | 1 | **2** | 1 | 2 | **3** |
| Free Pars. | | 11 | 11 | 14 | **12** | 15 | 14 | **12** |
| Log-likelihood | | -370.05 | -373.31 | -362.53 | **-363.31** | -351.97 | -351.97 | **-352.07** |
| BIC | | 812.54 | 819.07 | -817.26 | **805.66** | 802.73 | 796.14 | **783.17** |
| $TT_C$ | est. | -4.4541 | -1.1490 | -6.2586 | -6.2414 | -25.1388 | -25.1615 | -25.6064 |
| | rob. t-rat. | -3.88 | -3.70 | -2.35 | -3.03 | -2.24 | -2.20 | -2.24 |
| $TT_O$ | est. | -2.0021 | -0.5272 | -2.7685 | -2.8922 | -4.0627 | -4.0592 | -4.0716 |
| | rob. t-rat. | -2.46 | -2.47 | -2.42 | -3.36 | -2.76 | -2.73 | -2.87 |
| $TT_{EB1}$ | est. | -3.7769 | -0.9341 | -4.2679 | -4.4185 | -7.4737 | -7.4681 | -7.4193 |
| | rob. t-rat. | -4.63 | -4.66 | -3.40 | -5.03 | -4.67 | -4.69 | -5.01 |
| $TT_{EB2}$ | est. | -5.7016 | -1.4342 | -7.2535 | -7.3995 | -10.7129 | -10.7074 | -10.7462 |
| | rob. t-rat. | -7.10 | -6.85 | -3.74 | -6.09 | -6.92 | -6.74 | -7.01 |
| $TC$ | est. | -2.2127 | -0.6152 | -2.2127 | -2.2127 | -4.2257 | -4.2246 | -4.3372 |
| | rob. t-rat. | -8.52 | -9.27 | **fixed** | **fixed** | -48.19 | -5.35 | -4.73 |
| $HW$ | est. | -0.1267 | -0.0532 | -0.1255 | -0.1240 | -0.1189 | -0.1191 | -0.1301 |
| | rob. t-rat. | -0.64 | -0.75 | -0.75 | -0.73 | -0.62 | -0.62 | -0.40 |
| $ASC_C$ | est. | 0.7549 | 0.6813 | 2.9787 | 3.9385 | 1.5041 | 1.5036 | 1.5100 |
| | rob. t-rat. | 2.75 | 2.52 | 1.85 | 2.49 | 4.36 | 4.28 | 4.30 |
| $ASC_M$ | est. | -0.4882 | -0.5251 | -0.6376 | -0.5124 | -0.4468 | -0.4472 | -0.4929 |
| | rob. t-rat. | -1.86 | -1.96 | -0.41 | -0.39 | -1.75 | -1.53 | -1.64 |
| $ASC_N$ | est. | -0.4879 | -0.5164 | -0.4111 | -0.4429 | -0.4304 | -0.4308 | -0.4606 |
| | rob. t-rat. | -1.65 | -1.68 | -0.28 | -0.30 | -1.30 | -1.31 | -1.29 |
| $\lambda_{inc}$ | est. | 0.4563 | 0.4690 | 0.5596 | 0.5411 | 0.5857 | 0.5857 | 0.5953 |
| | rob. t-rat. | 4.43 | 4.46 | 4.48 | 4.48 | 4.63 | 4.80 | 6.76 |
| $\lambda_{miss}$ | est. | 0.4844 | 0.4939 | 0.8020 | 0.7951 | 0.5813 | 0.5798 | 0.5685 |
| | rob. t-rat. | 1.13 | 1.18 | 1.04 | 1.09 | 0.70 | 0.71 | 0.78 |
| $\phi_1$ | est. | | | 1.1530 | 0.0000 | | | |
| | rob. t-rat. | | | 0.64 | **fixed** | | | |
| $\phi_2$ | est. | | | -0.0865 | 0.0000 | | | |
| | rob. t-rat. | | | -1.71 | **fixed** | | | |
| $\sigma_\varepsilon$ | est. | | | 1.1588 | 1.1959 | | | |
| | rob. t-rat. | | | 2.53 | 2.99 | | | |
| $t$ | est. | | | 8.2073 | 8.1599 | | | |
| | rob. t-rat. (*vs* 1) | | | 2.54 | 3.03 | | | |
| $\chi$ | est. | | | | | 2.3344 | 2.3319 | 2.0000 |
| | rob. t-rat. (*vs* 1) | | | | | 11.88 | 1.36 | **fixed** |
| $I_0$ | est. | | | | | 0.0000 | 0.0009 | 0.0000 |
| | rob. t-rat. | | | | | 0.84 | 1.26 | **fixed** |
| $\lambda_1$ | est. | | | | | 0.1107 | 0.1107 | 0.1073 |
| | rob. t-rat. | | | | | 11.20 | 7.96 | 12.12 |
| $\lambda_2$ | est. | | | | | 4018.6448 | Inf | Inf |
| | rob. t-rat. | | | | | 0.76 | **fixed** | **fixed** |

With this data, we again see that MLBA obtains a lower BIC value than DFT, with both DFT and MLBA outperforming MNL and RRM, thus demonstrating that they work well for RP data as well as SC data.

With a view to not just focussing on model fit, Table 6 contrasts the cost and time elasticities on the RP data for the four models. We see that the elasticities for MNL and RRM are quite similar to each other. DFT obtains visibly higher time and cost elasticities than MNL and RRM. For MLBA, the cost elasticity is in between MNL/RRM and DFT, while the time elasticity is the lowest across all models. These results again show that DFT and MLBA offer more significant departures from standard models than for example RRM.

**TABLE 6** : Cost and time elasticities on RP data

| | elasticities | |
|---|---|---|
| | cost | time |
| MNL | -0.537 | -0.933 |
| RRM | -0.530 | -0.901 |
| DFT | -0.604 | -1.017 |
| MLBA | -0.584 | -0.881 |

We finally test all four models for their ability to make out-of-sample predictions. For each of the five data subsets, we take choices corresponding to a random 80% of the individuals in the data to be used for estimation, with the remaining 20% used for validation. We fit each model to each estimation subset and then calculate log-likelihoods for the remaining 20% of the data using the parameter estimates obtained for the first 80%. Table 7 gives the log-likelihoods of the estimation and validation subsets of the data under each model. Additionally, Figure 2 gives the average probability that the models assign to the chosen alternatives in the out-of-sample observations.

We see that DFT and MLBA outperform MNL and RRM across all five subsamples in both estimation and performance on the holdout sample except for DFT in holdout sample 4. MNL outperforms RRM in estimation and holdout across all samples, while MLBA always outperforms DFT. Overall, these findings confirm the results on the full sample.

### 3.4. Comparison of results

To summarise the results, Table 8 shows the BIC for the final recommended specification for each model type on each dataset. We see that DFT and MLBA consistently offer better performance than MNL and RRM. While MLBA marginally outperforms DFT on the Danish SC data, the differences are more substantial on the remaining two datasets, with DFT performing best on the UK SC data and MLBA best on the RP data.

An additional benefit of the new scaling method we use for DFT is that it allows us to more directly compare parameter estimates across different models, notwithstanding the different meaning of the parameters. This is possible as a result of the new specifications of both MLBA and DFT having attribute-specific scaling coefficients, which have a role analogous to marginal utility

**TABLE 7** : Out-of-sample estimation and holdout log-likelihoods for the RP data

| | *MNL* (11 pars) | | | *RRM* (11 pars) | | |
|---|---|---|---|---|---|---|
| | estimated | holdout | sum | estimated | holdout | sum |
| Full Data | | | -370.05 | | | -373.31 |
| Dataset 1 | -302.88 | -68.92 | -371.81 | -306.05 | -69.27 | -375.31 |
| Dataset 2 | -298.59 | -72.76 | -371.35 | -301.04 | -73.59 | -374.62 |
| Dataset 3 | -296.70 | -75.08 | -371.78 | -299.31 | -75.76 | -375.07 |
| Dataset 4 | -302.29 | -68.18 | -370.47 | -304.81 | -68.84 | -373.65 |
| Dataset 5 | -296.64 | -75.74 | -372.39 | -299.41 | -76.28 | -375.69 |
| | *DFT* (12 pars) | | | *MLBA* (12 pars) | | |
| | estimated | holdout | sum | estimated | holdout | sum |
| Full Data | | | -363.31 | | | -352.07 |
| Dataset 1 | -296.90 | -67.80 | -364.70 | -286.58 | -67.26 | -353.84 |
| Dataset 2 | -293.80 | -70.64 | -364.43 | -283.05 | -70.33 | -353.38 |
| Dataset 3 | -293.12 | -71.75 | -364.87 | -282.30 | -71.35 | -353.65 |
| Dataset 4 | -295.41 | -68.29 | -363.70 | -285.90 | -65.73 | -351.63 |
| Dataset 5 | -293.23 | -72.90 | -366.12 | -282.78 | -71.67 | -354.45 |

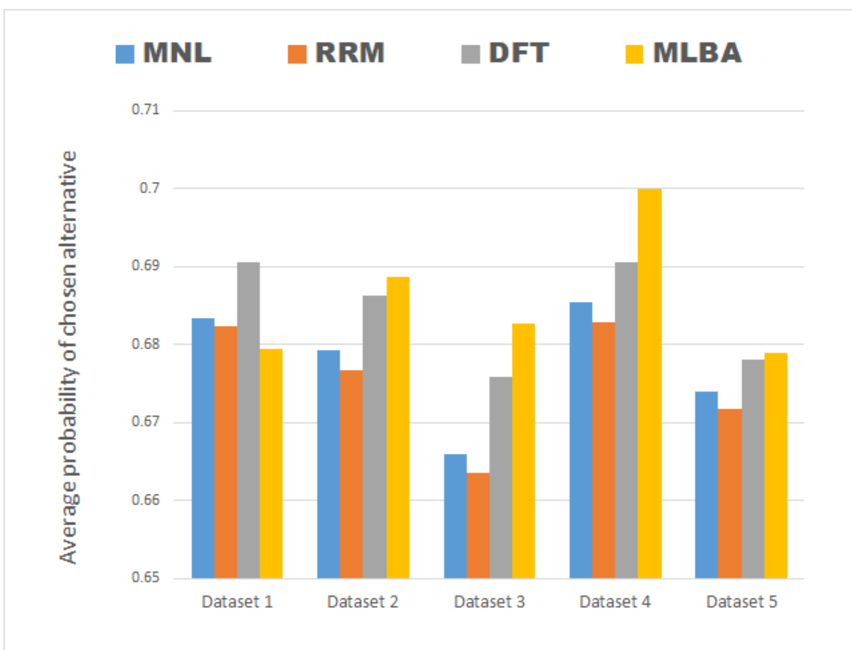



**FIGURE 2** : Average probabilities of the chosen alternatives for each holdout subset of the RP data.

coefficients in RUM models. Although these scaling coefficients cannot be directly translated into measures such as the value of travel time, we can calculate 'relative importance of travel time with respect to travel cost'. In Table 9, we set the calculated MNL values to a base rate of 1 (with the rates being based on the MNL value for commuters in the RP dataset). Consequently we can

**TABLE 8** : Model fit (BIC) comparison across models and datasets

| | MNL | RRM | DFT | MLBA |
|---|---|---|---|---|
| Danish SC | 4,457.58 | - | 4,062.50 | 4,058.83 |
| UK SC | 6,802.97 | 6,809.92 | 6,706.41 | 6,743.26 |
| RP | 812.54 | 819.07 | 805.66 | 783.17 |

compare whether DFT and MLBA assign more or less importance to travel time with respect to travel cost.

**TABLE 9** : The relative importance of travel time compared to cost across different models in comparison to MNL

| | | MNL | RRM | DFT | MLBA |
|---|---|---|---|---|---|
| SP | Danish | 1.000 | 1.000 | 0.777 | 0.785 |
| | UK | 1.000 | 0.992 | 0.920 | 0.667 |
| RP | Commuters | 1.000 | 0.928 | 1.401 | 2.931 |
| | Other Non-Work | 0.449 | 0.426 | 0.649 | 0.466 |
| | Employees' Business 1 | 0.848 | 0.754 | 0.992 | 0.849 |
| | Employees' Business 2 | 1.280 | 1.158 | 1.661 | 1.230 |

Across the SP datasets, it appears that MNL tends to assign higher importance to travel time with respect to travel cost relative to DFT and MLBA. The opposite is the case for the RP datasets. RRM always estimates lower ratios than MNL, while DFT has fairly similar values to MLBA, with an exception being the UK data and commuters in the RP data, for which DFT is more similar to MNL.

## 3.5. Scaling of attributes

In this section, we compare our new method (see Section 2.3.1) to scaling methods that have been used in previous DFT applications. The different scaling methods are:

1. Unity-based normalisation, as used by Berkowitsch et al. [2014], where we rescale the attributes values to a range between 0 and 1.

2. No scaling method other than taking the negative value for all 'negative' attributes (as DFT can only capture 'positive' effects of attributes as the relative importance weights must be positive[12] - see Section 4.3.1 of Hancock et al. [2018] for an illustration of the results of failing to do this for DFT models)

[12] Note that here we adjust the attributes accordingly so that our new scaling method does not have an unfair advantage for attributes which have a positive sign.

3. Standard score normalisation, as previously found to be effective for DFT (see results in Hancock et al. [2018]).

4. Minimum rescaling (dividing each attribute by the smallest value for that attribute across the choice set), as previously shown to be effective for a previous version of MLBA [Trueblood et al., 2013a]

5. Maximum rescaling (dividing each attribute by the largest value for that attribute across the choice set), as previously shown to be effective for a previous version of MLBA [Trueblood et al., 2013a]

6. Our new method detailed in Section 2.3.1, which removes the scale-variant nature of DFT.

For both datasets, it appears that scale 6 has the best model fit. This is regardless of whether we include DFT's feedback matrix. Crucially, scale 6 appears to better capture the impact of the feedback matrix for the UK data, resulting in an improvement by more than 40 log-likelihood units, where this improvement is much smaller with the other scalings. On the other hand, with the Danish data, the feedback matrix is needed for some of the other scalings to obtain model fits more in line with the new scaling. Additionally, the results here demonstrate that scales 4 and 5 offer relatively poor performance for DFT, which could in part explain why these scaling methods resulted in MLBA outperforming DFT previously [Trueblood et al., 2013a].

## 4. SIMULATED DATA EXPERIMENTS

The work in Section 3 has provided initial insights about the potential benefits of DFT and MLBA compared to more traditional structures. Of course, these results are dataset specific and the advantages might be a result of the true (and unobserved) data generation process. In this section, we provide some further evidence based on simulated data, where we have a number of aims. In particular, we test the impacts of considering choices generated by different models, compare the ability of the different accumulator models at capturing various complexities in the data, and finally consider parameter recoverability.

### 4.1. Generation of simulated data

We use an efficient design to generate $5,000$ mode choice observations where each choice task has four alternatives (car, air, rail and high-speed rail), each described by travel cost (TC) and travel time (TT). Additionally, all alternatives other than car have an access time (AT) attribute.

We then generate choices four times using a MNL model, a RRM model, a DFT model and an MLBA model. The aim of this exercise is to see how robust each of the models is to the case where the data stems from a different model.

For our MNL model, we define the utility a respondent $n$ obtains from alternative $j$ in choice task $t$ as:

$$U_{jnt} = ASC_j + ASC_{F_j} \cdot z_{F,n} + \beta_{TT} \cdot \alpha_{TT_j} \cdot TT_{jnt} + \beta_{TC} \cdot TC_{jnt} \cdot \alpha_{IE,n} + \beta_{AT} \cdot AT_{jnt} + \varepsilon_{jnt} \tag{26}$$

**TABLE 10** : The log-likelihood (LL) values obtained from models for the two stated choice datasets, with different types of scaling for DFT

| | Danish | | | | UK | | | |
|---|---|---|---|---|---|---|---|---|
| | with feedback | | without feedback | | with feedback | | without feedback | |
| | free pars. | LL | free pars. | LL | free pars. | LL | free pars. | LL |
| scale 1 | 5 | -2,020.24 | 3 | -2,020.24 | 11 | -3,404.09 | 9 | -3,405.60 |
| scale 2 | 5 | -2,034.24 | 3 | -2,040.94 | 11 | -3,395.50 | 9 | -3,400.88 |
| scale 3 | 5 | -2,021.86 | 3 | -2,021.86 | 11 | -3,390.31 | 9 | -3,400.41 |
| scale 4 | 5 | -2,112.62 | 3 | -2,112.62 | 11 | -3,419.43 | 9 | -3,420.48 |
| scale 5 | 5 | -2,139.50 | 3 | -2,146.38 | 11 | -3,442.14 | 9 | -3,443.22 |
| scale 6 | 5 | **-2,018.73** | 3 | **-2,018.73** | 11 | **-3,346.23** | 9 | **-3,387.38** |

where $ASC_j$ and $ASC_{F_j}$ are alternative specific constants, with the latter capturing the difference between male and female participants through the use of an appropriate dummy term which takes a value of 1 if individual $n$ is female. $TT_{jnt}$ is the travel time, $TC_{jnt}$ is the travel cost and $AT_{jnt}$ is the access time, all for alternative $j$ in choice situation $t$ for respondent $n$. There are coefficients for travel cost, access time and mode-specific coefficients for travel time, which are defined as $\beta_{TT} \cdot \alpha_{TT_j}$. A general value $\beta_{TT}$ is estimated, with appropriate adjustments applied by multiplying by $\alpha_{TT_j}$ for mode $j$ (for identification purposes we fix this coefficient for cars, $\alpha_{TT_{car}} = 1$). We additionally have an income effect, $\alpha_{IE,n}$, which is defined as $\alpha_{IE,n} = (\frac{income_n}{2500})^{\alpha_I}$, where $income_n$ is the income for individual $n$ and $\alpha_I$ is an estimated income elasticity.

These additional coefficients are simple to add in for psychological choice models too following our modifications. For the DFT simulated dataset, we incorporate underlying preferences by setting $P_{0_{jnt}} = ASC_j + ASC_{F_j} \cdot z_{F,n}$, with this having been effective previously (see results in Hancock et al. [2018]). The alternative specific travel time coefficients can be included in DFT and MLBA by multiplication of the attribute values, as we use our new scaling method (see Section 2.3.1) which means that these coefficients will have an equivalent impact on the attributes in DFT and MLBA as they would in a RUM model. Finally, in the MLBA models, we incorporate alternative specific constants ($\delta_j$) by adding them to the mean drift rates as in Equation 22. All of the values used for the parameters to generate probabilities for each alternative are given in Table 12.

### 4.2. Results for simulated data

We next test the performance of the different models across the four datasets, i.e. seeing also how well each model performs on data generated with a different model, thus giving an indication of robustness to the underlying data generation process. We conduct these tests for three different specifications of each model, namely:

1. A basic model with three alternative specific constants and three parameters for the importance of the attributes: $\beta_{AT}$, $\beta_{TC}$ and a single coefficient for travel times across all alternatives, $\beta_{TT}$. Additionally we have a parameter for income effects, $\alpha^I$.
2. The basic model with three additional mode-specific travel time coefficient multipliers, $\alpha_{TT_j}$.
3. The second model with three additional alternative specific constants, segmenting these by gender.

This gradual build up of model complexity mirrors a process that would happen in an actual specification search, allowing us to test the behaviour of DFT and MLBA in what is a common process when using more typical discrete choice models. The log-likelihood and BIC values obtained from these models are displayed in Table 11, with a plot of these values given in Figure 3. For all of the MLBA models in this section, we find that fixing $\chi$ has no significant impact on model fit.

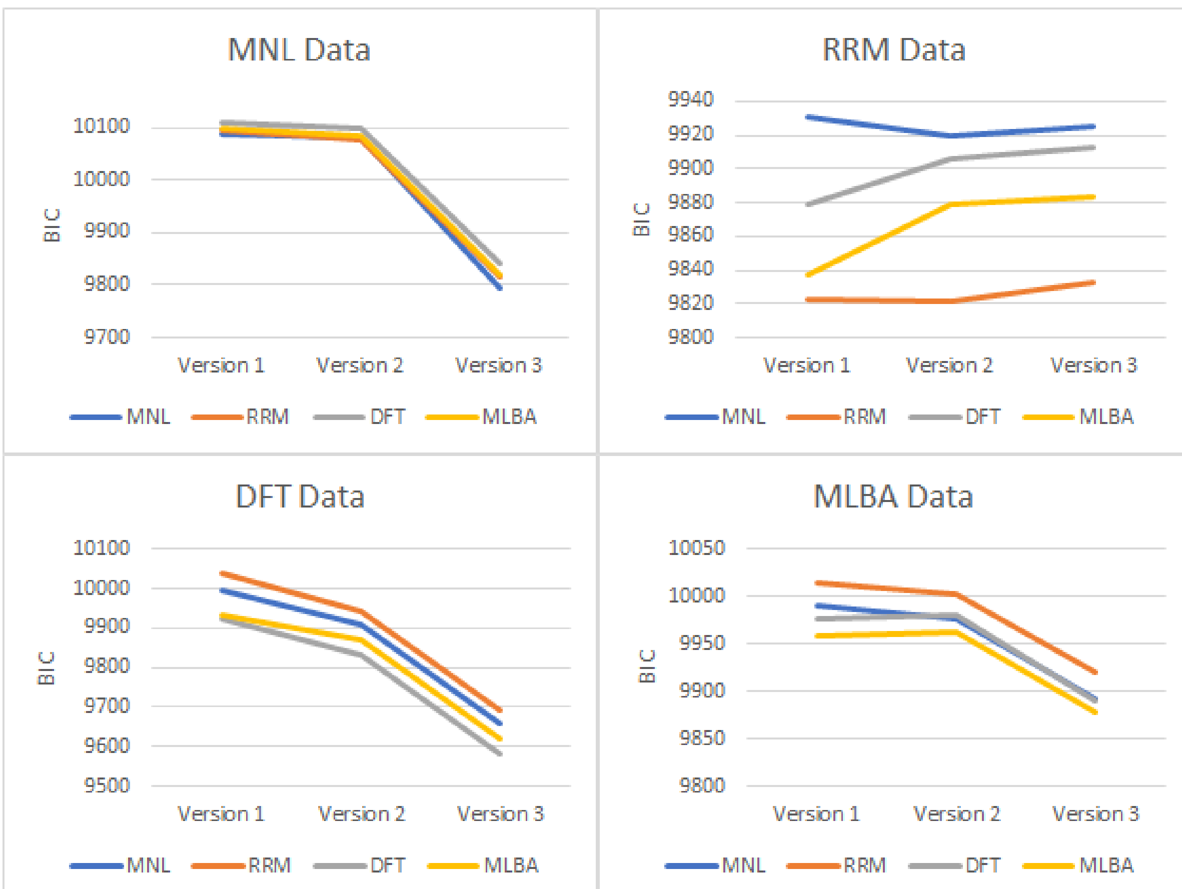
For the dataset with choices generated by MNL, the best log-likelihood is found by a full specification of MLBA, although MNL also performs well and has the lowest BIC value (highlighted in

**TABLE 11** : The BIC values obtained from models for the simulated datasets

| Model | Version | free pars. | dataset | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | MNL | | RRM | | DFT | | MLBA | |
| | | | LL | BIC | LL | BIC | LL | BIC | LL | BIC |
| MNL | 1 (basic) | 7 | -5,014.46 | 10,088.54 | -4,966.87 | 9,993.37 | -4,935.70 | 9,931.03 | -4,965.63 | 9,990.88 |
| | 2 (+times) | 10 | -4,998.31 | 10,081.79 | -4,912.66 | 9,910.49 | -4,917.08 | 9,919.33 | -4,945.91 | 9,976.99 |
| | 3 (+gender) | 13 | -4,842.60 | **9,795.92** | -4,773.88 | 9,658.48 | -4,907.49 | 9,925.70 | -4,891.11 | 9,892.95 |
| RRM | 1 (basic) | 7 | -5,018.68 | 10,096.98 | -4,912.76 | 9,885.14 | -4,957.61 | 9,974.84 | -4,977.44 | 10,014.51 |
| | 2 (+times) | 10 | -4,996.02 | 10,077.21 | -4,863.61 | 9,812.39 | -4,932.74 | 9,950.66 | -4,958.59 | 10,002.35 |
| | 3 (+gender) | 13 | -4,853.38 | 9,817.48 | **-4,727.57** | **9,565.86** | -4,923.43 | 9,957.58 | -4,904.63 | 9,919.98 |
| DFT | 1 (basic) | 10 | -5,010.61 | 10,106.40 | -4,923.17 | 9,931.52 | -4,883.07 | 9,851.31 | -4,941.57 | 9,968.31 |
| | 2 (+times) | 13 | -4,991.18 | 10,093.09 | -4,893.21 | 9,897.15 | -4,861.43 | **9,833.59** | -4,928.41 | 9,967.54 |
| | 3 (+gender) | 16 | -4,847.94 | 9,832.15 | -4,751.80 | 9,639.88 | **-4,853.03** | 9,842.34 | -4,873.10 | 9,882.48 |
| MLBA | 1 (basic) | 10 | -5,007.84 | 10,100.85 | -4,907.00 | 9,899.17 | -4,891.89 | 9,868.95 | -4,936.65 | 9,958.47 |
| | 2 (+times) | 13 | -4,987.72 | 10,086.16 | -4,879.46 | 9,869.64 | -4,884.07 | 9,878.86 | -4,925.39 | 9,961.50 |
| | 3 (+gender) | 16 | **-4,841.66** | 9,819.60 | -4,739.89 | 9,616.06 | -4,875.03 | 9,886.34 | **-4,870.56** | **9,877.40** |

Table 11). RRM unsurprisingly has the best model fit for the dataset generated by RRM, although both DFT and in particular MLBA provide markedly better fit for this dataset than MNL[13]. It also appears that, as expected, MLBA fits the MLBA generated choices with a higher log-likelihood and DFT fits the DFT generated choices best. MLBA provides significantly better fit than MNL and RRM for both the DFT and MLBA datasets, with the gap remaining fairly constant as the models become more complex (see Figure 3). The main difference between the RRM and MNL datasets compared to the MLBA and DFT datasets is that there are parameters for competition between psychologically similar alternatives in the MLBA and DFT models. It appears that MNL and RRM cannot capture this effect and thus have worse model fits for these datasets, but have far more similar model fits for the MNL generated datasets[14]. RRM appears to be the most inconsistent model, with the best fit for the RRM dataset but the worst for the DFT and MLBA datasets.

**FIGURE 3** : BIC values of the models for the simulated data



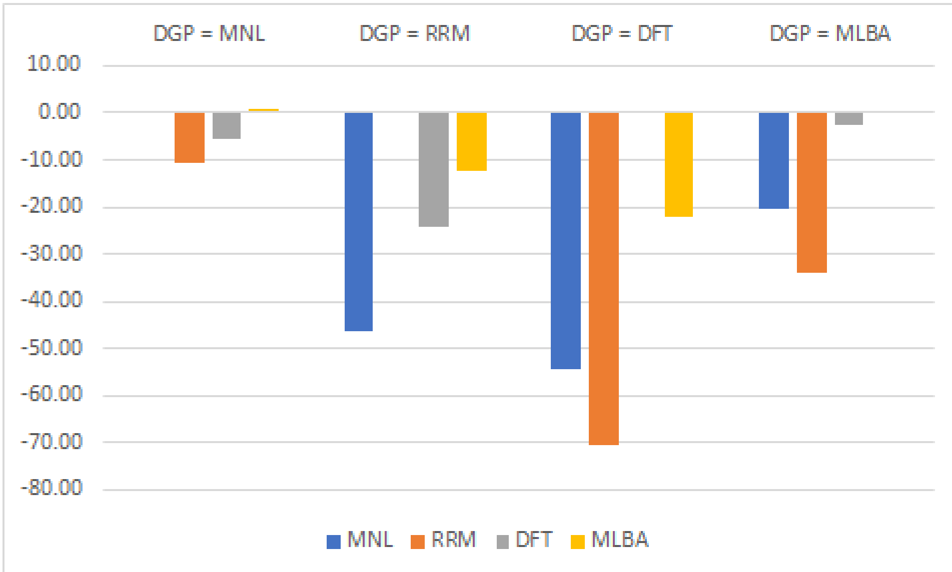

We finally contrast the fit of the full specification for each of the four models with that of the model type used for data generation. These results are shown in Figure 4 and show that DFT

[13]It is also worth noting that this is the only dataset in the paper in which RRM significantly outperforms other models. It achieves worse fit than both DFT and MLBA on all SP and RP datasets.

[14]Note that this is also suggested by the fact that the removal of DFT's feedback matrix results in a loss of 10.28 log-likelihood units for the DFT dataset, but only 0.56 units for the MNL dataset.

and MLBA show much smaller differences in fit compared to the model consistent with the data generating process (DGP). This suggests that they are more robust to potential misspecification.

**FIGURE 4** : Log-likelihood of estimated models compared to model consistent with data generating process (DGP)




### 4.3. Recovery of parameters from simulated datasets

We next consider how well the different models recover the parameter values that were used to generate the simulated datasets for the same model. Table 12 gives the parameters used in simulating the data (labelled as 'setup') as well as the parameters produced in estimation, and the difference between those two. As each model is tested against a dataset generated by the same model, we can test the stability of the parameters. Using our new scaling method allows us to use similar parameter setup values across models, with the exception that parameters are adjusted such that the data generation process has similar amounts of noise across all datasets no matter which model is used to generate the choices.

All four models appear to accurately recover the three $\beta$-coefficients associated with the explanatory variables. These appear to be more recoverable than the alternative specific constants. All four models, however, additionally perform well at recovering the attribute-specific travel time coefficients. Most crucially for DFT and MLBA, the process parameters are fairly stable.

## 5. CONCLUSIONS

In this paper, we consider two alternate accumulator choice models, developed in mathematical psychology, and compare them against models typically used in choice modelling. The models in questoin are decision field theory (DFT), a model where preferences for alternatives stochastically update over time, and the multi-attribute linear ballistic accumulator (MLBA), where the preferences for alternatives *race* towards a threshold.

**TABLE 12** : Parameter values used to generate datasets and estimates for full models for their respective datasets

| Parameter | MNL Setup | MNL Estimate | MNL Change | RRM Setup | RRM Estimate | RRM Change | DFT Setup | DFT Estimate | DFT Change | MLBA Setup | MLBA Estimate | MLBA Change |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\beta_{TT}$ | -0.0050 | -0.0044 | -12% | -0.0030 | -0.0029 | -3% | -0.0050 | -0.0050 | **fixed** | -0.0030 | -0.0023 | -23% |
| $\beta_{TC}$ | -0.0280 | -0.0279 | 0% | -0.0160 | -0.0162 | 1% | -0.0280 | -0.0308 | 10% | -0.0160 | -0.0157 | -2% |
| $\beta_{AT}$ | -0.0060 | -0.0053 | -12% | -0.0040 | -0.0046 | 15% | -0.0060 | -0.0057 | -5% | -0.0040 | -0.0040 | 0% |
| $ASC_{car}$ | -0.5000 | -0.8238 | 65% | -0.5000 | -0.6965 | 39% | -0.5000 | -1.4346 | 187% | -0.5000 | -1.3676 | 174% |
| $ASC_{air}$ | -1.5000 | -1.8053 | 20% | -1.5000 | -1.8363 | 22% | -1.5000 | -2.7972 | 86% | -1.5000 | -2.2476 | 50% |
| $ASC_{rail}$ | -1.0000 | -0.9036 | -10% | -1.0000 | -1.0067 | 1% | -1.0000 | -1.9222 | 92% | -1.0000 | -1.1817 | 18% |
| $ASC_{car_{fem}}$ | -0.5000 | -0.4752 | -5% | -0.5000 | -0.4020 | -20% | -0.5000 | -0.6330 | 27% | -0.5000 | -0.4227 | -15% |
| $ASC_{air_{fem}}$ | 0.5000 | 0.6952 | 39% | 0.5000 | 0.6822 | 36% | 0.5000 | 0.8624 | 72% | 0.5000 | 0.9142 | 83% |
| $ASC_{rail_{fem}}$ | 1.0000 | 1.1188 | 12% | 1.0000 | 1.1855 | 19% | 1.0000 | -0.1289 | -113% | 1.0000 | 1.1815 | 18% |
| $\beta_{TT_{air}}$ | 1.2500 | 1.1041 | -12% | 1.2500 | 1.7576 | 41% | 1.2500 | 1.1109 | -11% | 1.2500 | 0.6942 | -44% |
| $\beta_{TT_{rail}}$ | 2.0000 | 2.3845 | 19% | 2.0000 | 2.2579 | 13% | 2.0000 | 1.9180 | -4% | 2.0000 | 2.5291 | 26% |
| $\beta_{TT_{hsr}}$ | 1.5000 | 1.7723 | 18% | 1.5000 | 1.3528 | -10% | 1.5000 | 1.8307 | 22% | 1.5000 | 1.8596 | 24% |
| $\alpha_I$ | -0.5000 | -0.5106 | 2% | -0.5000 | -0.4962 | -1% | -0.5000 | -0.3584 | -28% | -0.5000 | -0.5218 | 4% |
| $\phi_1$ | | | | | | | 0.0500 | 0.0356 | -29% | | | |
| $\phi_2$ | | | | | | | 0.1000 | 0.1379 | 38% | | | |
| $\sigma_\varepsilon$ | | | | | | | 1.4142 | 1.4314 | 1% | | | |
| t | | | | | | | 10.0000 | 8.8319 | -12% | | | |
| A | | | | | | | | | | 2.5000 | 2.5000 | **fixed** |
| $\chi$ | | | | | | | | | | 7.5000 | 6.3023 | -16% |
| s | | | | | | | | | | 2.0000 | 2.0000 | **fixed** |
| $I_0$ | | | | | | | | | | 10.0000 | 9.5697 | -4% |
| $\lambda_1$ | | | | | | | | | | 0.1000 | 0.0872 | -13% |
| $\lambda_2$ | | | | | | | | | | 0.2000 | 0.1702 | -15% |

We first make a number of methodological developments to improve the suitability of the models for studying travel behaviour and other non-laboratory based choices. For DFT, we implement a new scaling method on the attributes, which results in a number of benefits such as the modeller not having to know the sign of the attributes before running the model. This has an immediate benefit for the UK dataset, for which one attribute (whether the delay information service is free) is a positive attribute. A comparison with other available scaling approaches in Section 3.5 also highlights the benefits of this approach.

We also consider the impacts of including parameters to capture underlying preferences in MLBA and DFT. Results from our UK dataset suggest that MLBA and DFT make substantial gains when these parameters are included and can consequently capture status quo biases. We have, however, only considered one method for incorporating preferences in these models. Whilst we add parameters to the drift rate in MLBA, alternative specifications would allow for an adjustment of the starting point $A$ or the threshold $\chi$, such that alternatives had different values for these parameters. It is easily possible that some alternatives may not require as much evidence to be chosen (for example, a commuter's usual route to work), meaning that an MLBA model including alternative specific thresholds may work well. This could be investigated in future research. The operationalisation of these two models in this paper provides promising results, and paves the way for the incorporation of data on the processes of decision-making in these models, such as eye-tracking information, response times and EEG data.

We also consider in detail the relative importance of different parameters of our models. Whereas additionally fixing the threshold parameter for MLBA does not have a significant impact for our simulated datasets, it does have an impact for our SP data. The opposite is true for the drift rate constant, $I_0$, which is important for our simulated datasets but is less important for our SP data. It is possible that the importance of these parameters varies according to how deterministic the data is and further work could test datasets with specified variations in the level of noise. This could help an analyst determine which parameters are important for MLBA for complex choice data. For DFT, it appears that our new method for the scaling of attributes significantly improves the impact of the feedback matrix parameters. It appears that the feedback matrix is not relevant for choices where there are only two alternatives. However, regardless of whether the feedback matrix has an impact or not, DFT outperforms MNL and RRM for our SP and RP datasets.

We test the models extensively using simulated data, where the findings suggest that DFT and MLBA may be less sensitive to model misspecifcation (i.e. if the estimated model differs substantially from that used for data generation) than the corresponding RUM and RRM models. Crucially, both DFT and MLBA outperform MNL and RRM across the two SP datasets and the RP dataset, including in out of sample validation for the latter, which is to the best of our knowledge the first use of both DFT and MLBA on RP data. The good model fits for both DFT and MLBA for our second stated survey dataset suggest that if there is competition between psychologically similar alternatives (when there are two alternatives that have attributes that are more similar than those of a third alternative), a move towards a choice model with psychological foundations becomes more appealing.

Moving away from RUM has obvious pitfalls, especially in terms of the use of models for welfare analysis [see e.g. Hess et al., 2018]. The evidence in this paper suggests that if an analyst is willing to accept these pitfalls, then moving further away from RUM than for example with a RRM model, may be beneficial, and models from mathematical psychology provide an interesting avenue for such work. Of course, more research is needed in terms of additional comparisons, including on larger datasets with more alternatives and attributes. Also, whilst we have considered DFT and MLBA, future research should also consider models from mathematical psychology that do not have likelihood functions. A large number of models from mathematical psychology such as the drift diffusion model [Wiecki et al., 2013], the leaky competing accumulator [Usher and McClelland, 2001] and the feed-forward inhibition model [Turner et al., 2016] can be estimated using hierarchical Bayesian estimation combined with probability density approximation [Turner and Sederberg, 2014]. This means that there is large scope for further comparisons between psychological and mainstream choice models using hierarchical Bayesian estimation, a method already popular in traditional choice modelling for mixed logit models [Train, 2001, Burda et al., 2008, Dumont et al., 2015].

Additionally, the linear ballistic accumulator [Brown and Heathcote, 2008], a simplified version of MLBA for alternatives without multiple attributes, has been demonstrated to work well with dynamic datasets where the drift rates change over time [Holmes et al., 2016]. A similar concept could be applied to both DFT and MLBA, for which changing attributes could easily be incorporated. Thus DFT and MLBA may work well with dynamic revealed preference datasets such as the lane merging decisions made by drivers, where typical choice models may not do so well due to their static nature. Complex datasets such as these, as well as datasets with additional process or psychometric measures, would also be useful for further testing the functionality and usefulness of the process parameters within both DFT and MLBA. Additionally, given that in Hancock et al. [2018], we demonstrate that DFT can efficiently incorporate random parameters, it is possible that similar adjustments could also be made for MLBA. All of these potential extensions of DFT and MLBA, combined with the results in this paper, demonstrate that accumulator models such as DFT and MLBA are attractive alternative approaches to random utility models, particularly when it comes to forecasting. It therefore appears that these models, as well as others, may hold significant promise in improving the behavioural realism in choice models, in both transport and beyond.

## ACKNOWLEDGEMENTS

The authors acknowledge the financial support by the European Research Council through the consolidator grant 615596-DECISIONS. The data used for the revealed preference analysis comes from a study conducted for the UK Department for Transport. The results reported, findings inferred, and views expressed are the responsibility of the authors, and not a statement of the Department's policy or guidance. We would also like to thank Jennifer Trueblood and two anonymous reviewers for their comments on a previous version of this paper.

**REFERENCES**

Arup, ITS Leeds and Accent (2015). Provision of market research for value of time savings and reliability. Phase 2 report to the Department for Transport.

Berkowitsch, N. A., Scheibehenne, B., and Rieskamp, J. (2014). Rigorously testing multialternative decision field theory against random utility models. *Journal of Experimental Psychology: General*, 143(3):1331.

Berkowitsch, N. A., Scheibehenne, B., Rieskamp, J., and Matthäus, M. (2015). A generalized distance function for preferential choices. *British Journal of Mathematical and Statistical Psychology*, 68(2):310–325.

Bhatia, S. (2013). Associations and the accumulation of preference. *Psychological Review*, 120(3):522.

Bierlaire, M., Thémans, M., and Zufferey, N. (2010). A heuristic for nonlinear global optimization. *INFORMS Journal on Computing*, 22(1):59–70.

Brown, S. D. and Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology*, 57(3):153–178.

Bruno, H. A., Cebollada, J., and Chintagunta, P. K. (2015). Brand preference and state-dependence with intra-household heterogeneity.

Burda, M., Harding, M., and Hausman, J. (2008). A bayesian mixed logit–probit model for multinomial choice. *Journal of Econometrics*, 147(2):232–246.

Busemeyer, J. R. and Diederich, A. (2002). Survey of decision field theory. *Mathematical Social Sciences*, 43(3):345–370.

Busemeyer, J. R. and Townsend, J. T. (1992). Fundamental derivations from decision field theory. *Mathematical Social Sciences*, 23(3):255–282.

Busemeyer, J. R. and Townsend, J. T. (1993). Decision field theory: a dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review*, 100(3):432.

Cataldo, A. M. and Cohen, A. L. (2018). Reversing the similarity effect: The effect of presentation format. *Cognition*, 175:141–156.

Chernev, A. (2004). Extremeness aversion and attribute-balance effects in choice. *Journal of consumer research*, 31(2):249–263.

Chorus, C. G. (2010). A new model of random regret minimization. *EJTIR, 10 (2), 2010*.

Chorus, C. G., Arentze, T. A., and Timmermans, H. J. (2008). A random regret-minimization model of travel choice. *Transportation Research Part B: Methodological*, 42(1):1–18.

Cohen, A. L., Kang, N., and Leise, T. L. (2017). Multi-attribute, multi-alternative models of choice: Choice, reaction time, and process tracing. *Cognitive Psychology*, 98:45–72.

Daly, A., Hess, S., and de Jong, G. (2012). Calculating errors for measures derived from choice modelling estimates. *Transportation Research Part B*, 46(2):333–341.

Diederich, A. (2003). Mdft account of decision making under time pressure. *Psychonomic Bulletin & Review*, 10(1):157–166.

Dumont, J., Keller, J., and Carpenter, C. (2015). Rsghb: functions for hierarchical bayesian estimation: a flexible approach. *R package version*, 1(2).

Eddelbuettel, D., François, R., Allaire, J., Ushey, K., Kou, Q., Russel, N., Chambers, J., and Bates, D. (2011). Rcpp: Seamless r and c++ integration. *Journal of Statistical Software*, 40(8):1–18.

Fosgerau, M. (2006). Investigating the distribution of the value of travel time savings. *Transportation Research Part B: Methodological*, 40(8):688–707.

Giergiczny, M., Hess, S., Dekker, T., and Chintakayala, P. K. (2013). Testing the consistency (or lack thereof) between choices in best-worst surveys. In *Third International Choice Modelling Conference, Sydney*.

Gold, J. I. and Shadlen, M. N. (2000). Representation of a perceptual decision in developing oculomotor commands. *Nature*, 404(6776):390–394.

González-Valdés, F. and Ortúzar, J. d. D. (2017). The stochastic satisficing model: A bounded rationality discrete choice model. *Journal of Choice Modelling*, 27:74–87.

Hancock, T. O., Hess, S., and Choudhury, C. F. (2018). Decision field theory: Improvements to current methodology and comparisons with standard choice modelling techniques. *Transportation Research Part B: Methodological*, 107:18–40.

Hawkins, G. E., Marley, A., Heathcote, A., Flynn, T. N., Louviere, J. J., and Brown, S. D. (2014). Integrating cognitive process and descriptive models of attitudes and preferences. *Cognitive Science*, 38(4):701–735.

Heathcote, A. and Love, J. (2012). Linear deterministic accumulator models of simple choice. *Frontiers in psychology*, 3:292.

Henningsen, A. and Toomet, O. (2011). maxlik: A package for maximum likelihood estimation in R. *Computational Statistics*, 26(3):443–458.

Hess, S., Daly, A., and Batley, R. (2018). Revisiting consistency with random utility maximisation: theory and implications for practical work. *Theory and Decision*, 84(2):181–204.

Hess, S. and Palma, D. (2019). Apollo: a flexible, powerful and customisable freeware package for choice model estimation and application, www.apollochoicemodelling.com.

Hess, S., Rose, J. M., and Hensher, D. A. (2008). Asymmetric preference formation in willingness to pay estimates in discrete choice models. *Transportation Research Part E: Logistics and Transportation Review*, 44(5):847–863.

Hess, S. and Stathopoulos, A. (2013). A mixed random utility - random regret model linking the choice of decision rule to latent character traits. *Journal of Choice Modelling*, 9:27–38.

Holmes, W. R., Trueblood, J. S., and Heathcote, A. (2016). A new framework for modeling decisions about changing information: The piecewise linear ballistic accumulator model. *Cognitive Psychology*, 85:1–29.

Hotaling, J. M., Busemeyer, J. R., and Li, J. (2010). Theoretical developments in decision field theory: comment on Tsetsos, Usher, and Chater (2010). *Psychological Review*, 117(4):1294–1298.

Kaufman, B. E. (1990). A new theory of satisficing. *Journal of Behavioral Economics*, 19(1):35–51.

Krajbich, I., Lu, D., Camerer, C., and Rangel, A. (2012). The attentional drift-diffusion model extends to simple purchasing decisions. *Frontiers in psychology*, 3:193.

Liu, Y. and Cirillo, C. (2018). A generalized dynamic discrete choice model for green vehicle adoption. *Transportation Research Part A: Policy and Practice*, 114:288–302.

Masiero, L. and Hensher, D. A. (2010). Analyzing loss aversion and diminishing sensitivity in a freight transport stated choice experiment. *Transportation Research Part A: Policy and Practice*, 44(5):349–358.

McFadden, D. (1974). Conditional Logit Analysis of Qualitative Choice Behaviour. In Frontiers in Econometrics, ed. P. Zarembka. (New York: Academic Press).

McNair, B. J., Hensher, D. A., and Bennett, J. (2012). Modelling heterogeneity in response behaviour towards a sequence of discrete choice questions: a probabilistic decision process model. *Environmental and Resource Economics*, 51(4):599–616.

Miletić, S., Turner, B. M., Forstmann, B. U., and van Maanen, L. (2017). Parameter recovery for the leaky competing accumulator model. *Journal of Mathematical Psychology*, 76:25–50.

Noguchi, T. and Stewart, N. (2014). In the attraction, compromise, and similarity effects, alternatives are repeatedly compared in pairs on single dimensions. *Cognition*, 132(1):44–56.

Otter, T., Johnson, J., Rieskamp, J., Allenby, G. M., Brazell, J. D., Diederich, A., Hutchinson, J. W., MacEachern, S., Ruan, S., and Townsend, J. (2008). Sequential sampling models of choice: Some recent advances. *Marketing letters*, 19(3-4):255–267.

Qin, H., Guan, H., and Wu, Y.-J. (2013). Analysis of park-and-ride decision behavior based on decision field theory. *Transportation Research Part F: Traffic Psychology and Behaviour*, 18:199–212.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, 85(2):59.

Roe, R. M., Busemeyer, J. R., and Townsend, J. T. (2001). Multialternative decision field theory: A dynamic connectionist model of decision making. *Psychological Review*, 108(2):370.

Rose, J. M. (2014). Interpreting discrete choice models based on best-worst data: a matter of framing. *The Transportation Research Board (TRB) 93rd Annual Meeting*.

Sanderson, C. and Curtin, R. (2016). Armadillo: a template-based c++ library for linear algebra. *Journal of Open Source Software*.

Schwartz, B., Ward, A., Monterosso, J., Lyubomirsky, S., White, K., and Lehman, D. R. (2002). Maximizing versus satisficing: happiness is a matter of choice. *Journal of Personality and Social Psychology*, 83(5):1178.

Stathopoulos, A. and Hess, S. (2012). Revisiting reference point formation, gains–losses asymmetry and non-linear sensitivities with an emphasis on attribute specific treatment. *Transportation Research Part A: Policy and Practice*, 46(10):1673–1689.

Stern, E. and Richardson, H. W. (2005). Behavioural modelling of road users: current research and future needs. *Transport Reviews*, 25(2):159–180.

Stewart, N. and Simpson, K. (2008). A decision-by-sampling account of decision under risk. *The probabilistic mind: Prospects for Bayesian cognitive science*, pages 261–276.

Swait, J. (2001). A non-compensatory choice model incorporating attribute cutoffs. *Transportation Research Part B: Methodological*, 35(10):903–928.

Train, K. (2001). A comparison of hierarchical bayes and maximum simulated likelihood for mixed logit. *University of California, Berkeley*, pages 1–13.

Trueblood, J., Brown, S., and Heathcote, A. (2013a). The multi-attribute linear ballistic accumulator model of decision-making. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 35.

Trueblood, J. S., Brown, S. D., and Heathcote, A. (2014). The multiattribute linear ballistic accumulator model of context effects in multialternative choice. *Psychological review*, 121(2):179–205.

Trueblood, J. S., Brown, S. D., Heathcote, A., and Busemeyer, J. R. (2013b). Not just for consumers context effects are fundamental to decision making. *Psychological science*, 24(6):901–908.

Turner, B. M., Schley, D. R., Muller, C., and Tsetsos, K. (2017). Competing theories of multialternative, multiattribute preferential choice.

Turner, B. M. and Sederberg, P. B. (2014). A generalized, likelihood-free method for posterior estimation. *Psychonomic bulletin & review*, 21(2):227–250.

Turner, B. M., Sederberg, P. B., and McClelland, J. L. (2016). Bayesian analysis of simulation-based models. *Journal of Mathematical Psychology*, 72:191–199.

Tversky, A. (1972). Elimination by aspects: A theory of choice. *Psychological Review*, 79(4):281.

Tversky, A. (1977). Features of similarity. *Psychological Review*, 84(4):327.

Tversky, A. and Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, 5(2):207–232.

Usher, M. and McClelland, J. L. (2001). The time course of perceptual choice: the leaky, competing accumulator model. *Psychological Review*, 108(3):550.

Wiecki, T. V., Sofer, I., and Frank, M. J. (2013). HDDM: hierarchical bayesian estimation of the drift-diffusion model in python. *Frontiers in neuroinformatics*, 7:14.

Zhao, G., Wu, C., and Ou, B. (2011). Mathematical modeling of average driver speed control with the integration of queuing network-model human processor and rule-based decision field theory. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 55, pages 856–860. Sage Publications.

# A. APPENDIX: ALTERNATIVE VERSIONS OF MLBA

Whilst we use the mainstream version of MLBA [Trueblood et al., 2014] in this paper, it should be noted that the original version of MLBA [Trueblood et al., 2013a] has also not been tested on large-scale consumer choice data. Whilst this version of MLBA, here denoted '$MLBA_0$' uses the same start, threshold and standard deviation for its drift rates, it differs in the specification for the value of the mean drift rate:

$$d_j = \frac{10}{1 + exp(-\gamma \cdot v_j)} \tag{27}$$

where $v_j$ is a valence function and $\gamma$ is a logistic parameter. Small values of the logistic parameter $\gamma$ would result in $exp(-\gamma \cdot v_j) \to 1$, meaning that the valences, $v_j$, are less influential and the probabilities of the alternatives become more similar, resulting in a less deterministic choice. The valences are similar to a decision field theory model's valences with the exception that they attempt to additionally capture the comparison process achieved by DFT's feedback matrix. Thus we have

$$V = C \cdot M \cdot W \tag{28}$$

where $W$ is a vector comprising of a set of attribute importance weights that sum to 1, $M$ is the attribute matrix and C is a $n \times n$ comparison matrix ($n$ being the number of alternatives) with diagonal entries of 1 and off-diagonal elements:

$$C_{i,j \neq i} = \frac{exp(-\phi \cdot Dist_{i,j}) - 1}{n-1}. \tag{29}$$

Finally, $\phi$ is a sensitivity parameter such that high values result in the distance between the attributes of the alternatives becoming insignificant. Low values allow for more similar alternatives to compete more with each other relative to less similar alternatives.

Results from applying the previous version of MLBA to both of the SP datasets and the RP dataset are given in Table 13 below.

**TABLE 13** : Comparison of different versions of MLBA

| Dataset | $MLBA_0$ | MLBA | Difference |
|---|---|---|---|
| Danish | -2,189.78 | -2,010.46 | -179.32 |
| UK | -3,394.36 | -3,322.36 | -72.00 |
| RP | -375.24 | -352.07 | -23.17 |

From these results, it appears that the old version of MLBA has far inferior fits compared to that of the mainstream MLBA. Consequently, it would appear that modellers should focus on the mainstream version of MLBA.