Australasian Transport Research Forum 2021 Proceedings
8-10 December, Brisbane, Australia
Publication website: http://www.atrf.info

Decision field theory: an extension for real-world settings

Thomas O. Hancock[1], Stephane Hess[1], Charisma F. Choudhury[1], Panagiotis Tsoleridis[1]

[1]Choice Modelling Centre and Institute for Transport Studies, 34-40 University Road, University of Leeds, Leeds, LS2 9JT, United Kingdom.

Email for correspondence: T.O.Hancock@leeds.ac.uk

Abstract

Decision field theory (DFT) is a model that was originally developed within the context of cognitive psychology to explain phenomena not expected under classical choice models. This meant that the model was initially designed to explain choice behaviour observed under controlled laboratory settings. Recent work has improved the mathematical foundations of DFT, such that it has become a tractable and more rigorous model that is easier to apply to a wider variety of choice contexts. In particular, the inclusion of attribute importance parameters has led to successful applications to stated preference data including travel mode choice. However, thus far, implementations to real-life behaviour have been limited. The aim of this paper is to extend decision field theory such that it can take further steps towards accounting for real-world behaviour and a wider variety of contexts, in general. First, we give theoretical extensions for the model, demonstrating that relaxing the assumptions around the normal error term within DFT can lead to more flexible structures. Second, we demonstrate on two large-scale case studies of revealed preference mode choice behaviour in the UK that DFT can incorporate a range of sociodemographic variables. Thirdly, we demonstrate that our new 'heteroskedastic' DFT model substantially outperforms the original version of DFT, as well as alternative econometric choice models.

1. Introduction

Decision field theory (DFT) is a dynamic, stochastic model, introduced by Busemeyer and Townsend (1992, 1993). The key idea within DFT is that the preferences for different alternatives update over time whilst the decision-maker considers the different alternatives and their attributes. At some timepoint, the decision-maker reaches a conclusion when the preference for an alternative reaches some (internal) threshold, or the decision-maker runs out of time in which to deliberate on their options (an external threshold). A graphical representation of the preference evolution process under a DFT model is given in Figure 1. Different alternatives may be chosen depending on whether an internal (the grey horizontal line representing a preference value of 3) or an external threshold (represented by a vertical black line at 10 preference updating steps) is assumed.

DFT was originally developed to explain decision-making in an uncertain environment or under time pressure, but has subsequently been reformulated and adapted for modelling choice behaviour when an individual can choose multiple alternatives (Roe et al., 2001). There are multiple conceptual reasons for adopting DFT models, as they specifically attempt to capture the decision-making process, with Busemeyer et al. (2006) amongst others arguing that a better representation of the decision-making process may also lead to a better representation of outcomes. In particular, DFT attempts to mimic the mental deliberation process through preference changes for alternatives whilst the decision-maker considers different attributes of these

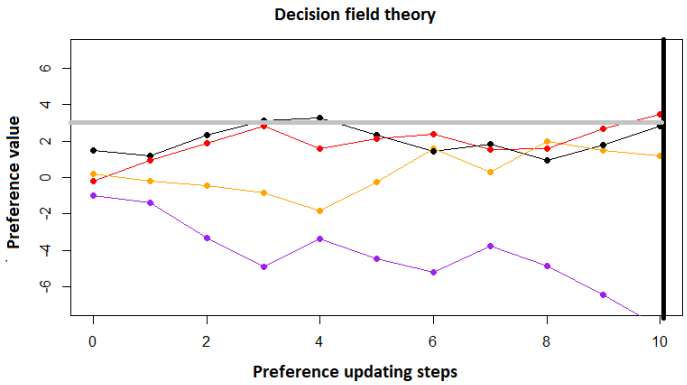

FIGURE 1 : An example of the evolution of preferences for different alternatives under a DFT model (adapted from Hancock et al. 2021).

.

alternatives, with the preferences accumulating until some threshold is reached. Notably, neurobiologists have demonstrated that neural activation patterns appear similar to the accumulation of preference process under a DFT model (Gold and Shadlen, 2000; Ratcliff et al., 2003). There are also empirical benefits, with DFT capturing contextual phenomena that are not typically accounted for by standard choice theory, such as the similarity, attraction and compromise effects. This has typically been the focus of implementations of DFT within the context of cognitive psychology: how well it can explain these contextual effects (Roe et al., 2001; Pettibone, 2012; Berkowitsch et al., 2014). However, a key limitations of the model at this point was computational restrictions (Otter et al., 2008) that often led to restricted forms of DFT being applied. For example, the probabilities of different alternatives can be easily calculated if it is assumed that decision-makers only make decisions when their preference values for the different alternatives have stabilised (Berkowitsch et al., 2014). More recently, these restrictions have been lifted through improvements to the underlying mathematical mechanisms (Hancock et al., 2019). This allows for easier implementation of DFT models, and has led to its inclusion in the popular choice modelling R package *apollo* (Hess and Palma, 2019).

The current most flexible form of the model is the specification by Hancock et al. (2021), which incorporates relative importance scaling parameters for different attributes, allowing for the inclusion of sociodemographic effects with equivalent functions to those used in random utility models. This means that, for example, cost sensitivity can vary as a function of an individual's income, without the importance of other attributes being impacted (as was the case in the original scale-variant version of DFT, see Hancock et al. (2019) for details). This is an important step towards applications of DFT to real-world data, with Hancock et al. (2021) demonstrating that DFT could be applied to model the chosen service provider for train trips to London. This also led to DFT outperfoming standard multinomial logit models, as well as models based on random regret minimisation. However, whilst Hancock et al. (2019) demonstrated that random parameters could be incorporated within DFT, it still does not have the same flexibility as standard choice models, particularly in regard to the possible substitution patterns that can exist between alternatives. The aim of this paper is to make further extensions towards increasing the flexibility of the model. We focus on the development of a 'heteroskedastic' decision field theory, which relaxes the assumption of identical normal error terms being added to the different preference values for each alternative during the deliberation process. The rest of this paper is arranged as follows. First, we describe the current implementation of decision field theory. We then detail the new theory and extensions for the model. Next, we consider two real-world case studies of mode choice in the UK. Finally, we draw some conclusions and discuss options for future

research.

2. Decision field theory: standard implementation and new extensions

In this section, we first describe in detail the current state of the art for implementing a decision field theory (DFT) model. We then discuss how DFT generates probabilistic choices before detailing the new theory for extending DFT, such that heteroskedasticity can be accounted for.

2.1. Standard implementation

Under a decision field theory (DFT) model, it is assumed that for an individual choice context (s), a decision-maker (n) has a 'preference value' for each of the different J alternatives. This can be represented by a preference vector $\boldsymbol{P}_{ns,0}$, which gives the initial preference towards the different alternatives prior to any consideration of them. This vector is often assumed to be a set of zeros, but could also be influenced by the status quo or prior beliefs regarding the alternatives held by the decision-maker. It is then further assumed that a decision-maker 'updates' their preferences by considering the levels of one particular attribute across the full set of alternatives, such that the preference at the next step is:

$$\boldsymbol{P}_{ns,\tau+1} = S_{ns} \cdot \boldsymbol{P}_{ns,\tau} + \boldsymbol{V}_{ns,\tau+1}, \tag{1}$$

where τ represents the completed number of preference updating steps, S is a feedback matrix and $\boldsymbol{V}$ is a random 'valence' vector, which gives information on the attribute attended to in the current step $\tau+1$.

The feedback matrix S contains two 'process parameters' that define how the preference state evolves over time:

$$S_{ns} = I_{ns} - \phi_2 \times \exp(-\phi_1 \times D_{ns}^2), \tag{2}$$

where ϕ_1 is a sensitivity parameter, which allows for similar (subjective) alternatives to subtract more preference from each other, thus controlling for the level of contextual effects predicted by the model (Roe et al., 2001). ϕ_2 is a memory parameter, which controls the rate of decay of the previously accumulated preference, meaning that more recently attended attributes will be more important. I_{ns} is an identity matrix of size $J \times J$ and D_{ns} is a matrix containing the Euclidean distances between the full set of pairs of alternatives measured with respect to the attribute-levels and the relative importance of the different attributes.

We next detail the random valence vector. At step τ, we have $\boldsymbol{V}_{ns,\tau}$, which can be calculated as:

$$\boldsymbol{V}_{ns,\tau} = C_{ns} \cdot M_{ns} \cdot \beta \cdot \boldsymbol{W}_{ns,\tau} + \boldsymbol{\varepsilon}_{ns,\tau}, \tag{3}$$

where C_{ns} is a contrast matrix used to rescale the attributes, such that they sum to zero (see Roe et al. 2001). M_{ns} is the matrix of attribute values for all of the alternatives, where each element is multiplied by a corresponding attribute scaling coefficient, contained on the diagonal matrix β. We also have a random column vector, $\boldsymbol{W}_{ns,\tau}$, of size K, where K is the number of attributes. It is comprised of a set of zeros with a 1 on the k^{th} entry, corresponding to the attribute that is (randomly) attended to by the decision-maker at preference updating step τ. Uniform draws can be used to simulate a DFT process and weights for the likelihood of attending the different attributes, w_k, where $\sum_{k=1}^{K} w_k = 1$, and w_k represents the probability that the decision-maker attends to attribute k. Finally, we have a 'process noise' entered through the variance of the error term, with $\boldsymbol{\varepsilon}_{ns,\tau} = [\varepsilon_1..\varepsilon_J]'$, and $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$, distributed identically and independently across alternatives, steps, individuals and choice tasks.

To estimate the probabilities of different alternatives being chosen under DFT, we require the expected value and the covariance of the preference values ($\boldsymbol{\xi}_{ns,\tau}$ and $\Omega_{ns,\tau}$), where

$$E[\boldsymbol{P}_{ns,\tau}] = \boldsymbol{\xi}_{ns,\tau} = (I_{ns} - S_{ns})^{-1}(I_{ns} - S_{ns}^{\tau}) \cdot \boldsymbol{\mu}_{ns} + S_{ns}^{\tau} \cdot \boldsymbol{P}_{ns,0}, \tag{4}$$

and

$$Cov[\boldsymbol{P}_{ns,\tau}] = \Omega_{ns,\tau} = \sum_{r=0}^{\tau-1} \left[S_{ns}^{r} \cdot \Phi_{ns} \cdot S_{ns}^{r'} \right]. \tag{5}$$

with $\boldsymbol{\mu}_{ns}$ and Φ_{ns} the expectation and covariance of the valence vector, respectively (Roe et al., 2001). These are calculated $\boldsymbol{\mu}_{ns} = C_{ns} \cdot M_{ns} \cdot \beta \cdot \boldsymbol{w}$, where $\boldsymbol{w}$ is the vector of attribute attention weights w_k, and:

$$Cov[\boldsymbol{W}_{ns,\tau}] = \Phi_{ns} = C_{ns} \cdot M_{ns} \cdot \beta \cdot \Psi_{ns} \cdot \beta \cdot M'_{ns} \cdot C'_{ns} + \sigma_{\varepsilon}^2 \cdot I. \tag{6}$$

Full formulations and derivations for these and the above equations given by Roe et al. (2001); Hancock et al. (2019, 2021). This results in a choice probability for choosing alternative j from the set CS_{ns} at step τ of:

$$\begin{aligned} &Prob\left[\boldsymbol{P}_{ns,\tau}[j] = \max_{i \in CS_{ns}} \boldsymbol{P}_{ns,\tau}[i]\right] = \\ &\int_{\boldsymbol{X}_{ns,\tau}>0} exp\left[-(\boldsymbol{X}_{ns,\tau} - \boldsymbol{\Gamma}_{ns,\tau})' \Lambda_{ns,\tau}^{-1} (\boldsymbol{X}_{ns,\tau} - \boldsymbol{\Gamma}_{ns,\tau})/2\right] / (2\pi |\Lambda_{ns,\tau}|^{0.5}) dX. \end{aligned} \tag{7}$$

with full details of this calculation again given by Hancock et al. (2021). As these probabilities are estimated with the use of multivariate normal distributions, there are requirements for the normalisation of scale and location. Numerous possibilities exist and are discussed by Hancock et al. (2021). Under typical implementations of DFT and the work in this paper, only one normalisation is required, with the size of the variance of the process noise, σ_{ε}^2, fixed to a value of 1.

2.2. Heteroskedastic DFT

Mathematically, there are of course similarities between DFT and probit, which are both in part reliant on normal error terms. As the sum of normal variables are themselves normal, the fact that the interpretation of the error under DFT (it is 'process' noise that occurs at each preference updating step, see Equation 3) differs from that of probit is of little consequence. This is particularly the case for DFT models, which estimate a large number of preference updating steps (e.g. Hancock (2019), Chapter 6) or are based on situations, where it is assumed that preferences stabilise over time (e.g. Berkowitsch et al. 2014). However, there are also key differences in the psychological assumptions of the model that also result in mathematical differences. Firstly, DFT incorporates two 'sources of error': the random attribute attendance, as well as the normal error. Secondly, the covariance under DFT is explicitly dependent on the attributes of the alternatives in the current choice set (M_{ns} enters Φ_{ns}), whereas standard implementations of probit models estimate covariance terms that are at an aggregate level, i.e. not necessarily dependent on attribute levels in individual choice contexts. This is an issue for DFT in that it cannot capture observed correlations between alternatives at an aggregate level, as its normal error term is distributed identically and independently across alternatives, deliberation steps, individuals and choice tasks. For the work in this paper, we consider the possibility of relaxing this assumption, such that the error terms are not identical across alternatives. We thus define an error matrix Δ, which replaces $\sigma_{\varepsilon}^2 \cdot I$ in Equation 6. It is a diagonal matrix with different elements $\sigma_{\varepsilon,j}^2$ for each alternative j. This essentially means that the size of the variance of the error term for each alternative is different. For the calculation of the probabilities of choosing different alternatives under this more flexible 'heteroskedastic' DFT, the only change is indeed in the calculation of the covariance of $\boldsymbol{W}_{ns,\tau}$ in Equation 6. Fixing one of these elements in estimation will avoid identification issues in line with

the previous version of DFT (see Table 1 of Hancock et al. 2021). The original version of DFT is of course a restricted form of the heteroskedastic version, where insignificant differences between different $\sigma^2_{\varepsilon,j}$ result in the latter specification collapsing back to the original DFT specification.

3. Empirical work

In this section, we first give an outline of the different models that we use for comparison purposes with our proposed DFT specifications. We then present the two different case studies in turn. For each, we present the dataset before detailing the specifications for our different models. We then initially compare our base DFT model against a standard multinomial logit model before giving results of our extended versions of DFT against a number of alternative models.

3.1. Models for comparison

We first discuss some different possible models that we use and compare against our DFT models. For our basic DFT model without heteroskedastic errors, we use a standard random utility model (multinomial logit, MNL, with the formula for generating probabilities of alternatives being chosen under this and all other competing models given in Appendix A). This is a comparable model as in the context of choice-only data (as used in the two case studies here) both models accept the same inputs (choice data and attribute levels for the different alternatives) and return probabilities. For the heteroskedastic DFT model, additional flexibility in the formulation of the process noise allows for more complex patterns of behaviour, thus we require comparisons with alternative choice models that allow for a similar flexibility in error structure. We thus use two extensions of the MNL model: nested logit and a heteroskedastic MNL model with equivalent flexibility (through the addition of a corresponding scaling parameter for the utility of each alternative). Furthermore, asymmetric choice models, which are variants of traditional random utility maximisation models (e.g. MNL), aim to capture the asymmetry in the distribution of the error term that arises when there are large differences in the observed shares of different alternatives in the data. In the two datasets we have used, preliminary analyses show that there is a presence of imbalance in the choices, with 43% and 48% shares for private car, respectively, in comparison to only 3% shares for cycling. This prompt us to explore asymmetric logit models. We chose Scobit, uneven logit and asymmetric logit as they are the most commonly used asymmetric models (Brathwaite and Walker, 2018).

3.2. First case study

For the first case study, we use the freely available London Passenger Mode Choice (LPMC), which was collected via the London travel demand survey, where for each observation where an individual chose to travel by one of the four modes (walk, cycle, transit and drive), the data was augmented such that attributes for these four alternatives were derived for all observations (Hillel et al., 2018). We restrict observations to home-based trips reported by individuals who are at least 12 years old, meaning that we use the same subset as Krueger et al. (2020). The resulting dataset contains a total of 58,584 trip observations. 10% of the sample is set aside for out-of-sample validation, resulting in 52,726 observations in our estimation subset. For full details of the dataset, readers should refer to (Hillel et al., 2018).

3.2.1. Model specification

For our utility-based models, we follow the specification by Krueger et al. (2020), including a number of marginal utility parameters β for alternative-specific attributes, as well as parameters for sociodemographic variables α. Note that although we do estimate alternative specific constants, they are not included in the equations below.

The utility for walking for individual n in choice scenario s is defined:

$$V_{ns,walk} = \beta_{tt} \cdot x_{walk,tt}, \tag{8}$$

where $x_{walk,tt}$ is the travel time for walking.

The utility for cycling is defined:

$$\begin{aligned} V_{ns,cycle} &= \beta_{tt} \cdot x_{cycle,tt} && \text{(9a)} \\ &+ \alpha_{cycle,fem} \cdot n_{fem} + \alpha_{cycle,age} \cdot n_{age} + \alpha_{cycle,winter} \cdot z_{winter}, && \text{(9b)} \end{aligned}$$

where $x_{cycle,tt}$ is the travel time for cycling, n_{fem} is a dummy variable set to a value of 1 if individual n is female, n_{age} is a dummy variable set to a value of 1 if individual n is either under 18 or over 64 years old, and z_{winter} is a dummy variable that is included if the trip was made between November and March.

The utility for transit is defined:

$$\begin{aligned} V_{ns,transit} &= \beta_{tt} \cdot x_{transit,ovtt} + \beta_{ivtt} \cdot x_{transit,ivtt} + \beta_{cost} \cdot x_{transit,cost} + \beta_{ch} \cdot x_{transit,ch} && \text{(10a)} \\ &+ \alpha_{transit,fem} \cdot n_{fem} + \alpha_{transit,u18} \cdot n_{u18} + \alpha_{transit,o64} \cdot n_{o16}, && \text{(10b)} \end{aligned}$$

where $x_{transit,ovtt}$ is the out-of-vehicle transit time corresponding to access time plus interchange time, $x_{transit,ivtt}$ is the in-vehicle travel time, $x_{transit,cost}$ is the cost, $x_{transit,ch}$ is the number of required interchanges and n_{u18} and n_{o64} are dummy variables equal to a value of 1 if individual n is under the age of 18 or over the age of 64, respectively.

The utility for driving is defined:

$$\begin{aligned} V_{ns,drive} &= \beta_{ivtt} \cdot x_{drive,ivtt} + \beta_{cost} \cdot x_{drive,cost} + \beta_{tv} \cdot x_{drive,tv} && \text{(11a)} \\ &+ \alpha_{drive,fem} \cdot n_{fem} + \alpha_{drive,u18} \cdot n_{u18} + \alpha_{drive,o64} \cdot n_{o16} + \alpha_{drive,cars} \cdot n_{cars}, && \text{(11b)} \end{aligned}$$

where $x_{drive,ivtt}$ is the in-vehicle travel time, $x_{drive,cost}$ is the cost, $x_{drive,tv}$ is the traffic variability and n_{cars} is the number of household cars for individual n.

3.2.2. DFT specifications

For our DFT models, we define the attribute matrix M_{ns} using the same set of variables as in our utility specifications. We thus have:

$$M_{ns} = \begin{bmatrix} \eta_{walk} & x_{walk,tt} & 0 & 0 \\ \eta_{cycle} & x_{cycle,tt} & 0 & 0 \\ \eta_{transit} & x_{transit,ovtt} & x_{transit,ivtt} & x_{transit,cost} \\ \eta_{drive} & 0 & x_{drive,ovtt} & x_{drive,cost} \end{bmatrix}, \tag{12}$$

where all variables not relating to cost or time are summed within the alternative-specific η (i.e. $\eta_{cycle} = \alpha_{cycle,fem} \cdot n_{fem} + \alpha_{cycle,age} \cdot n_{age} + \alpha_{cycle,winter} \cdot z_{winter}$). The marginal utility parameters for cost and time are then included within the DFT β matrix (see Equation 3, which is thus a diagonal matrix with elements: 1, β_{tt}, β_{ivtt}, β_{cost}). Note that we follow Table 1 of Hancock et al. (2021) in fixing the error term, thus our DFT model has four additional estimated parameters over the base MNL model. These are the three estimated psychological parameters and an additional initial preference parameter (δ_{all}, which is added to all initial preferences such that we can still estimate whether the initial preferences are significantly different from each other). We do not need to normalise an initial preference parameter as they are multiplied by a factor of the feedback matrix S when they enter the expectation of the preference vector after τ steps (see Equation 4). As S has a value that depends on the attribute-levels in a given choice context, adding the same value to all parameters for the initial preferences results in a change that is also dependent on the choice context. Our heteroskedastic DFT model has three additional parameters corresponding to the estimated alternative-specific error terms.

3.2.3. Basic model results

We first give the results of our basic MNL and DFT models, with model outputs and parameter estimates given in Table 1.

TABLE 1 : Basic model results for the London passenger mode choice dataset.

	MNL		DFT	
Log-likelihood (estimation)	-39,917.79		-38,647.77	
Estimated parameters	18		22	
Adj. ρ^2	0.4539		0.4710	
BIC	80,031.29		77,534.75	
Log-likelihood (holdout)	-4,428.57		-4,285.31	
	est	rob. t-rat.	est	rob. t-rat.
δ_{all}	0.0000	NA	-422.0176	-3.57
δ_{walk}	0.0000	NA	0.0000	NA
δ_{cycle}	-3.0983	-50.57	-13.9375	-30.96
$\delta_{transit}$	-1.5149	-35.67	-8.8786	-27.15
δ_{drive}	-2.7170	-42.13	-15.2355	-26.47
β_{cost}	-0.1890	-21.68	-0.1293	-10.45
β_{tt}	-6.6805	-60.77	-6.5441	-18.14
β_{ivtt}	-3.1821	-28.26	-1.7036	-13.96
β_{tv}	-3.4302	-34.24	-2.3404	-11.72
β_{ch}	0.7484	26.00	0.3669	7.34
$\alpha_{cycle,fem}$	-1.1009	-12.74	-0.7310	-7.83
$\alpha_{cycle,winter}$	-0.3343	-3.86	-0.2932	-4.55
$\alpha_{cycle,age}$	-0.7924	-6.57	-0.5240	-3.87
$\alpha_{transit,fem}$	0.1995	4.93	0.1369	4.03
$\alpha_{transit,u18}$	0.3303	5.31	0.2967	5.34
$\alpha_{transit,o64}$	0.5631	9.84	0.5161	9.53
$\alpha_{drive,fem}$	0.1152	2.50	0.1136	2.66
$\alpha_{drive,u18}$	-1.0330	-13.02	-0.9269	-8.86
$\alpha_{drive,o64}$	0.5492	8.49	0.6639	7.54
$\alpha_{drive,cars}$	1.5264	64.13	1.5609	11.37
τ			15.7116	27.54
ϕ_1			0.0040	12.33
ϕ_2			0.0154	3.38
σ_ε			1.0000	NA

These results demonstrate that the DFT model substantially outperforms the MNL model, with an additional

3 parameters resulting in an improvement in log-likelihood of 1,270 units. It is notable that all estimates are of the expected sign with no discrepancies between MNL and DFT, which also have similar parameter ratios. Both ϕ parameters in the DFT model are significantly different from zero, suggesting that the driving factor in DFT's better performance may be the fact that it can capture the competition between similar (at the attribute-level) alternatives.

3.2.4. Heteroskedastic DFT results

We next look at the model outputs from across our full set of models, with results given in Table 2. This table also gives outputs from models including elasticities, the relative importance of attributes, average probabilities for choosing each alternative under the model and some policy forecasts (formulae for estimating these are given in Appendix B.

TABLE 2 : Results from all models for the LPMC dataset.

		MNL	DFT	CNL	Het-MNL	Het-DFT	Scobit	Un. Logit	As. Logit
Model fit	Log-likelihood (estimation)	-39,917.79	-38,647.77	-38,634.73	-38,947.79	-38,055.73	-38,919.04	-38,811.26	-39,050.19
	Estimated parameters	18	22	29	21	25	22	22	21
	Adj. ρ^2	0.4539	0.4710	0.4714	0.4669	0.4790	0.4672	0.4687	0.4655
	BIC	80,031.29	77,534.75	77,584.78	77,937.58	76,344.50	78,077.28	77,861.72	78,328.72
	Log-likelihood (holdout)	-4,428.57	-4,285.31	-4,275.70	-4,308.77	-4,206.09	-4,285.57	-4,272.43	-4,306.87
Elasticities for car	EC_{cost}	-0.0717	-0.0682	-0.0672	-0.0547	-0.0593	-0.0566	-0.0512	-0.0575
	EC_{tt}	-0.3004	-0.2248	-0.2916	-0.4424	-0.2843	-0.3718	-0.3813	-0.3289
	EC_{tv}	-0.3674	-0.2682	-0.2902	-0.2577	-0.2363	-0.2865	-0.2844	-0.3425
Relative importances	RI_{tt} (£/hr)	16.84	13.17	17.75	33.61	19.03	27.48	31.62	23.29
	RI_{ivtt} (£/hr)	35.35	50.60	30.92	79.45	53.69	88.27	77.22	39.17
	RI_{tv}	18.15	18.10	15.47	16.68	19.24	18.30	20.47	21.39
Average probability	P_{Walk}	16.64%	16.49%	16.40%	16.64%	16.52%	16.64%	16.64%	16.64%
	P_{Cycle}	3.20%	3.57%	3.21%	3.20%	3.19%	3.20%	3.20%	3.20%
	$P_{Transit}$	37.14%	36.59%	37.44%	37.14%	37.16%	37.14%	37.14%	37.14%
	P_{Drive}	43.02%	43.35%	42.95%	43.02%	43.14%	43.02%	43.02%	43.02%
Forecasts	5% transit cost	-0.59%	-0.62%	-0.64%	-0.18%	-0.50%	-0.17%	-0.17%	-0.36%
	No bias winter cycling	8.94%	11.64%	9.12%	8.67%	6.62%	8.39%	8.62%	9.08%

With regards to model fit, the best performing model is our new heteroskedastic DFT (het-DFT) model, which finds a gain in log-likelihood of 592 units at the cost of only 3 additional parameters over the basic DFT model. The next best performing model is the cross nested logit (CNL) model,[1] which substantially outperforms the basic MNL model. Scobit, uneven logit and asymmetric logit all perform similarly but slightly worse than the basic DFT model, and record substantially worse model fits that the heteroskedastic DFT or CNL models. Out-of-sample log-likelihoods return similar results to the estimation results.

Table 2 also gives arc elasticities for a 10% increase in car cost, car time or the amount of traffic variability on roads. All models return relatively low car cost elasticities. Some differences across the models are observed for time and traffic elasticities, with all other models predicting less change relative to MNL under increased traffic variability. The relative importance measures (which do not have an econometric interpretation under DFT models) also vary substantially, with DFT models suggesting higher costs associated with out-of-vehicle travel time relative to the base MNL model. All models give approximately 2-3 times higher ratios for out-of-vehicle compared to in-vehicle travel times. We also compare model forecasts for a 5% increase in transit cost (where DFT models again give similar forecasted reductions in choice of transit to MNL) and the forecast for the increase in cycling if there was not a bias against choosing to cycle in winter. This is where the DFT and het-DFT models differ most significantly, with DFT giving the largest increase in cycling, but het-DFT giving the smallest increase. Given the variable impact of initial preference parameters in DFT models, the average probability of choosing each alternative under the model does not reflect observed

[1] Note that we fix nesting parameters to a value of 1 if they are unacceptable values during testing.

market shares. However, results in Table 2 also demonstrate that mean probabilities under het-DFT are closer to observed shares than they are under DFT.

3.3. Second case study

The second dataset comes from a large-scale survey conducted as part of the DECISIONS project carried out by the Choice Modelling Centre, Leeds University (Calastri et al., 2020). A number of elements of life and travel patterns were recorded including participants' mobility patterns, in-home and out of-home-activities and information on their social networks. The data used for this work corresponds to the observed mode choice behaviour where after extensive data cleaning and data enrichment (Tsoleridis et al., 2019), 12,524 trips made by 540 individuals remained. For each trip, individuals travelled by one of six modes: car, bus, rail, taxi, cycling or walking. Again, we set aside the observations for 10% of individuals, leaving 11,176 observations in our estimation subset.

3.3.1. Utility model specifications

For our utility models, we follow the specification by Tsoleridis (2019). In this case, we include a number of attribute-specific marginal utility parameters, as well as sociodemographic variables, where we again do not list alternative specific constants in the utilities below, though they are estimated.

The utility for walking is defined:

$$V_{ns,walk} = \beta_{walk,tt} \cdot x_{walk,tt} \tag{13a}$$

$$+ \alpha_{walk,age} \cdot n_{age,18,29} + \alpha_{walk,edu} \cdot n_{edu}, \tag{13b}$$

where $n_{age,18,29}$ is a dummy variable taking the value of 1 if the individual is between the ages of 18 and 29, and n_{edu} is a dummy variable set to 1 for individuals, who do not have a university degree.

The utility for cycling is defined:

$$V_{ns,cycle} = \beta_{cycle,tt} \cdot x_{cycle,tt}. \tag{14}$$

The utility for driving is defined:

$$V_{ns,drive} = \beta_{drive,ivtt} \cdot x_{drive,ivtt} + \beta_{logcost} \cdot ln(x_{drive,cost}) \tag{15}$$

where the logarithm of cost is used after a cost damping effect (Daly, 2010) was found in initial specification testing (Tsoleridis, 2019).

The utility for train is defined:

$$V_{ns,train} = \beta_{train,ivtt} \cdot x_{train,ivtt} + \beta_{train,ovtt} \cdot x_{train,ovtt} + \beta_{logcost} \cdot ln(x_{train,cost}) \tag{16}$$

where $x_{train,ivtt}$ and $x_{train,ovtt}$ are in-vehicle and out-of-vehicle travel times for travelling by train, respectively.

The utility for bus is defined:

$$V_{ns,bus} = \beta_{bus,ivtt} \cdot x_{bus,ivtt} + \beta_{bus,ovtt} \cdot x_{bus,ovtt} + \beta_{logcost} \cdot ln(x_{bus,cost}) \tag{17}$$

where $x_{bus,ivtt}$ and $x_{bus,ovtt}$ are in-vehicle and out-of-vehicle travel times for travelling by bus, respectively.

Finally, the utility for taxi is defined:

$$V_{ns,taxi} = \beta_{taxi,tt} \cdot x_{taxi,tt} + \beta_{logcost} \cdot ln(x_{taxi,cost}) \tag{18a}$$

$$+ \alpha_{taxi,age1} \cdot n_{age,18,24} + \alpha_{taxi,age2} \cdot n_{age,25,29} + \alpha_{taxi,male} \cdot (1 - n_{fem}) + \alpha_{taxi,edu} \cdot n_{olevel}, \tag{18b}$$

where $x_{taxi,tt}$ is the travel time for travelling by taxi, $n_{age,18,24}$ and $n_{age,25,29}$ are dummy variables that take the value of 1 if the participant is aged 18-24 and 25-29, respectively, and n_{olevel} is a dummy variable that takes a value of 1 if the highest level of education obtained by the individual is O-level or equivalent.

3.3.2. DFT Specifications

For our DFT model, we define the attribute matrix M_{ns} using the same set of variables as in our utility specifications. Note that as we have attribute-specific marginal utility parameters, we incorporate these directly within M_{ns}. We thus have:

$$M_{ns} = \begin{bmatrix} \beta_{walk,tt} \cdot x_{walk,tt} & 0 & 0 \\ \beta_{cycle,tt} \cdot x_{cycle,tt} & 0 & 0 \\ \beta_{drive,ivtt} \cdot x_{drive,ivtt} & 0 & \beta_{logcost} \cdot ln(x_{drive,cost}) \\ \beta_{train,ivtt} \cdot x_{train,ivtt} & \beta_{train,ovtt} \cdot x_{train,ovtt} & \beta_{logcost} \cdot ln(x_{train,cost}) \\ \beta_{bus,ivtt} \cdot x_{bus,ivtt} & \beta_{bus,ovtt} \cdot x_{bus,ovtt} & \beta_{logcost} \cdot ln(x_{bus,cost}) \\ \beta_{taxi,tt} \cdot x_{taxi,tt} & 0 & \beta_{logcost} \cdot ln(x_{drive,taxi}) \end{bmatrix}. \tag{19}$$

For this version of DFT, we incorporate the alternative specific constants and the sociodemographic effects (α) in the initial preference matrix $P_{ns,\tau}$.

3.3.3. Basic model results

We first give the results of our basic MNL and DFT models, with model outputs and parameter estimates given in Table 3.

As was the case with the first dataset, we observe that DFT substantially outperforms MNL despite having just four additional parameters, though the difference for the holdout sample is small. However, as a contrast to the first dataset, we observe some significant differences in the estimates between MNL and DFT. Firstly, the alternative specific constants for rail and taxi are not statistically significant under the DFT model. Secondly, whilst the α and β parameter estimates are of the same sign, we observe some clear differences in scale, particularly for the β parameters. For example, the ratio $\beta_{bus_{ovtt}}/\beta_{bus_{ivtt}}$ is 2.3 for MNL and 4.2 for DFT, implying that the DFT model suggests a larger taste difference between the disutility for in and out-of-vehicle travel times. Furthermore, the ratio for $\beta_{cycle,tt}/\beta_{walk,tt}$ is 0.51 for MNL and 1.05 for DFT, which suggests that under a DFT model, the time sensitivity of walking is equivalent to cycling, but that cycling time sensitivity is lower under the MNL model.

3.3.4. Heteroskedastic DFT results

We next look at the model performance from our full set of models, with results detailed in Table 4. As before, this table also gives a wider set of model outputs.

The heteroskedastic DFT model again outperforms alternative models, with a 46 unit improvement in log-likelihood over the basic DFT model at a cost of 5 parameters. Results are similar to those from Table 2,

TABLE 3 : Basic model results for the Decisions mode choice dataset.

	MNL		DFT	
Log-likelihood (estimation)	-4,483.69		-4,387.46	
Estimated parameters	20		24	
Adj. ρ^2	0.6635		0.6704	
BIC	9,153.80		8,998.64	
Log-likelihood (holdout)	-584.55		-581.26	
	est	rob. t-rat.	est	rob. t-rat.
δ_{all}	0.0000	NA	1.2394	0.2908
δ_{drive}	0.0000	NA	0.0000	NA
δ_{bus}	-1.9785	-6.86	-4.3699	-6.31
δ_{rail}	-1.0203	-2.57	-0.2952	-0.30
δ_{taxi}	-2.3479	-6.96	0.8738	0.46
δ_{cycle}	-3.2842	-7.68	-7.1348	-4.26
δ_{walk}	1.2853	4.79	3.2219	3.75
$\alpha_{taxi,male}$	-0.7888	-2.35	-2.3638	-2.54
$\alpha_{taxi,age1}$	1.4047	4.51	5.6161	5.12
$\alpha_{taxi,age2}$	0.7568	1.86	3.3656	2.48
$\alpha_{taxi,edu}$	-1.2467	-2.52	-3.1019	-1.96
$\alpha_{walk,age}$	0.8909	3.54	3.8467	4.61
$\alpha_{walk,edu}$	-0.9748	-4.36	-2.8104	-3.93
$\beta_{drive,ivtt}$	-0.1064	-7.65	-0.0606	-5.25
$\beta_{bus,ivtt}$	-0.0431	-6.61	-0.0245	-4.88
$\beta_{bus,ovtt}$	-0.1008	-3.47	-0.1028	-5.92
$\beta_{train,ivtt}$	-0.0407	-3.18	-0.0313	-3.41
$\beta_{train,ovtt}$	-0.0978	-7.01	-0.0834	-6.90
$\beta_{taxi,tt}$	-0.1356	-5.13	-0.2056	-3.20
$\beta_{cycle,tt}$	-0.0753	-5.91	-0.0724	-4.10
$\beta_{walk,tt}$	-0.1490	-14.18	-0.0691	-8.32
$\beta_{logcost}$	-0.8923	-10.69	-0.5097	-7.16
τ			10.9586	5.39
ϕ_1			0.2013	1.76
ϕ_2			0.0683	1.23
σ_ε			1.0000	NA

TABLE 4 : Results from all models for the Decisions mode choice dataset.

		MNL	DFT	CNL	Het-MNL	Het-DFT	Scobit	Un. Logit	As. Logit
Model fit	Log-likelihood (estimation)	-4,483.69	-4,387.46	-4,468.06	-4,445.07	-4,341.58	-4,438.04	-4,443.44	-4,466.01
	Estimated parameters	20	24	25	23	29	26	26	25
	Adj. ρ^2	0.6635	0.6704	0.6643	0.6662	0.6735	0.6665	0.6661	0.6645
	BIC	9,153.80	8,998.64	9,169.16	9,104.54	8,953.48	9,118.45	9,129.24	9,165.05
	Log-likelihood (holdout)	-584.55	-581.26	-580.01	-582.70	-578.02	-582.96	-582.48	-581.94
Elasticities for car	EC_{cost}	-0.0700	-0.0682	-0.0671	-0.0836	-0.0620	-0.0862	-0.0785	-0.0636
	EC_{time}	-0.1625	-0.1467	-0.1635	-0.1393	-0.1479	-0.1324	-0.1366	-0.1560
Relative importances (£/hr)	$RI_{drive,ivtt}$	10.83	10.79	11.29	7.70	11.90	7.16	7.59	10.24
	$RI_{bus,ivtt}$	4.39	4.37	4.58	4.35	4.86	4.55	4.17	4.01
	$RI_{bus,ovtt}$	10.26	18.32	10.75	9.57	19.88	10.95	9.15	8.88
	$RI_{train,ivtt}$	4.14	5.58	4.55	12.75	6.33	14.91	11.59	5.80
	$RI_{train,ovtt}$	9.96	14.86	9.63	24.64	15.79	31.09	22.01	12.48
	$RI_{taxi,tt}$	13.80	36.64	14.27	3.56	37.01	3.70	4.25	17.95
	$RI_{cycle,tt}$	7.66	12.91	7.53	5.73	11.58	6.30	30.11	9.68
	$RI_{walk,tt}$	15.16	12.31	15.38	8.47	12.46	21.98	14.66	19.09
Average probability	P_{Drive}	48.49%	48.49%	48.49%	48.49%	48.65%	48.49%	48.49%	48.49%
	P_{Bus}	14.75%	14.65%	14.91%	14.75%	14.69%	14.75%	14.75%	14.74%
	P_{Rail}	5.04%	4.92%	4.92%	5.04%	4.91%	5.04%	5.04%	5.04%
	P_{Taxi}	3.11%	3.26%	3.11%	3.11%	3.20%	3.11%	3.11%	3.11%
	P_{Cycle}	3.20%	3.34%	3.23%	3.20%	3.18%	3.20%	3.20%	3.20%
	P_{Walk}	25.41%	25.33%	25.34%	25.41%	25.37%	25.41%	25.41%	25.41%
Forecasts	5% rail cost	-1.87%	-1.70%	-1.92%	-0.74%	-1.63%	-0.61%	-0.82%	-1.47%
	5% bus cost	-1.95%	-1.86%	-1.91%	-1.96%	-1.75%	-1.87%	-2.00%	-2.05%

with the exception that scobit this time outperforms uneven logit. The holdout log-likelihoods are, however, very similar across all models.

Table 4 also gives outputs including elasticities, which are similar across models. Notably, the elasticities for car cost are remarkably similar to those from the LPMC dataset. As a contrast, car time elasticities have a smaller magnitude. Relative importances are also often similar, though DFT models assign greater importance to out-of-vehicle travel times. A clear exception is the value for taxi, which varies substantially across models. Finally, 5% increases in the cost of rail or bus results in around 2% decreases in the uptake of these alternatives. This suggests that survey participants in the second case study are more price sensitive than those in the first case study, which is unsurprising given the high costs associated with living in London. As before, het-DFT returns average probabilities that are closer to market shares than values observed under DFT.

4. Conclusions and next steps

In this paper, we detail theoretical steps for improving the flexibility of decision field theory. In particular, we focus on relaxing the assumption of identical normal error terms of across alternatives. This 'heteroskedastic' version of DFT substantially outperforms the previous version of DFT on two large-scale revealed preference mode choice datasets. Notably, tests of the model on stated preference datasets (the Danish and UK datasets tested in Hancock et al. 2021) did not result in significant improvements in model fit, implying that this model is a step towards capturing real-life behaviour, where the relative shares of chosen alternatives may be very unbalanced. Further tests are of course required. For example, tests on simulated datasets may confirm the importance of unbalanced shares of chosen alternatives on finding significant improvements through moving to a heteroskedastic DFT. Additionally, measures from models such as elasticities could be compared in future applications.

Future work should also further investigate the concept of a 'nested DFT', which may be able to incorporate the ideas of nested logit models through allowing correlations between the error terms for different alternatives within the DFT model. Future models could also test, for example, the incorporation of random parameters or error components, which may alter the relative performance of the different models. Though it is clear that there are many routes for future work, the results in this paper prove that decision field theory is no longer a model just for laboratory or stated preference data.

References

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.

Brathwaite, T. and Walker, J. L. (2018). Asymmetric, closed-form, finite-parameter models of multinomial choice. *Journal of choice modelling*, 29:78–112.

Busemeyer, J. R., Jessup, R. K., Johnson, J. G., and Townsend, J. T. (2006). Building bridges between neural models and complex decision making behaviour. *Neural Networks*, 19(8):1047–1058.

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.

Calastri, C., dit Sourd, R. C., and Hess, S. (2020). We want it all: experiences from a survey seeking to cap-

ture social network structures, lifetime events and short-term travel and activity planning. *Transportation*, 47(1):175–201.

Daly, A. (2010). Cost damping in travel demand models. Technical report, RAND Corporation.

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

Hancock, T. O. (2019). *Travel behaviour modelling at the interface between econometrics and mathematical psychology*. PhD thesis, University of Leeds.

Hancock, T. O., Hess, S., and Choudhury, C. F. (2019). An accumulation of preference: two alternative dynamic models for understanding transport choices. *Submitted*.

Hancock, T. O., Hess, S., Marley, A., and Choudhury, C. F. (2021). An accumulation of preference: Two alternative dynamic models for understanding transport choices. *Transportation Research Part B: Methodological*, 149:250–282.

Hess, S. and Palma, D. (2019). Apollo: A flexible, powerful and customisable freeware package for choice model estimation and application. *Journal of Choice Modelling*, 32:100170.

Hillel, T., Elshafie, M. Z., and Jin, Y. (2018). Recreating passenger mode choice-sets for transport simulation: A case study of London, UK. *Proceedings of the Institution of Civil Engineers-Smart Infrastructure and Construction*, 171(1):29–42.

Krueger, R., Bansal, P., Bierlaire, M., and Gasos, T. (2020). Robust discrete choice models with t-distributed kernel errors. *arXiv preprint arXiv:2009.06383*.

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.

Pettibone, J. C. (2012). Testing the effect of time pressure on asymmetric dominance and compromise decoys in choice. *Judgment and Decision making*, 7(4):513.

Ratcliff, R., Cherian, A., and Segraves, M. (2003). A comparison of macaque behavior and superior colliculus neuronal activity to predictions from models of two-choice decisions. *Journal of neurophysiology*, 90(3):1392–1407.

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.

Tsoleridis, P. (2019). Contrasting and combining discrete choice modelling and machine learning algorithms in the context of mode-destination choices. *Institute for Transport Studies, University of Leeds. PhD Transfer Report*.

Tsoleridis, P., Choudhury, C. F., and Hess, S. (2019). Capturing heterogeneity in mode choice decisions: Comparing and combining machine learning and discrete choice models. *Institute for Transport Studies, University of Leeds. Working paper*.

Wen, C.-H. and Koppelman, F. S. (2001). The generalized nested logit model. *Transportation Research Part B: Methodological*, 35(7):627–641.

Appendix A: Models for comparison

A.1. Multinomial logit

For both case studies, we start with a standard random utility model with the typical multinomial logit formulation, i.e, where we have utilities (V_{nsj}) for the different alternatives and the assumption of type I

extreme value error terms results in probabilities of choosing each alternative:

$$P_{nsj} = \frac{\mu_J \cdot exp(V_{nsj} + \delta_j)}{\sum_{i=1}^{J_{ns}} (\mu_J \cdot exp(V_{nsi} + \delta_i))}, \tag{20}$$

where there are J_{ns} alternatives for individual n in choice context s, and we have alternative specific constants, δ_j. Under our multinomial logit model, the alternative specific scaling parameters are fixed to 1.

A.2. Heteroskedastic MNL

There are multiple possibilities for specifying heteroskedastic extensions of multinomial logit models. For the work in this paper, we define a heteroskedastic MNL model such that it has equivalent flexibility to our heteroskedastic DFT model, meaning that it has an error structure that allows for different variances across the different alternatives. We thus use Equation 20 but estimate the alternative specific scaling parameters, μ_J, where one must be fixed for model identification.

A.3. Cross-nested logit

For the cross-nested logit models, we follow the 'generalised nested logit' model by Wen and Koppelman (2001). We thus have all nesting parameters freely estimated and we put constraints on the allocation parameters, where $0 \leq \alpha_{j,m} \leq 1$ and $\sum_j \alpha_{j,m} = 1, \forall m$, with each α giving the membership of alternative j in nest m. For our nesting structure, we define a nest (S_1 to S_M) for each and every pair of alternatives, meaning that we have a total of $M = (J)(J-1)/2$ nests, where J is the number of alternatives. The probability for choosing an alternative j can then be calculated as a sum over nests with:

$$P_{nsj} = \sum_{m=1}^{M} \left(P_{S_m,ns} \cdot P_{(j|S_m),ns} \right), \tag{21}$$

where we have the probability of being within nest m given by:

$$P_{S_m,ns} = \frac{\left(\sum_{i \in S_m} (\alpha_{i,m} e^{(V_{nsi} + \delta_i)})^{\frac{1}{\lambda_m}} \right)^{\lambda_m}}{\sum_{l=1}^{M} \left(\sum_{i \in S_l} (\alpha_{i,l} e^{(V_{nsi} + \delta_i)})^{\frac{1}{\lambda_l}} \right)^{\lambda_l}}, \tag{22}$$

and the probability of choosing alternative j within nest m given by:

$$P_{(j|S_m),ns} = \frac{(\alpha_{j,m} e^{(V_{nsj} + \delta_j)})^{\frac{1}{\lambda_m}}}{\sum_{i \in S_m} (\alpha_{i,m} e^{(V_{nsi} + \delta_i)})^{\frac{1}{\lambda_m}}}, \tag{23}$$

where λ_m is the estimated nesting parameter for nest m.

A.4. Scobit

The probability of choosing alternative j under a multinomial scobit model (Brathwaite and Walker, 2018) is defined:

$$P_{nsj} = \frac{exp\left(\delta_j - ln[(1 + e^{-V_{nsj}})^{\gamma_j} - 1]\right)}{\sum_{i=1}^{J_{ns}} \left(exp\left(\delta_i - ln[(1 + e^{-V_{nsi}})^{\gamma_i} - 1]\right)\right)}, \tag{24}$$

where each alternative j has an associated estimated scalar γ_j.

A.5. Uneven logit

The probability of choosing alternative j under an uneven logit model (Brathwaite and Walker, 2018) is defined:

$$P_{nsj} = \frac{exp\left[\delta_j + V_{nsj} + ln(1+e^{-V_{nsj}}) - ln(1+e^{-\gamma_j \cdot V_{nsj}})\right]}{\sum_{i=1}^{J_{ns}} \left(exp\left[\delta_i + V_{nsi} + ln(1+e^{-V_{nsi}}) - ln(1+e^{-\gamma_i \cdot V_{nsi}})\right]\right)}, \tag{25}$$

where each alternative j has an associated estimated shape parameter γ_j.

A.6. Asymmetric logit

The final model that we test is also introduced by Brathwaite and Walker (2018) and is called the asymmetric logit model. Under this model, the probability of choosing alternative j is defined:

$$P_{nsj} = \frac{exp\left[\delta_j + S_{nsj}\right]}{\sum_{i=1}^{J_{ns}} \left(exp\left[\delta_j + S_{nsi}\right]\right)}, \tag{26}$$

where $S_{nsj} = ln(\gamma_j) - V_{nsj} \cdot ln(\gamma_j)$ if $V_{nsj} \geq 0$ and $S_{nsj} = ln(\gamma_j) - V_{nsj} \cdot ln(\frac{1-\gamma_j}{J-1})$ if $V_{nsj} < 0$, where $J = max(J_{ns})$. Unlike scobit and uneven logit, asymmetric logit has restrictions on the shape parameters with $0 \leq \gamma_j \leq 1$ and $\sum_J \gamma_j = 1$.

Appendix B: Model outputs

B.1. Elasticities

For all models, we estimate arc elasticities for car (AEC) with:

$$AEC = \log\left(\frac{\text{Forecasted Car Trips}}{\text{Base Car Trips}}\right) / \log(1.1), \tag{27}$$

where 'Base Car Trips' is calculated as the sum over the probabilities of choosing car across all choice tasks in the dataset, with 'Forecasted Car Trips' calculated equivalently but with adjusted attributes. For both datasets, we estimate cost and time elasticities by using a 10% increase of the controlling factor. For the LPMC dataset, we additionally test for a 10% increase in traffic variability (see (Hillel et al., 2018) for details on how variability is estimated).

B.2. Relative importances

The relative importance (RI) of an attribute, k, is calculated with respect to cost, such that:

$$RI_k = \frac{\beta_k}{\beta_{cost}}. \tag{28}$$

For the Decisions dataset, these values are multiplied by 60 to convert from minutes to hours, and by a representative cost of £1.51 (to account for the fact that logarithmic transforms of cost are used). Note that the relative importance of time is equivalent to the value of time under all models except the DFT and Het-DFT models.

B.3. Average probabilities

Average probabilities are calculated by averaging across the probability of an alternative being chosen under a model across the full set of choice tasks.

B.4. Forecasts

We also obtain forecasts given a 5% increase in the cost of public transport. We calculate the forecasted relative reduction (*RR*) in the amount the alternative (*i*) is chosen with:

$$RR_i = \left(\frac{\text{Forecasted Trips}_\text{i}}{\text{Base Trips}_\text{i}} \right) - 1, \tag{29}$$

where the trips are calculated as the sum over the probabilities of choosing alternative *i* across all choice tasks in the dataset. For the LPMC, we also estimate the relative increase in share of cycling if there is no winter bias. For this scenario, 'Forecasted Trips' are calculated with the model estimates but with $\alpha_{cycle,winter} = 0$.